

Example-Based Photometric Stereo: Shape Reconstruction with General, Varying BRDFs

Aaron Hertzmann and Steven M. Seitz, *Member, IEEE*

Abstract—This paper presents a technique for computing the geometry of objects with general reflectance properties from images. For surfaces with varying material properties, a full segmentation into different material types is also computed. It is assumed that the camera viewpoint is fixed, but the illumination varies over the input sequence. It is also assumed that one or more example objects with similar materials and known geometry are imaged under the same illumination conditions. Unlike most previous work in shape reconstruction, this technique can handle objects with arbitrary and spatially-varying BRDFs. Furthermore, the approach works for arbitrary distant and unknown lighting environments. Finally, almost no calibration is needed, making the approach exceptionally simple to apply.

Index Terms—Photometric stereo, shape reconstruction, shape-from-shading, bidirectional reflectance distribution function (BRDF), specular materials, clustering materials.

1 INTRODUCTION

AN important unsolved problem in computer vision is modeling scenes with real-world materials. Although real objects reflect light in a wide range of interesting ways [8], most dense shape reconstruction techniques in the computer vision literature assume that the scene is Lambertian. The human visual system operates effectively with a much wider variety of materials—we have no problem interpreting shiny objects and other surfaces that reflect light in varied ways. This suggests that other cues may exist and significant improvements in shape reconstruction algorithms are possible.

In this paper, we introduce *orientation-consistency*, a cue for interpreting scenes with arbitrary reflectance properties. Orientation-consistency states that, under the right conditions, two points with the same surface orientation must have the same or similar appearance in an image. Based on this cue, we show how to reconstruct the surface normals of a target object, when imaged together with one or more reference objects of similar materials and known shape. Our approach operates by finding, for each point on the target object, a matching point (and normal) on the reference object(s) that is most orientation-consistent. Because orientation-consistency does not depend on illumination direction, we can capture several images with different illuminations and require orientation-consistency over *all* images to achieve very robust matches. A dense shape

reconstruction is obtained by integrating the recovered normal field.

This approach has the following features:

- The BRDF, illumination, and shape may *all* be unknown. Moreover, the BRDF may be arbitrary, and can vary (with some restrictions, described below) over the surface. The approach will work with any number of distant point or area light sources.
- Almost no calibration of the camera or lighting environment is needed.
- A segmentation of the object into different materials is computed.
- The algorithm is extremely simple to implement, yet operates on a broader class of objects than any previous photometric stereo technique, and achieves impressive results for a wide variety of challenging materials ranging from specular objects to velvet and brushed fur. The quality of results on shiny objects compares favorably to the performance of commercial laser scanners on diffuse objects.

We also make the following assumptions: first, one or more reference objects of the same or similar materials must be imaged under the same illumination as the target object. All BRDFs on the target object must (approximately) lie in the linear span of the BRDFs of the reference object(s). This assumption conveniently limits the underlying parameter space, while still allowing for an infinite range of BRDFs. In practice, we have found it to be an accurate model for a wide range of real-world objects. Second, the camera is assumed to be orthographic. Third, shadows, interreflections, and other nonlocal reflectance or transmittance effects are ignored. Finally, like other photometric stereo methods, the scene is assumed to be continuously differentiable.

1.1 Relation to Previous Work

Our use of reference objects builds upon early work on photometric stereo. In particular, Silver [26] used images of

• A. Hertzmann is with the Computer Science Department, University of Toronto, 10 King's College Road, Room 3302, Toronto, ON M5S 3G4 Canada. E-mail: hertzman@dgp.toronto.edu.

• S.M. Seitz is with the Department of Computer Science and Engineering, University of Washington, Box 352350, Seattle, WA 98195-2350. E-mail: seitz@cs.washington.edu.

Manuscript received 4 Dec. 2003; revised 24 June 2004; accepted 24 Sept. 2004; published online 13 June 2005.

Recommended for acceptance by R. Basri.

For information on obtaining reprints of this article, please send e-mail to: tpami@computer.org, and reference IEEECS Log Number TPAMI-0398-1203.



Fig. 1. Orientation-consistency. A sphere and a bottle are painted with the same material and lit the same way. The orientation-consistency cue says that any two surface points with the same normal have the same color. In particular, for each of the above images (acquired with a single distant point light source), all points that are in highlight have the same surface normal.

a wooden sphere to reconstruct other wooden objects imaged under the same lighting, based on an idea proposed by Woodham [37]. By specifying the outgoing radiance function for a hemisphere of directions, the sphere images provide an empirical model of reflection for wood. More recently, Woodham [34] developed a technique for real-time photometric stereo using a reference object. For a good overview of early work on photometric stereo, including the use of reference objects, see Horn's classic book [15].

Despite its simplicity, the use of reference objects in shape reconstruction was not widely adopted in the research community, due in part to the following problems: 1) the scene must have a single BRDF, 2) the reference object must be made of the same material as the target object, and 3) the need for a calibration object is undesirable. In an effort to overcome these limitations, most subsequent work on photometric stereo turned to using analytic rather than empirical models of surface reflection [3], [7], [12], [16], [22], [33] (for a good overview, see [31]). While some of these techniques have been shown to yield good results, they are currently restricted to simple diffuse and specular materials. It is known, however, that a wide range of real world materials do not fit these traditional BRDF models [8].

In this paper, we show that these difficulties are overcome by some simple observations and the use of a small number of reference objects (typically two) instead of one. The resulting approach operates on a broader class of objects than any previous photometric stereo technique.

In terms of its generality and the quality of the results obtained, our approach is similar to Helmholtz Stereopsis [18], [30], [35], [36] which also enables reconstructing surfaces with arbitrary BRDFs in the absence of shadows and interreflections. The two approaches have different requirements, however. Some advantages of the Helmholtz approach are that it does not require reference objects, orthography, or distant illumination. In addition, Helmholtz techniques can exploit multiple camera viewpoints to compute *absolute* depth, which is not possible with our photometric stereo technique. A disadvantage is that the Helmholtz approach requires a more complex experimental setup: both the camera and the light source must be calibrated and moved in a precise and controlled fashion (alternatively, several static cameras and light sources may be used) [35]. In contrast, our approach does not require that the viewpoint or illumination are known, and is flexible enough to be used in relatively unconstrained environments, e.g., for outside imaging using the sun and environment as a light source. By exploiting a fixed camera viewpoint, we directly obtain a reconstruction at the resolution of the input images and, thus, avoid errors due to camera calibration, image discretization,

and resampling that affect Helmholtz and other stereo techniques. An additional advantage of our approach is that it enables the use of area light sources, whereas the Helmholtz approach is restricted to point sources. Area sources are potentially useful for highly specular surfaces that do not have a strong diffuse component—reconstructing such surfaces with the Helmholtz method would require a very large number of images, enough for the highlight to appear at every surface point.

2 SHAPE BY EXAMPLE

Surface shading provides a great deal of information about surface geometry. However, shading also depends on illumination, camera geometry, and reflectance—varying any one of these factors may modify the appearance of the object. Without some way of separating the contributions of each of these factors, computing shape is difficult.

To address this difficulty, we propose a cue that, given certain assumptions, is *invariant* to the light source distribution, camera parameters, and reflectance function and depends purely on shape.

Orientation-consistency cue: two points with the same surface orientation reflect the same light toward the viewer.

