

Shape and Materials by Example: A Photometric Stereo Approach

Aaron Hertzmann

University of Toronto, Toronto, Canada

Steven M. Seitz

University of Washington, Seattle, WA

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.

1 Introduction

An important unsolved problem in computer vision is modeling scenes with realistic materials. Although real objects reflect light in a wide range of interesting ways [3], most shape reconstruction techniques in the computer vision literature work only for Lambertian scenes. 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 normals of an object, when imaged together with one or more reference objects of similar materials and known shape.

This approach has the following features:

- The BRDF, illumination, and shape may *all* be unknown. Moreover, the BRDF may be arbitrary and can

vary 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. Second, the camera is assumed to be orthographic. Finally, shadows, interreflections, and other non-local radiance effects are ignored.

1.1 Relation to Previous Work

Our use of reference objects builds upon early work on photometric stereo. In particular, Silver [10] used images of a wooden sphere to reconstruct other wooden objects imaged under the same lighting. By specifying the outgoing radiance function for a hemisphere of directions, the sphere images provide an empirical model of reflection for wood.

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 albedo, (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 [2, 5, 8, 10, 14] (for a good overview, see [12]). While some of these techniques have been shown to yield good results, they are currently restricted to simple diffuse and specular materials. It is

This work was performed while Aaron Hertzmann was at the University of Washington. The authors' email addresses are hertzman@dgp.toronto.edu and seitz@cs.washington.edu.



Figure 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, all points that are in highlight have the same surface normal.

known, however, that a wide range of real world materials do not fit these traditional BRDF models [3].

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 [7, 15] which also enables reconstructing surfaces with arbitrary BRDF’s in the absence of shadows and interreflections. The two approaches have different requirements, however. An advantage of the Helmholtz approach is that it does not require reference objects, orthography, or distant illumination. 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) [15]. In contrast, our approach does not require that the viewpoint or illumination are known, and is flexible enough to be used in unconstrained environments, e.g., for outside imaging using the sun and environment as a light source.

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 non-local 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 from the principles of radiometry [4] 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 Figure 1.

If we happen to *know* the normals for some points in the scene, orientation-consistency allows us to propagate this knowledge to other points. In the example shown in Figure 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.

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 [10]; 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 run-time.

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:

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

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

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

Given a pixel \mathbf{p} on the target object, the normal at \mathbf{p} is determined simply by searching for the point \mathbf{q} on the reference object with the best matching OV, i.e., \mathbf{q} minimizes $\|\mathbf{V}_p - \mathbf{V}_q\|$. 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 [14], 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 [2, 5, 8, 14] can provide a more efficient alternative to the explicit search method presented here.

Implementation Issues. In order to determine the geometry of the reference sphere, we manually segmented 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. 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; we use the ANN technique [1] for this purpose. To reduce the dimensionality of the search, we apply principle component analysis to the reference OVs before building the ANN search structure.

We compute a 3D surface from the estimated surface orientations by formulating a quadratic energy function incorporating both surface orientation fit at each pixel and a smoothness term [11]. This results in a sparse system of linear equations that we solve using conjugate gradient.

Note that the calibration requirements are extremely minimal—other than correcting the images for radial distortion or vignetting, 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

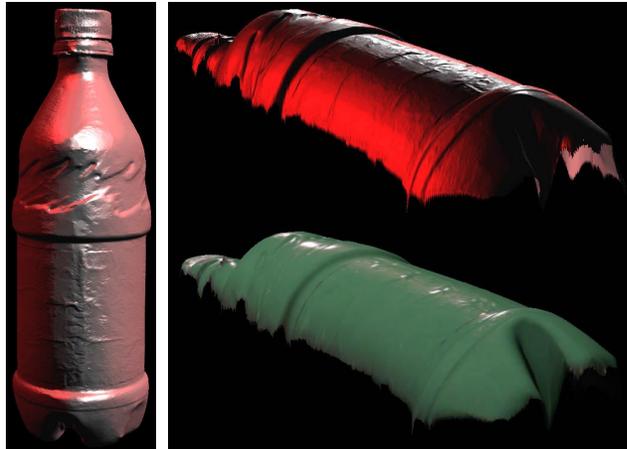


Figure 2: Bottle reconstruction. Shaded mesh renderings of the result are shown at left and top. A texture-mapped version is below. Note the fine details such as the wrinkles in the label that are accurately captured.

light shone at the objects from a distance of 10 to 15 feet. The objects were all less than one foot high.