Orientation-consistency holds when the following assumptions are satisfied: Both points have the same BRDF, the light sources are directional (i.e., distant), the camera is orthographic, and there are no shadows, interreflections, transparency effects, or other nonlocal effects that do not depend purely on the BRDF of a single point. Note that under these conditions, the BRDF, surface orientation, incident illumination, and viewer direction are the same for both points. It follows [15] that the outgoing radiance is the same for both points. Orientation-consistency is used implicitly in the photometric stereo literature, but in a more restricted form. We propose its use as a generic cue for image analysis and, to this end, describe generalizations that greatly increase its applicability.

This cue is helpful in shape perception, as it tells us quite a bit about the distribution of highlights on an object. For example, if a point is in highlight, any other point with the same orientation should also have a highlight. An example of this effect is shown in Fig. 1.

If we happen to *know* the surface orientation for some points in the scene, orientation-consistency allows us to propagate this knowledge to other points. In the example shown in Fig. 1, we can determine the orientation of each point on the bottle in highlight by finding the corresponding point on the sphere that is also in highlight. In general,

any reference object may be substituted for the sphere, provided the shape of the reference is known and it contains a sufficient distribution of surface orientations. However, the correspondence is ambiguous when there are multiple highlights on the sphere, and for points not in highlight. More specifically, ambiguities arise when two points on the sphere have the same color, resulting in two surface orientations that cannot be distinguished from one another.

2.1 A Correspondence Approach to Photometric Stereo

We now describe an algorithm for reconstructing the shape of an object from multiple images, using a reference object. This algorithm is nearly identical to that presented by Silver [26]; however, we show that it can be applied to a surprisingly wide variety of objects including anisotropic materials (Silver's original formulation assumed isotropic materials and was demonstrated only for simple wooden objects). We also describe an effective technique for reducing runtime.

Suppose that we capture multiple images of the reference and target objects from the same viewpoint but under different illuminations. Let I_1^r, \dots, I_n^r be the reference images and I_1^t, \dots, I_n^t the target images. It is assumed that corresponding reference I_i^r and target I_i^t images are captured under the same illumination. For each pixel position $\mathbf{p} = (x, y)$ in a reference or target image, let $I_{\mathbf{p}}$ be the intensity of that pixel and define the *observation vector* (abbreviated OV, or as a vector \mathbf{V}) to be the set of intensities observed at that pixel over the n images:

$$\mathbf{V}_{\mathbf{p}} = [I_{1,\mathbf{p}}, \dots, I_{n,\mathbf{p}}]^T. \quad (1)$$

For RGB color images, there is a component OV for each color channel. Let $\mathbf{R}_{\mathbf{p}}$, $\mathbf{G}_{\mathbf{p}}$, and $\mathbf{B}_{\mathbf{p}}$ be the OVs for the red, green, and blue channel, respectively. Define the cumulative OV for color images $\mathbf{V}_{\mathbf{p}}$ to be the concatenation of $\mathbf{R}_{\mathbf{p}}$, $\mathbf{G}_{\mathbf{p}}$, and $\mathbf{B}_{\mathbf{p}}$ into a single vector

$$\mathbf{V}_{\mathbf{p}} = [\mathbf{R}_{\mathbf{p}}^T, \mathbf{G}_{\mathbf{p}}^T, \mathbf{B}_{\mathbf{p}}^T]^T. \quad (2)$$

Given a pixel \mathbf{p} on the target object, the surface normal at \mathbf{p} is determined simply by searching for the pixel \mathbf{q} on the reference object with the best matching OV, i.e., \mathbf{q} minimizes $\|\mathbf{V}_{\mathbf{p}} - \mathbf{V}_{\mathbf{q}}\|$. A complete correspondence determines the normal for every pixel on the target object.

It is interesting to observe that this formulation casts photometric stereo as a two-image stereo matching problem—find pixel correspondences between an image of the reference object and an image of the target object. Although it is often argued that one of the virtues of photometric stereo is that it avoids the correspondence problem present in binocular stereo [33]; in fact, photometric stereo and binocular stereo can be viewed as addressing very similar correspondence problems. In cases where the reflectance map has a simple parametric form, model-fitting techniques [3], [16], [22], [33] can provide a more efficient alternative to the explicit search method presented here.

2.1.1 Implementation Issues

In order to determine the geometry of the reference sphere, we manually segment it from the background and then fit a circle to the silhouette. The target object is also manually segmented. These tasks could be fully automated, by careful choice of the background [27], [34]. The task of finding the

reference OV closest to the target OV is a nearest-neighbor search problem that may be accelerated with appropriate data structures. Several fast algorithms have been developed for nearest-neighbors computations based on k -d trees [1], [5], [11], [23]. We use the ANN algorithm [1], one such method with freely-available source code.

We then compute a 3D surface from the estimated surface orientations [3]. Given the set of normals $\{n_x, n_y, n_z\}$ for each point, we solve for the height field $z(x, y)$ that minimizes

$$\Psi(z) = \sum_{x,y} \left(n_z \frac{\partial z(x,y)}{\partial x} + n_x \right)^2 + \left(n_z \frac{\partial z(x,y)}{\partial y} + n_y \right)^2, \quad (3)$$

using the approximations

$$\frac{\partial z(x,y)}{\partial x} = z(x+1,y) - z(x,y), \quad (4)$$

$$\frac{\partial z(x,y)}{\partial y} = z(x,y+1) - z(x,y). \quad (5)$$

Note that the calibration requirements are extremely minimal—other than correcting the images for radial distortion or vignetting (steps that we did not perform), no geometric or radiometric calibration of the camera or light sources is needed.

The objects were imaged with a Canon D60 camera, using a 640mm lens to attain a good approximation to orthography. The scene was illuminated with a hand held spot light shone at the objects from a distance of 10 to 15 feet. The objects were all less than one foot high.

2.1.2 Results

In our first test, we spray-painted a sphere and a plastic bottle with the same shiny green paint, and applied the above algorithm. The eight input images are shown in Fig. 2; the resulting reconstruction is shown in Fig. 3. The algorithm accurately reconstructed finely-detailed geometry of the bottle, including wrinkles on the label of the bottle and indentations on the cap. Some artifacts occur in places where the assumptions of the algorithm are not satisfied, at self-shadows below the cap and underneath the bottle, and where paint has flaked off the cap. Nevertheless, the shape reconstruction is very accurate—the median distance from each point on the laser scan to our reconstructed mesh is 0.5mm. A plot comparing a slice through the two scans is shown in Fig. 3. Note that the laser scan has artifacts due to specular reflections, manifested as sharp discontinuities (spikes) in Fig. 3. To avoid such artifacts, shiny objects are commonly painted with a diffuse coating prior to laser scanning [6], [32]—we did not do this for the bottle, as we wanted to demonstrate these artifacts. Later in the paper, we show a comparison with a laser scan of an object that was painted, and problems that the painting process can introduce. Our approach avoids such artifacts and the need to paint the object.

Fig. 4 shows the impact that the number of images and distribution of light sources has on reconstruction quality. Using as few as three images can produce a reasonable reconstruction if the light source directions are broadly distributed, and adding more images further improves the result. Using even two images yields results that capture the rough overall shape.

One challenge is that the images are quite large: $398 \times 1,176$ for the bottle and 328×332 for the sphere. A brute

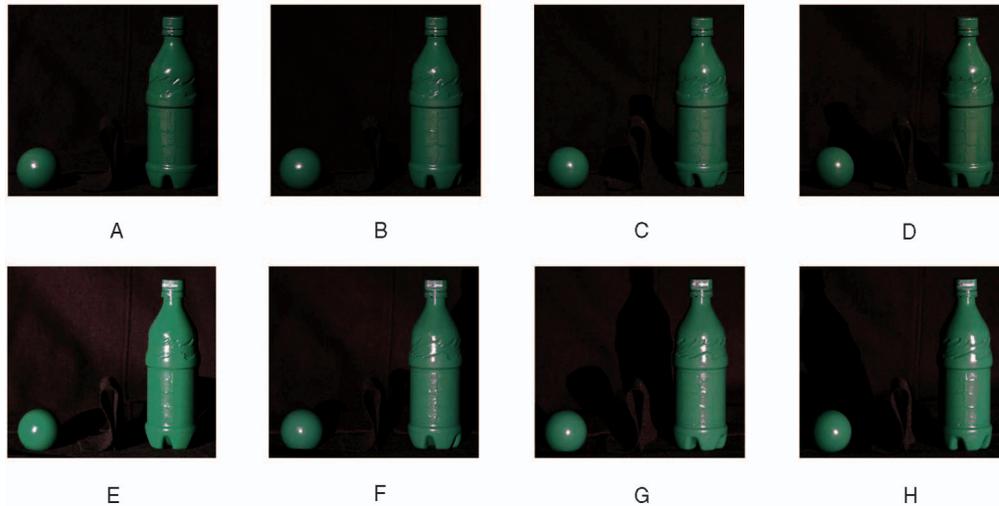


Fig. 2. The eight source images used for this experiment. The hand-held light moves from left to right across each row. For each horizontal position, the light is held at two heights, shown in the top and bottom rows, respectively.

force search to compute normals requires roughly one day of compute time. The accelerated method using ANN completed in about 5 minutes on a 2GHz Intel Xeon processor. An additional 30 minutes was required to compute the high-resolution mesh from the normals, but this could be accelerated using multigrid techniques or other fast linear solvers [28].

The same algorithm is applied to reconstructing a velvet surface in Fig. 5. Since we did not have access to a velvet sphere, we used a cylinder reference object instead, and constrained the surface to bend roughly along the vertical axis. Velvet is known to have a particularly unusual reflectance map that is brightest at grazing angles [38]. This particular velvet sample has a two-toned appearance—from oblique angles only the red threads are visible, while from the front, the material has a blue tone as more of the backing material shows through. The algorithm captured the shape very well, as seen in the figure (top view). The reconstruction is smooth except for a number of small wrinkles, many of which appear to reproduce features on the actual fabric—the velvet sample itself was wrinkled, having been folded and refolded several times. There are also a few wrinkles on one region of the cylinder reference object—this did not pose a problem, however, as we simply masked out that region of the reference images so it was not used in matching.

2.1.3 Anisotropic Materials

This algorithm can also be used to reconstruct materials with anisotropic BRDFs. In this case, orientation-consistency dictates that two pixels will have the same OV if they have the same surface normal *and* orientation about the normal. This means that the reference object should contain samples of many surface normals and rotations, so that there will be a good match for each possible surface orientation on the target.

Fig. 5 shows reconstruction of an anisotropic brushed-fur-like fabric. Note that the entire strip is made from a *single* material—the interesting pattern of highlights is due to varying orientations of the fibers, causing the BRDF to be rotated at different positions on the fabric. As in the velvet example above, we used a cylindrical reference object, thus restricting the problem to vertical surfaces.

To our knowledge, this is the first time in the computer vision literature that a shape reconstruction technique has been successfully applied to an anisotropic material. The reflectance in this example is sufficiently complex that it is very difficult for humans to perceive the shape, yet the algorithm performs quite well, as shown in the top view. Note that the brushed-fur fabric contained a couple of sharp creases, visible in Fig. 5b, and these were captured in the reconstruction (front view). Moreover, the algorithm qualitatively preserves fine-scale textural variations over the surface, as smoother regions appear smoother in the reconstruction. While not perfect, it is remarkable that the reconstruction is this good, given that the sample itself is not exactly vertical, whereas the reconstructed normals are all constrained to be perpendicular to the vertical axis.

Although we show results only for vertically-oriented surfaces, note that both components of the surface normal can be computed with this algorithm, given a reference object with the full range of surface orientations (or by imaging the cylinder rotated through multiple orientations).

3 MODELING MATERIAL VARIATION

Two limitations of the technique presented in Section 2 are 1) the target object must have a single albedo and 2) the reference object must be composed of exactly the same material as the target object. In this section, we describe how both of these restrictions are removed.

In the next section, we consider the case of surface texture due to albedo variations over a purely diffuse or specular object. The following section considers the case of more general material variations.

3.1 Color Variation

Consider a diffuse target object with a texture pattern that varies over the surface. The appearance of such an object may be approximated using the Lambertian formula:

$$I_p^t = \rho_p^t \mathbf{n}_p \cdot \sum_{\text{light } l} \mathbf{l}, \quad (6)$$

where \mathbf{n}_p is the surface normal at that point, \mathbf{l} encodes the light source direction and intensity, and ρ_p is a reflection coefficient that varies over the surface.

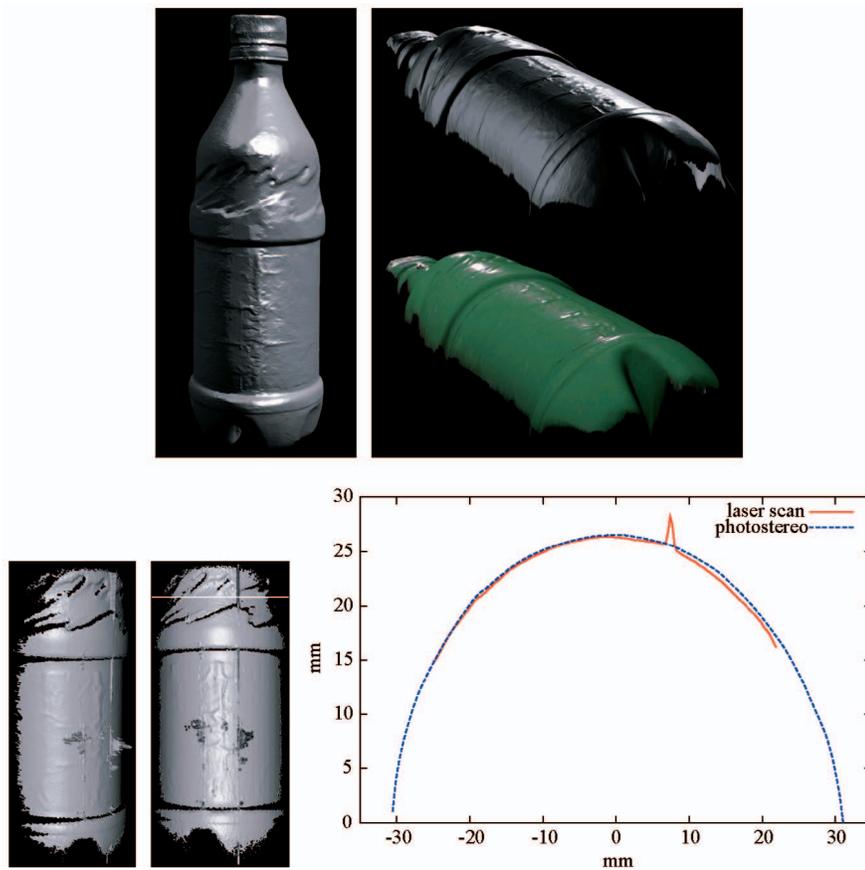


Fig. 3. Bottle reconstruction. *Top*: Shaded and texture-mapped renderings of the reconstructed bottle. Note the fine details such as the wrinkles in the label that are accurately captured. *Bottom*: Laser scan of the same bottle rendered from the side and front. Specular reflections cause spike artifacts in the laser scan. The graph at right compares both reconstructions, for the slice marked in white (note spike in the laser scan). Median error over the entire scan is 0.5mm.