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. Two of the eight input images are shown in Figure 1; the resulting reconstruction is shown in Figure 2. 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. The images were quite large— 398×1176 for the bottle and 328×332 for the sphere. A brute force search to compute normals requires roughly one day of compute time. The accelerated method using ANN completed in about 5 minutes. 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 [11].

The same algorithm is applied to reconstructing a velvet surface in Figure 3. 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 when the surface normal is nearly orthogonal to the light source direction [3]. 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

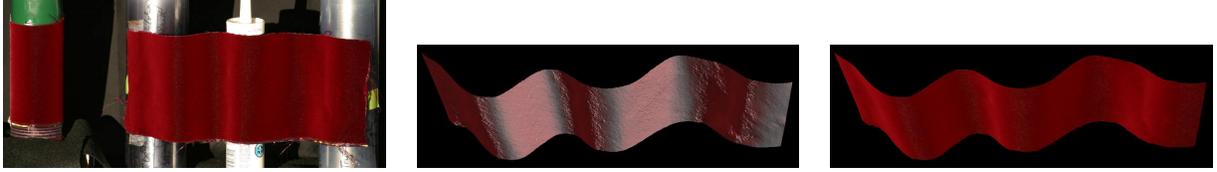


Figure 3: Velvet reconstruction. *Left*: One photograph (of fourteen) used for reconstruction. The reference object, in the shape of a cylinder is on the left, and the target object is on the right. *Center*: Reconstructed target shape, rendered from above. *Right*: Reconstructed target shape, textured with the target image

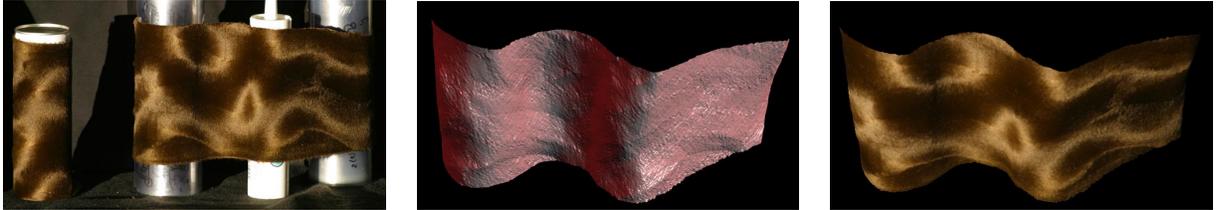


Figure 4: Brushed fur reconstruction. This anisotropic material is made of a single BRDF that is rotated due to fiber orientation, causing an interesting pattern of highlights. *Left*: One photograph (of fourteen) used for reconstruction. The reference object, in the shape of a cylinder is on the left, and the target object is on the right. *Center*: Reconstructed target shape, rendered from above. *Right*: Reconstructed target shape, textured with the target image

very well, as seen in the figure—note the three protrusions from the three supporting cylinders.

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.

Figure 4 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. Note that the fabric was placed in a different configuration than in the velvet example.

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. 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 rotating the cylinder).

3 Modeling Material Variation

Two limitations of the technique presented in the previous section 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 first subsection, we consider the case of surface texture due to albedo variations over a purely diffuse or specular object. The next subsection 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:

$$I_{\mathbf{p}}^t = \rho_{\mathbf{p}}^t \mathbf{n}_{\mathbf{p}} \cdot \sum_{\text{light } \mathbf{l}} \mathbf{l} \quad (3)$$

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

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

$$I_{\mathbf{p}}^r = \rho^r \mathbf{n}_{\mathbf{p}} \cdot \sum_{\text{light } \mathbf{l}} \mathbf{l} \quad (4)$$

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 equations (3) and (4) that

$$\mathbf{V}_{\mathbf{p}}^t = \frac{\rho_{\mathbf{p}}^t}{\rho_{\mathbf{q}}^r} \mathbf{V}_{\mathbf{q}}^r \quad (5)$$

and therefore

$$\frac{\mathbf{V}_{\mathbf{p}}^t}{\|\mathbf{V}_{\mathbf{p}}^t\|} = \frac{\mathbf{V}_{\mathbf{q}}^r}{\|\mathbf{V}_{\mathbf{q}}^r\|} \quad (6)$$

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

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

$$I_{\mathbf{p}} = \rho_{\mathbf{p}} f(\mathbf{n}_{\mathbf{p}}, \mathbf{v}, \mathbf{L}) \quad (7)$$

where f is any reflectance map 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 [6, 9, 13].