Suppose we have a homogeneous diffuse reference object with reflection coefficient ρ^r . The appearance of this object may therefore be modeled as:

$$I_p^r = \rho^r \mathbf{n}_p \cdot \sum_{\text{light } l} \mathbf{l} \quad (7)$$

Now, suppose that a point \mathbf{p} on the target has the same surface orientation as a point \mathbf{q} on the reference object. It follows from (1), (6), and (7) that

$$\mathbf{V}_p^t = \frac{\rho_p^t}{\rho^r} \mathbf{V}_q^r \quad (8)$$

and, therefore,

$$\frac{\mathbf{V}_p^t}{\|\mathbf{V}_p^t\|} = \frac{\mathbf{V}_q^r}{\|\mathbf{V}_q^r\|} \quad (9)$$

Consequently, the algorithm in Section 2 can be applied if we normalize the target and reference OV's before matching.¹ In the case of color images, the OV's for each color channel can be normalized separately, before combining them into \mathbf{V}_p .

1. In general, we recommend the method in Section 3.3 over this normalization approach, as it will perform better in the presence of noise and dark albedos.

Note that the same approach can model attached shadows and purely specular surfaces. More generally, the approach applies for any surface having the form:

$$I_p = \rho_p f(\mathbf{n}_p, \mathbf{v}, \mathbf{L}), \quad (10)$$

where f is any reflectance map² as a function of \mathbf{v} , the direction to the viewer, and \mathbf{L} , the incident illumination field.

3.2 Material Variation

In the general case, the target object may be composed of multiple materials that vary over its surface. We make the assumption that all materials on a single object can be represented as a linear combination of k basis materials for some fixed value of k . For example, the Phong model is expressed as a linear combination of a diffuse and specular component, i.e., $k=2$. More generally, there is strong empirical evidence that a wide variety of reflectance maps may be represented as a linear combination of a small number of basis functions [19], [21], [25], [32].

Accordingly, suppose that the target surface has a reflectance map of the form:

2. We define *reflectance map* explicitly as a function of viewpoint and illumination, whereas in the traditional definition this dependence is implicit [15].

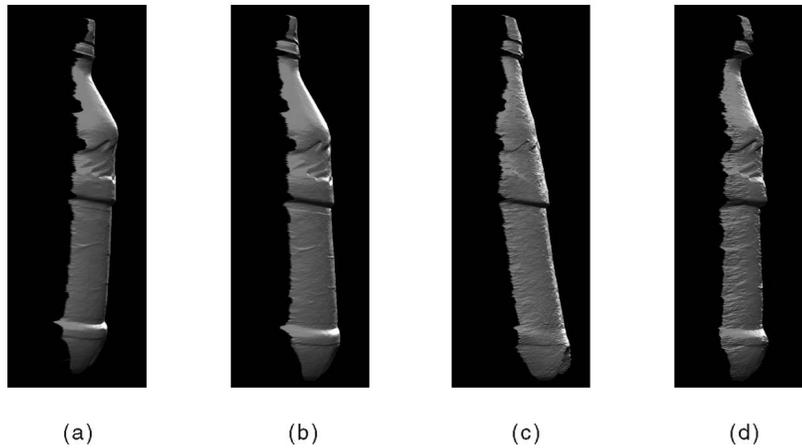


Fig. 4. Reconstruction with reduced numbers of illumination conditions. (a) All eight imaging conditions. (b) Three imaging conditions (using B, C, F). (c) Two source images, horizontally-separated light source (using B, C). (d) Two source images, vertically-separated light source (using C, H).

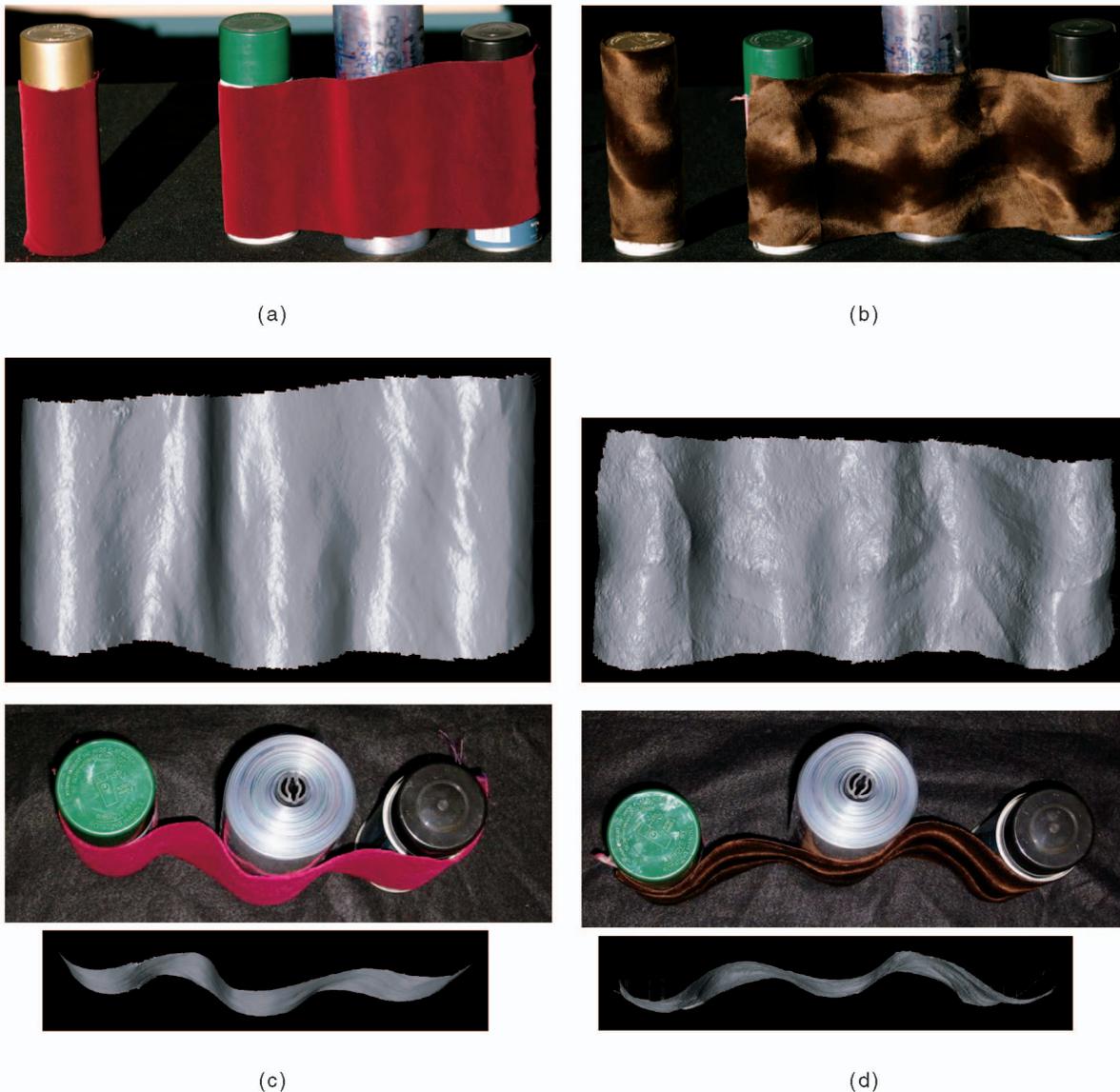


Fig. 5. Velvet and brushed fur reconstructions. (a) One of 10 input images of velvet. (b) One of 12 input images of brushed fur. In each input image, the cylindrical reference object is on the left, and the target object is on the right. (c) and (d) show frontal and overhead renderings of the reconstructions. The rendered viewpoints were manually chosen to match the photographs, to facilitate comparison. The reconstruction accurately captures both fine details such as creases (frontal view) and overall shape (overhead view).

$$I_{\mathbf{p}}^t = \sum_{i=1}^k \rho_{i,\mathbf{p}}^t f_i(\mathbf{n}_{\mathbf{p}}, \mathbf{v}, \mathbf{L}). \quad (11)$$

Note that here we assume that the image intensities measured by the camera are a linear function of scene radiances. It may be beneficial to calibrate and correct for nonlinearities [9], although we found it unnecessary for our experiments.

Now, instead of one reference object, suppose we had k reference objects r_1, \dots, r_k , with reflectance maps that are similar to the target object. Although the reference objects could in principle have spatially varying materials, we assume for simplicity that they are homogeneous, i.e.,

$$I_{\mathbf{q}}^{r_j} = \sum_{i=1}^k \rho_i^{r_j} f_i(\mathbf{n}_{\mathbf{q}}, \mathbf{v}, \mathbf{L}). \quad (12)$$

For notational convenience, let \mathbf{q} denote a point with the same surface orientation in every reference image, i.e., \mathbf{q} is a pointer to a pixel of a given orientation in each reference image rather than absolute image coordinates.

If the reference observation vectors $\mathbf{V}_{\mathbf{q}}^{r_1}, \dots, \mathbf{V}_{\mathbf{q}}^{r_k}$ are linearly independent, they form a k -dimensional vector space and it follows that the target OV must lie within their span, i.e.,

$$\mathbf{V}_{\mathbf{p}}^t = \sum_{j=1}^k m_{j,\mathbf{p}} \mathbf{V}_{\mathbf{q}}^{r_j}. \quad (13)$$

We refer to the vector $\mathbf{m}_{\mathbf{p}} = [m_{1,\mathbf{p}}, \dots, m_{k,\mathbf{p}}]^T$ as the *material index* for target point \mathbf{p} .

By stacking the reference OVs into a matrix $\mathbf{W}_{\mathbf{q}} = [\mathbf{V}_{\mathbf{q}}^{r_1}, \dots, \mathbf{V}_{\mathbf{q}}^{r_k}]$, we can rewrite (13) as

$$\mathbf{V}_{\mathbf{p}}^t = \mathbf{W}_{\mathbf{q}} \mathbf{m}_{\mathbf{p}}. \quad (14)$$

For the case of RGB color images, there are separate material indices $\mathbf{m}_{\mathbf{R},\mathbf{p}}$, $\mathbf{m}_{\mathbf{G},\mathbf{p}}$, and $\mathbf{m}_{\mathbf{B},\mathbf{p}}$ for each of the R, G, and B channels. We refer to the *color material index* $\mathbf{m}_{\mathbf{p}}$ as the concatenation of these component indices into one vector.

3.3 Generalized Orientation-Consistency

The orientation-consistency cue is generalized to handle multiple materials as follows: We say a point \mathbf{p} on the target is orientation-consistent with a point \mathbf{q} in each of the reference images if there exists a material index $\mathbf{m}_{\mathbf{p}}$ that satisfies (14).

The procedure in Section 2 is modified as follows: For each point \mathbf{p} on the target, each point \mathbf{q} on the reference object is considered as a candidate match. The material index $\mathbf{m}_{\mathbf{p}}$ is computed for each candidate point \mathbf{q} by the pseudoinverse ($^+$) operation:

$$\mathbf{m}_{\mathbf{p}} = \mathbf{W}_{\mathbf{q}}^+ \mathbf{V}_{\mathbf{p}}^t. \quad (15)$$

The candidate \mathbf{q} is chosen for which the estimated $\mathbf{m}_{\mathbf{p}}$ minimizes

$$\|\mathbf{W}_{\mathbf{q}} \mathbf{m}_{\mathbf{p}} - \mathbf{V}_{\mathbf{p}}^t\|^2. \quad (16)$$

In the case of RGB color images, a separate $\mathbf{m}_{\mathbf{p}}$ is estimated for each color channel, and (16) is summed over the three channels.

Note that for some points on the surface, the material index is ambiguous, but *the surface normal is uniquely*



Fig. 6. Fish input data (one of 14 sets of images).

determined. For example, some points on the bottle in Fig. 2 do not appear in highlight in any of the input images. Although the surface is shiny, these points could also fit a purely Lambertian model. Consequently, if we were to attempt to reconstruct the bottle with two reference spheres, one diffuse, and one specular, the coefficient for the specular sphere would be ambiguous. However, the match on the diffuse sphere is uniquely determined and this uniquely specifies the surface normal. The use of the pseudoinverse in (15) resolves this ambiguity.

3.3.1 Results

To evaluate the method as applied to multiple materials, we imaged a 3-inch ceramic fish and a 6-inch ceramic cat under a range of illumination conditions. Fourteen images of the fish and thirteen images of the cat were used. Each image was captured with a different position of the hand-held light source. Both objects are shiny and contain multiple materials.

Light positions in front of the objects were used, in order to avoid cast shadows from the object onto itself. To model the different materials, we used two reference objects: a shiny sphere and a diffuse sphere. The reference spheres were imaged together with each target object, under the same illuminations. For the shiny reference sphere, we used a black snooker ball. For the diffuse reference sphere, we spray-painted a billiard ball with gray primer. Fig. 6 shows an image of the fish and the two reference objects. For comparison, Fig. 7 shows our reconstruction result along with the same fish scanned with a desktop CyberWare Model 15 laser scanner. In order to laser scan the object, it first had to be covered with a thin diffuse paint. Our method extracts distinctly more surface detail than the laser scanner. While some of this detail is likely due to the higher resolution of the D60 compared to the imager in the laser scanner, other differences may be due to the fact that the fish is coated with a thin layer of transparent lacquer. Our approach is likely capturing the surface under the lacquer³—which does contain a relief texture, while the laser scan of the painted fish captures the smoothed, lacquered, outer surface. Our reconstruction required roughly 5 hours of computation on a 2.8 GHz Intel Xeon.