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

$$I_{\mathbf{p}}^t = \sum_{i=0}^k \rho_{i,\mathbf{p}}^t f_i(\mathbf{n}_{\mathbf{p}}, \mathbf{v}, \mathbf{L}) \quad (8)$$

Note that here we assume that the *image intensities*, not the BRDF's are a linear combination of k basis functions. The distinction is significant if the camera has a nonlinear gain, i.e., if the pixel values are not linear functions of scene radiance. In some cases, it may be beneficial to calibrate and correct for such nonlinearities, 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

¹In 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.

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.,

$$I_{\mathbf{q}}^{r_j} = \sum_{i=0}^k \rho_i^{r_j} f_i(\mathbf{n}_{\mathbf{q}}, \mathbf{v}, \mathbf{L}) \quad (9)$$

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.,

$$\mathbf{V}_{\mathbf{p}}^t = \sum_{j=0}^k m_{j,\mathbf{p}} \mathbf{V}_{\mathbf{q}}^{r_j} \quad (10)$$

We refer to the vector $\mathbf{m}_{\mathbf{p}} = [m_{1,\mathbf{p}}, \dots, m_{k,\mathbf{p}}]^T$ as the real-valued *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 restate equation (10) as

$$\mathbf{V}_{\mathbf{p}}^t = \mathbf{W}_{\mathbf{q}} \mathbf{m}_{\mathbf{p}} \quad (11)$$

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 equation (11).

The procedure in Section 2 is modified to compute the material index $\mathbf{m}_{\mathbf{p}}$ for each candidate point \mathbf{q} by the pseudoinverse (+) operation:

$$\mathbf{m}_{\mathbf{p}} = \mathbf{W}_{\mathbf{q}}^+ \mathbf{V}_{\mathbf{p}}^t \quad (12)$$

For each point \mathbf{p} on the target, the candidate \mathbf{q} is chosen such that the estimated $\mathbf{m}_{\mathbf{p}}$ minimizes

$$\|\mathbf{W}_{\mathbf{q}} \mathbf{m}_{\mathbf{p}} - \mathbf{V}_{\mathbf{p}}^t\|^2 \quad (13)$$

In the case of RGB color images, a separate $\mathbf{m}_{\mathbf{p}}$ is estimated for each color channel, and equation (13) is summed over the three channels. The pseudoinverse is used in order to handle pixels whose normal is uniquely determined but whose material index is ambiguous. For example, some



Figure 5: Fish input data (one of fourteen sets of images).

points on the bottle in Figure 1 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. However, this ambiguity does not prevent estimating their surface normals, which are fully determined from the Lambertian component. The ambiguity in material estimation can be addressed by the material clustering algorithm in the next section.

Results. Figures 6 and 7 show reconstruction results of a 3-inch tall ceramic fish and a 6-inch tall ceramic cat. Both objects are shiny and contain multiple materials. We chose light source positions to minimize cast self-shadows. For the shiny reference object, we used a black snooker ball. For the diffuse reference object, we spray-painted a billiard ball with gray primer. For comparison, Figure 6 shows our 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. Nonetheless, 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 lacquer—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.

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 appear near where highlights appeared in the original images, since generalized consistency does not model saturation of the highlights.

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. We assume that the surface is composed of K distinct materials, with material c defined by a material index \mathbf{m}_c . We can then cluster the surface using clustering algorithm similar to k -means and to the clustering algorithm of Lensch et al. [6].

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

$$E(\lambda, \mathbf{m}, s) = \sum_{\mathbf{p}} \sum_c \lambda_{c,\mathbf{p}} \|\mathbf{V}_{\mathbf{p}}^t - \mathbf{W}_{s(\mathbf{p})} \mathbf{m}_c\|^2 \quad (14)$$

Our clustering algorithm is as follows. First, we run the algorithm described in the previous section² to obtain an initial estimate of $s(\mathbf{p})$. We initialize λ with a random labeling. 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 $\mathbf{m}_c = (\sum_{\mathbf{p}} \lambda_{c,\mathbf{p}} \mathbf{W}_{s(\mathbf{p})}^T \mathbf{W}_{s(\mathbf{p})})^+ (\sum_{\mathbf{p}} \lambda_{c,\mathbf{p}} \mathbf{W}_{s(\mathbf{p})}^T \mathbf{V}_{\mathbf{p}}^t)$ 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 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. 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.

²We do not directly use the materials produced by the algorithm described in the previous section, due to the ambiguity discussed in Section 3.3. Directly clustering the estimated materials produces poor results.

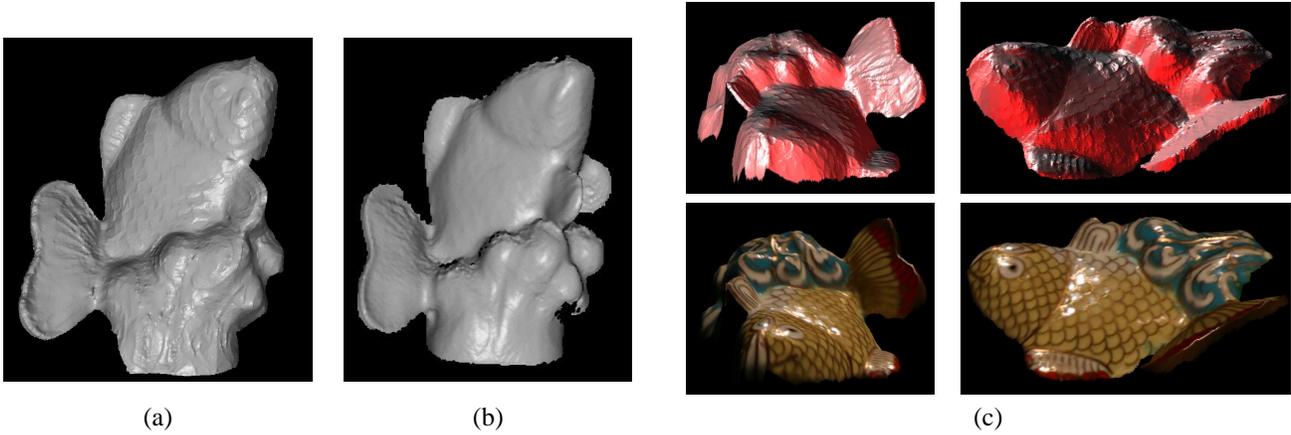


Figure 6: Fish reconstruction from data in Figure 5. (a) Reconstruction using our method. (b) Comparison with laser range scanner reconstruction from a similar viewpoint (the fish was painted with a diffuse paint in order to enable laser scanning). (c) Additional views of our reconstruction, using original image data for texture-mapping.



Figure 7: Cat reconstruction. *Left*: Input data (one of thirteen sets). *Right*: Views of the reconstructed model.