3. Except at highlights, where we capture a mixture of the inner and outer surfaces.

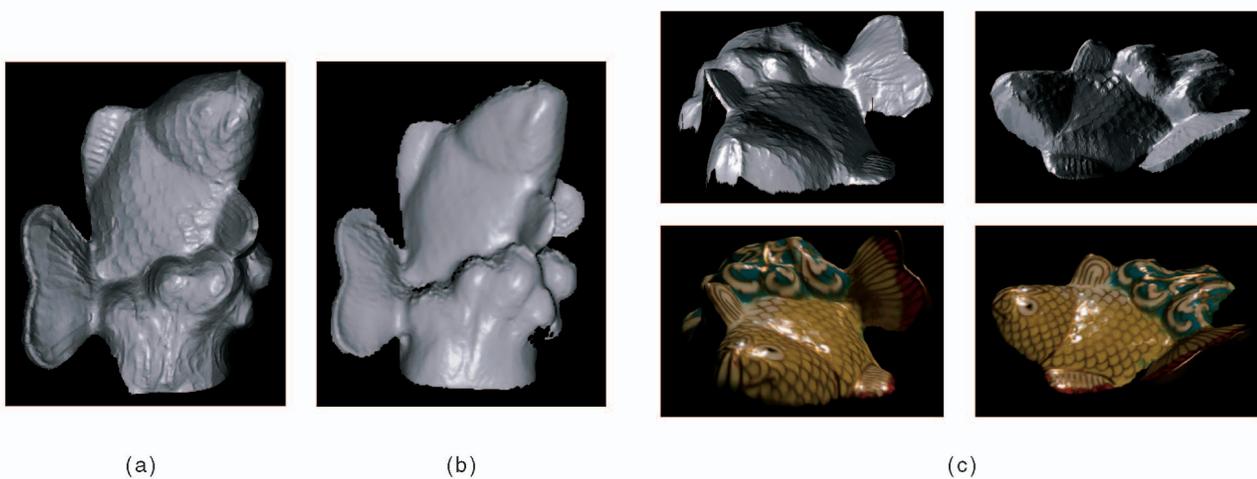


Fig. 7. Fish reconstruction from data in Fig. 6. (a) Reconstruction using our method. (b) Comparison with laser range scanner reconstruction from a similar viewpoint (the fish was painted with a diffuse paint in order to enable laser scanning). (c) Additional views of our reconstruction, using original image data for texture-mapping.

Fig. 8 shows one of the input images and our shape reconstruction of the cat model. Noticeable artifacts occur in our results where the assumptions of our algorithm do not hold. For example, cast shadows by the pink pendant on the cat result in indentations in the reconstructed surface. Also, some artifacts occur near where highlights appeared in the original images, since generalized orientation-consistency does not model saturation of the highlights. This saturation problem could potentially be resolved by using high-dynamic-range images.

4 CLUSTERING MATERIALS

We now describe a technique for segmenting surface materials. Most image segmentation techniques operate directly on image pixels, and are unable to distinguish shading variations from material (e.g., albedo) variations. Instead, we factor out shading variation and segment purely on albedo, in the spirit of classical work on physics-based vision [2], [13], [17]. Unlike previous work, however, we both exploit empirical reflectance models and operate on several images at once to achieve more robust results. We assume that the surface is composed of K distinct materials, with material c defined by a material index \mathbf{m}_c . We assume K is known a priori, although our approach could be extended to compute K automatically [10], [24]. We can then segment the surface materials using a technique similar to k -means and to the clustering algorithm of Lensch et al. [19].

Our goal is to estimate the material indices of the K materials, and to assign every target pixel \mathbf{p} to one of these materials. We also compute the surface orientations that best fit these estimates. Specifically, we define a labeling indicator variable $\lambda_{c,\mathbf{p}}$ that equals 1 if point \mathbf{p} is assigned to material c , and 0 otherwise. The labeling is mutually exclusive: $\sum_c \lambda_{c,\mathbf{p}} = 1$. Define $s(\mathbf{p})$ to be a point \mathbf{q} on the reference object with the same orientation as \mathbf{p} . The problem can then be stated as finding the labels λ , materials \mathbf{m} , and correspondences $s(\mathbf{p})$ to minimize

$$E(\lambda, \mathbf{m}, s) = \sum_{\mathbf{p}} \sum_c \lambda_{c,\mathbf{p}} \|\mathbf{V}_{\mathbf{p}}^t - \mathbf{W}_{s(\mathbf{p})} \mathbf{m}_c\|^2. \quad (17)$$

Our clustering algorithm works as follows. First, we run the algorithm described in Section 3.3 to obtain an initial estimate of $s(\mathbf{p})$. We initialize λ with a random labeling.⁴ We then optimize $E(\lambda, \mathbf{m}, s)$ by alternating between updates to \mathbf{m} and to λ and $s(\mathbf{p})$:

- The objective function is optimized with respect to the materials by setting

$$\mathbf{m}_c = \left(\sum_{\mathbf{p}} \lambda_{c,\mathbf{p}} \mathbf{W}_{s(\mathbf{p})}^T \mathbf{W}_{s(\mathbf{p})} \right)^+ \left(\sum_{\mathbf{p}} \lambda_{c,\mathbf{p}} \mathbf{W}_{s(\mathbf{p})}^T \mathbf{V}_{\mathbf{p}}^t \right) \quad (18)$$

for each c . Based on these new material indices, we can construct a *virtual reference object* (VRO) for each material by linearly combining the real reference objects: the OV at point \mathbf{q} of the VRO for material c is given by $\mathbf{W}_{\mathbf{q}} \mathbf{m}_c$. (The pseudoinverse is used in case of ambiguities, e.g., a cluster containing normals that are never in highlight on the calibration objects.)

- The labeling λ and correspondence $s(\mathbf{p})$ is computed for each target pixel \mathbf{p} by finding the material c and $s(\mathbf{p})$ for which $\|\mathbf{V}_{\mathbf{p}}^t - \mathbf{W}_{s(\mathbf{p})} \mathbf{m}_c\|$ is minimized. In practice, we compute all OVs from the VROs and place them in an ANN data structure [1], and search for the nearest neighbor to $\mathbf{V}_{\mathbf{p}}^t$ for each \mathbf{p} .

The objective function is guaranteed to be nonincreasing at each step and, since there is a finite set of possible labelings λ , the algorithm is guaranteed to converge to at least a local minimum. The extension to RGB color images is straightforward: material indices are computed separately for each color; the labelings are updated by searching for the color OV that best matches the target OV.

4.1 Results

Fig. 9 illustrates the clustering algorithm applied to the fish and cat images. Note that the algorithm correctly groups

4. We do not directly use the materials produced by the algorithm described in the Section 3.3, due to the ambiguity discussed in that section. Directly clustering the estimated materials produces poor results.



Fig. 8. Cat reconstruction. (a) Input data (one of 13 sets). (b) Views of the reconstructed model.

pixels with similar materials but different orientations (and, thus, different reflectances). For example, the white areas of the cat have been grouped in cluster 4, despite variations in shading due to illumination. Also, note that the algorithm has used cluster 1 on the cat as an “outlier material,” to model cast shadows and interreflections. Because there are not enough clusters to capture the smooth material variations on the surface, the clustering of the cat appears fragmented, but these fragments correspond well to the different albedos. For the same reason, the clustering algorithm gave slightly less accurate surface normals than the algorithm in the previous section. Because the fish object contained a discrete set of materials, the segmentation for this example is very clean.