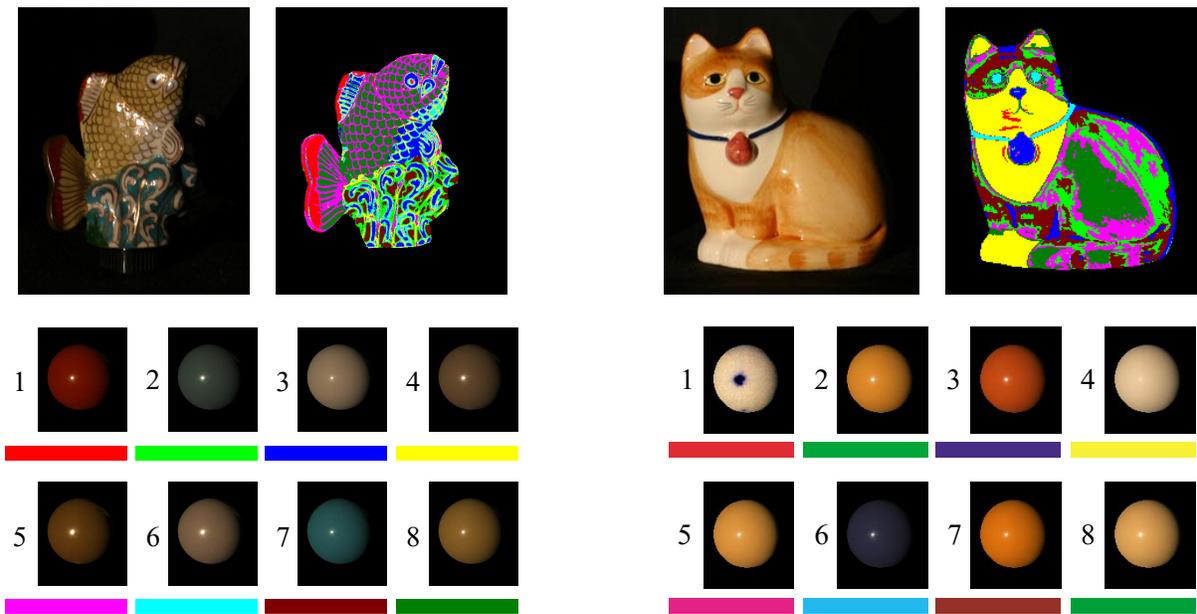


Figure 8: Material clustering and virtual reference objects of the fish and cat models. (Refer to the electronic version of this paper to view the images in color.)

Results. Figure 8 illustrates the clustering algorithm applied to the two surfaces of the previous section. Note that the algorithm correctly groups 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 remain. In particular, how many images are needed to reconstruct scenes using this example-based method? The answer will surely depend on the properties of the material and the environment. It is known, for instance, that three non-degenerate images are sufficient in the case of Lambertian scenes [14]. 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. Investigating the necessary illumination conditions (and designing reference objects) for anisotropic materials is another interesting area for future work.

The method makes several assumptions that one might want to relax. In particular, we assume orthography, distant lighting, no cast shadows, no interreflections, no sub-surface scattering, or 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 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. [3]).

Acknowledgements

We would like to thank Jiwon Kim, Li Zhang, and Daniel Wood for helpful suggestions and assistance. This work was supported in part by NSF 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. *Journal of the ACM*, 45(6):891–923, 1998.
- [2] R. Basri and D. Jacobs. Photometric stereo with general, unknown lighting. In *Proc. CVPR*, pages 374–381, 2001.
- [3] K. J. Dana, S. K. Nayar, B. van Ginneken, and J. J. Koenderink. Reflectance and texture of real-world surfaces. In *Proc. CVPR.*, pages 151–157, 1997.
- [4] B. K. P. Horn. *Robot Vision*. McGraw-Hill, New York, NY, 1986.
- [5] K. Ikeuchi. Determining surface orientations of specular surfaces by using the photometric stereo method. *IEEE PAMI*, 3(6):661–669, 1981.
- [6] H. P. Lensch, J. Kautz, M. Goesele, W. Heidrich, and H.-P. Seidel. Image-Based Reconstruction of Spatial Appearance and Geometric Detail. *Trans. on Graphics*, 22(2), Apr. 2003.
- [7] S. Magda, T. Zickler, D. Kriegman, and P. Belhumeur. Beyond Lambert: Reconstructing Surfaces with Arbitrary BRDFs. In *Proc. ICCV 2001*, pages 391–398, 2001.
- [8] S. K. Nayar, K. Ikeuchi, and T. Kanade. Determining shape and reflectance of hybrid surfaces by photometric sampling. *IEEE Trans. on Robotics and Automation*, 6(4):418–431, 1990.
- [9] R. Ramamoorthi and P. Hanrahan. Frequency Space Environment Map Rendering. *Trans. on Graphics*, 21(3):517–526, July 2002. (Proceedings of SIGGRAPH 2002).
- [10] W. M. Silver. Determining shape and reflectance using multiple images. Master’s thesis, MIT, Cambridge, MA, 1980.
- [11] D. Terzopoulos. The computation of visible-surface representations. *PAMI*, 10(4):417–438, 1988.
- [12] L. B. Wolff, S. A. Shafer, and G. E. Healey, editors. *Physics-based vision: Principles and Practice, Shape Recovery*. Jones and Bartlett, Boston, MA, 1992.
- [13] D. N. Wood, D. I. Azuma, W. Aldinger, B. Curless, T. Duchamp, D. H. Salesin, and W. Stuetzle. Surface Light Fields for 3D Photography. In *Proc. SIGGRAPH*, pages 287–296, 2000.
- [14] R. J. Woodham. Photometric method for determining surface orientation from multiple images. *Optical Engineering*, 19(1):139–144, 1980.
- [15] T. Zickler, P. N. Belhumeur, and D. J. Kriegman. Helmholtz Stereopsis: Exploiting Reciprocity for Surface Reconstruction. In *Proc. of ECCV*, volume 3, pages 869–884, 2002.