5 DISCUSSION

Many fundamental questions and possible extensions remain.

5.1 How Many Images Are Needed?

The answer will surely depend on the properties of the material and the environment. It is known, for instance, that three nondegenerate images are sufficient in the case of Lambertian scenes [33]. For purely specular scenes, the answer depends on the illumination, particularly in the case of a perfect mirror. In one extreme, a chrome object illuminated by a distant, moving point light source would require an *infinite* number of images to reconstruct the surface, enough for the highlight to pass over every point on the object. In contrast, if the chrome object is illuminated by a highly structured lighting environment, a *single* image could be sufficient to compute the correspondence of pixels to the reference object, i.e., if every incoming light ray has a unique intensity or color. A more practical scenario might be to use a small number of broad area light sources, generated, for instance, by light reflected off of an indoor or outdoor environment. In the case of an outdoor environment, the images could be captured at different times of day as the sun moves across the sky.

While chrome objects pose unique challenges, most realistic materials are not purely specular, but also exhibit a significant diffuse (or other low-frequency reflection) component. For example, the bottle in Section 2 is highly specular, yet may be reconstructed from only three images with a point light source due to the presence of the diffuse

component. In the spirit of [4], [25], we conjecture that many images are required only when the BRDF and the illumination lack low-frequency components, e.g., a mirror surface under point lighting. An interesting topic for future work is to determine the requisite number of images under general situations, perhaps using the frequency-domain analysis of Basri and Jacobs [4] and Ramamoorthi and Hanrahan [25].

5.2 Reference Objects

An important problem is how to choose a suitable set of reference objects to best fit a given target object or class of objects. For example, modeling surfaces where the polish and gloss vary continuously over the surface may require several reference spheres. Increasing the number of reference spheres would likely necessitate using more input images. Additional constraints may be needed to prevent invalid linear combinations (e.g., a material corresponding to a broad highlight *minus* a sharp highlight).

More generally, is there a “universal” set of reference objects that may be used for a wide range of target materials? Studies based on empirical BRDF measurements may help determine what objects would comprise such a set, in a similar vein to the work of Matusik et al. [20], [21].

5.3 Alternative Imaging Methods

Although the techniques in this paper allow area light sources, we have experimented only with point sources. Area light sources could provide more robust results with fewer images, particularly in the case of highly specular surfaces. However, one problem with using area sources is that cast shadows will be more difficult to avoid. For anisotropic materials, there remains a question of how to obtain sufficient samples of the reference material to model the full 3D space of surface orientations. One possibility is to image a cylindrical sample rotated to multiple orientations.

While we employed a fixed camera viewpoint, orientation-consistency also holds under changes in viewpoint; in [29], orientation-consistency is applied in the context of multiview stereo, in which both the illumination *and* the camera viewpoint may vary. Treuille et al. [29] also address the problem of cast shadows in the multiview case.

5.4 More General Surfaces

This paper makes several assumptions that one might want to relax. In particular, we assume orthography, distant

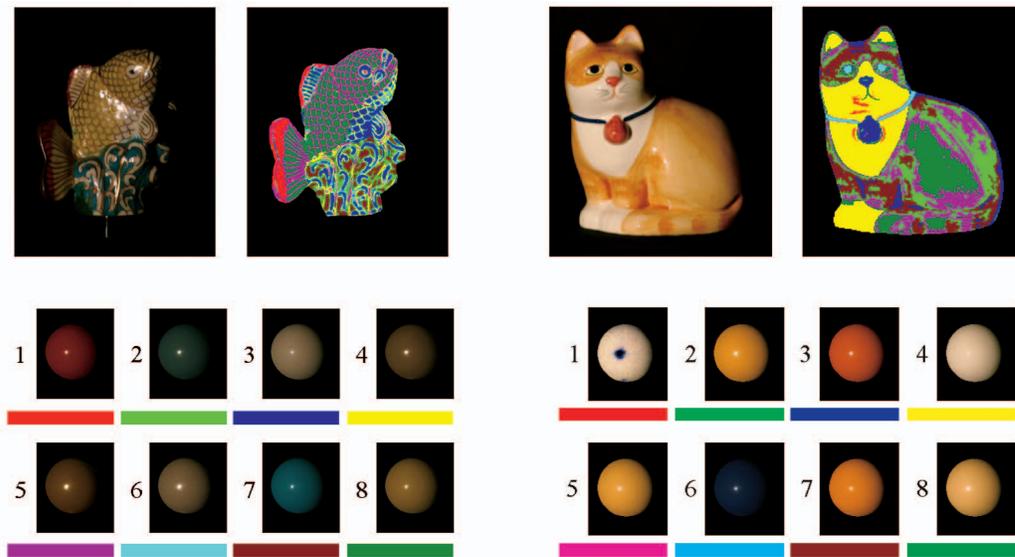


Fig. 9. Material clustering and virtual reference objects of the fish model. The pseudocolored images at top show the computed segmentation of the object into different materials. At the bottom, the recovered materials, i.e., different linear combinations of the reference spheres, are shown with the associated pseudocolor assignments.

lighting, no cast shadows, no interreflections, no subsurface scattering, and no transparency. For some of these assumptions, it may be possible to detect and mask out pixels that do not satisfy these assumptions, for example, by measuring reconstruction error for individual pixels. It may also be possible to model some of these effects by generalizing orientation-consistency. For example, our method should correctly handle interreflections when similar interreflections are observed in the reference object, for instance, to model chisel marks on a statue. Another interesting avenue for future research is to remove the need for a physical reference object by rendering a synthetic one, given a known lighting environment and a known BRDF model (e.g., [8]).

ACKNOWLEDGMENTS

An earlier version of this paper appeared in [14]. Portions of this work were conducted while Aaron Hertzmann was at the University of Washington. The authors would like to thank Loren Merritt for his help in preparing Fig. 3. Dan Goldman, Jiwon Kim, Li Zhang, and Daniel Wood also provided helpful suggestions and assistance. This work was supported in part by US National Science Foundation grants IIS-0049095 and IIS-0113007, an ONR YIP award, the U.W. Animation Research Labs, and a grant from Microsoft Corporation.

REFERENCES

- [1] S. Arya, D.M. Mount, N.S. Netanyahu, R. Silverman, and A.Y. Wu, "An Optimal Algorithm for Approximate Nearest Neighbor Searching in Fixed Dimensions," *J. ACM*, vol. 45, no. 6, pp. 891-923, 1998.
- [2] R. Bajcsy, S. Lee, and A. Leonardis, "Color Image Segmentation with Detection of Highlights and Local Illumination Induced by Inter-Reflections," *Proc. Int'l Conf. Pattern Recognition*, vol. 1, pp. 785-790, 1990.
- [3] R. Basri and D. Jacobs, "Photometric Stereo with General, Unknown Lighting," *Proc. Computer Vision and Pattern Recognition Conf.*, pp. 374-381, 2001.
- [4] R. Basri and D.W. Jacobs, "Lambertian Reflectance and Linear Subspaces," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 25, no. 2, pp. 218-233, Feb. 2003.
- [5] J.S. Beis and D.G. Lowe, "Indexing without Invariants in 3D Object Recognition," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 21, no. 10, pp. 1000-1015, Oct. 1999.
- [6] W.-C. Chen, J.-Y. Bouguet, M.H. Chu, and R. Grzeszczuk, "Light Field Mapping: Efficient Representation and Hardware Rendering of Surface Light Fields," *Proc. SIGGRAPH*, pp. 447-456, 2002.
- [7] E. Coleman Jr. and R. Jain, "Obtaining 3-Dimensional Shape of Textured and Specular Surfaces Using Four-Source Photometry," *Proc. Int'l Conf. Color in Graphics and Image Processing (CGIP)*, vol. 18, no. 4, pp. 309-328, Apr. 1982.
- [8] K.J. Dana, B. van Ginneken, S.K. Nayar, and J.J. Koenderink, "Reflectance and Texture of Real-World Surfaces," *ACM Trans. Graphics*, vol. 18, no. 1, pp. 1-34, 1999.
- [9] P.E. Debevec and J. Malik, "Recovering High Dynamic Range Radiance Maps from Photographs," *Proc. SIGGRAPH '97*, pp. 369-378, 1997.
- [10] C. Fraley and A.E. Raftery, "How Many Clusters? Which Clustering Method? Answers via Model-Based Cluster Analysis," *The Computer J.*, vol. 41, no. 8, pp. 578-588, 1998.
- [11] J.H. Friedman, J.L. Bentley, and R.A. Finkel, "An Algorithm for Finding Best Matches in Logarithmic Expected Time," *ACM Trans. Math. Software*, vol. 3, no. 3, pp. 209-226, 1977.
- [12] A.S. Georghiadis, "Recovering 3-D Shape and Reflectance from a Small Number of Photographs," *Proc. Eurographics Symp. Rendering: 14th Eurographics Workshop Rendering*, pp. 230-240, June 2003.
- [13] G. Healey, "Segmenting Images Using Normalized Color," *IEEE Trans. Systems, Man, and Cybernetics*, vol. 22, no. 1, pp. 64-73, 1992.
- [14] A. Hertzmann and S.M. Seitz, "Shape and Materials by Example: A Photometric Stereo Approach," *Proc. Conf. Computer Vision and Pattern Recognition*, vol. 1, pp. 533-540, June 2003.
- [15] B.K.P. Horn, *Robot Vision*. New York: McGraw-Hill, 1986.
- [16] K. Ikeuchi, "Determining Surface Orientations of Specular Surfaces by Using the Photometric Stereo Method," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 3, no. 6, pp. 661-669, 1981.
- [17] G.J. Klunker, S.A. Shafer, and T. Kanade, "A Physical Approach to Color Image Understanding," *Int'l J. Computer Vision*, vol. 4, no. 1, pp. 7-38, 1990.
- [18] D.J. Kriegman, T. Zickler, J. Ho, J. Ponce, and P.N. Belhumeur, "Binocular Helmholtz Stereopsis," *Proc. European Conf. Computer Vision*, pp. 1411-1417, Oct. 2003.

- [19] H.P. 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, Apr. 2003.
- [20] 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.
- [21] W. Matusik, H. Pfister, M. Brand, and L. McMillan, "Efficient Isotropic BRDF Measurement," *Proc. Eurographics Symp. Rendering: 14th Eurographics Workshop Rendering*, pp. 241-248, June 2003.
- [22] 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, 1990.
- [23] S.A. Nene and S.K. Nayar, "A Simple Algorithm for Nearest Neighbor Search in High Dimensions," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 19, no. 9, pp. 989-1003, Sept. 1997.
- [24] D. Pelleg and A.W. Moore, "X-Means: Extending K-Means with Efficient Estimation of the Number of Clusters," *Proc. Int'l Conf. Machine Learning*, pp. 727-734, 2000.
- [25] R. Ramamoorthi and P. Hanrahan, "Frequency Space Environment Map Rendering," *ACM Trans. Graphics*, vol. 21, no. 3, pp. 517-526, July 2002.
- [26] W.M. Silver, "Determining Shape and Reflectance Using Multiple Images," Master's thesis, MIT, Cambridge, Mass., 1980.
- [27] A.R. Smith and J.F. Blinn, "Blue Screen Matting," *Proc. SIGGRAPH '96*, pp. 259-268, Aug. 1996.
- [28] D. Terzopoulos, "The Computation of Visible-Surface Representations," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 10, no. 4, pp. 417-438, 1988.
- [29] A. Treuille, A. Hertzmann, and S.M. Seitz, "Example-Based Stereo with General BRDFs," *Proc. European Conf. Computer Vision*, pp. 457-469, 2004.
- [30] P. Tu and P.R.S. Mendonça, "Surface Reconstruction via Helmholtz Reciprocity with a Single Image Pair," *Proc. Conf. Computer Vision and Pattern Recognition*, vol. 1, pp. 541-547, June 2003.
- [31] *Physics-Based Vision: Principles and Practice, Shape Recovery*, L.B. Wolff, S.A. Shafer, and G.E. Healey, eds., Boston: Jones and Bartlett, 1992.
- [32] D.N. Wood, D.I. Azuma, W. Aldinger, B. Curless, T. Duchamp, D.H. Salesin, and W. Stuetzle, "Surface Light Fields for 3D Photography," *Proc. SIGGRAPH*, pp. 287-296, 2000.
- [33] R.J. Woodham, "Photometric Method for Determining Surface Orientation from Multiple Images," *Optical Eng.*, vol. 19, no. 1, pp. 139-144, 1980.
- [34] R.J. Woodham, "Gradient and Curvature from Photometric Stereo Including Local Confidence Estimation," *J. Optical Soc. Am. A*, vol. 11, no. 11, pp. 3050-3068, Nov. 1994.
- [35] 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.
- [36] T. Zickler, P.N. Belhumeur, and D.J. Kriegman, "Toward a Stratification of Helmholtz Stereopsis," *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, vol. 1, pp. 548-555, June 2003.
- [37] R.J. Woodham, "Reflectance Map Techniques for Analyzing Surface Defects in Metal Castings," Technical Report AI-TR-457, MIT, Artificial Intelligence Laboratory, Cambridge, Mass., 1978.
- [38] R. Lu, J.J. Koenderink, and A.M.L. Kappers, "Optical Properties (Bidirectional Reflection Distribution Functions) of Velvet," *Applied Optics-OT*, vol. 37, no. 25, pp. 5974-5984, 1998.



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 assistant professor of computer science at the University of Toronto. In the past, he has worked at the University of Washington, Microsoft Research, Mitsubishi Electric Research Lab, Interval Research Corporation, and at NEC Research Institute. In 2004, he cochaired the NPAR Conference on Nonphotorealistic Rendering, served on the jury for the Anney International Animated Film Festival, and was named to the TR100, a list of the top 100 innovators under 35 worldwide. His research interests include computer vision, computer animation, and machine learning.



Steven M. Seitz received the BA degree in computer science and mathematics at the University of California, Berkeley in 1991 and the PhD degree in computer sciences at the University of Wisconsin, Madison in 1997. He is an associate 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, an ONR Young Investigator Award, and an Alfred P. Sloan Fellowship. He is interested in problems in computer vision and computer graphics. His current research focuses on capturing the structure, appearance, and behavior of the real world from digital imagery. He is a 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.**