International Journal of Computer Vision 73(2), 123–138, 2007

DOI: 10.1007/s11263-006-9373-8

Shape and View Independent Reflectance Map from Multiple Views*

TIANLI YU, NING XU AND NARENDRA AHUJA
University of Illinois at Urbana-Champaign
tianli.yu@motorola.com

Received October 4, 2005; Revised May 24, 2006; Accepted June 2, 2006

First online version published in July 2006

Abstract. We consider the problem of estimating the 3D shape and reflectance properties of an object made of a single material from a set of calibrated views. To model the reflectance, we propose to use the View Independent Reflectance Map (VIRM), which is a representation of the joint effect of the diffuse+specular Bidirectional Reflectance Distribution Function (BRDF) and the environment illumination. The object shape is parameterized using a triangular mesh. We pose the estimation problem as minimizing the cost of matching input images, and the images synthesized using the shape and VIRM estimates. We show that by enforcing a constant value of VIRM as a global constraint, we can minimize the cost function by iterating between the VIRM and shape estimation. Experimental results on both synthetic and real objects show that our algorithm can recover both the 3D shape and the diffuse/specular reflectance information. Our algorithm does not require the light sources to be known or calibrated. The estimated VIRM can be used to predict the appearances of objects with the same material from novel viewpoints and under transformed illumination.

Keywords: reflectance model, 3d reconstruction, shape from shading, illumination model, BRDF

1. Introduction

This paper is concerned with recovering the geometric and photometric information from images of a 3D object acquired from multiple viewpoints. Many different methods have been proposed in the literature to solve this problem. Stereo methods, which involve a small number of views, usually employ a single depth image from the viewpoint of one of the cameras to represent the geometric information. While more recently, full 3D models such as deformable meshes (Samaras and Metaxas, 2003), volumetric (Kutulakos and Seitz, 2000) or implicit representations (Faugeras and Keriven, 1998) have been employed to incorporate more complicated visibility relations from a large number of viewpoints. These representations help to meaningfully combine the information present in multiple views. In many of these algorithms, objects are assumed to have Lambertian reflectance, i.e. their appearances do not change in different viewing directions. A simple consistency function can be defined whose maximization yields a 3D model that is most consistent with all the input views. However, many real world scenes do not obey the lambertian law and specular reflections in such cases cause severe problems for shape reconstruction. In addition, the non-lambertian reflectance of the surface needs to be properly modeled in order to fully reconstruct the object's appearance in different views.

To be robust to objects with non-lambertian reflectance, some methods remove the image areas with specularities as outliers during the reconstruction (Lin et al., 2002; Bhat and Nayar, 1998). This may be acceptable only when the specularities are sparse. Zickler

*The support of National Science Foundation under grant ECS 02-25523 is gratefully acknowledged. Tianli Yu was supported in part by a Beckman Institute Graduate Fellowship.

et al. (2002) use the Helmholtz reciprocal property for stereo matching, which holds even in specular areas. This requires a careful control of the illuminant and camera positions. Recently, Jin et al. (2003) propose a rank constraint on radiance tensor to handle more general non-lambertian reflectance. This constraint is incorporated in an energy minimization framework to solve for the 3D shape. Their method does not require the control of illumination and does not assume the sparseness of specularity. However, it requires every scene point to be visible in a substantial number of cameras. In addition, the estimates obtained by these image correspondence based algorithms are confined to individual pixels and cannot recover fine details of the shape, e.g., those encoded by shading.

Through explicit modeling of the photometric variation of the object surface, more scene details can be recovered from the captured images. Given the object shape, surface reflectance properties as well as the illumination can be estimated (Yu et al., 1999; Ramamoorthi and Hanrahan, 2001). A more difficult problem, however, is to obtain the object shape. Shape from shading or photometric stereo methods attempt to recover surface details by imposing constraints directly on the surface normal. However, shape from shading algorithms are usually developed for constrained environments, such as those involving single material objects, lambertian reflectance, a single viewpoint, known or very simple light source, orthographic projection, and absence of shadows or inter-reflections. Zhang et al. (1999) present a good survey of different shape from shading methods. Fua and Leclerc (1995) and Samaras et al. (2000) incorporate shape from shading method into multiple-view reconstruction. They consider lambertian objects and recover the piecewise constant albedo map as well as the surface shape. Jin et al. (2004) also propose a more complex illumination model, and use a level set framework to recover the surface shape of a lambertian object from the multi-view shading information while simultaneously estimating the illumination. For lambertian objects, complex lighting can be well modeled locally using a single point light source. This, however, is not the case for specular objects.

By controlling illumination, photometric stereo can recover object shape and surface reflectance while avoiding the difficulty of finding correspondences between images. Lee and Kuo (1997) use the Torrance-Sparrow BRDF model and a triangular mesh representation of the surface shape to recover both the BRDF and surface shape from images taken under known illumination settings. Georghiades (2003) also uses Torrance-Sparrow BRDF model and extracts the shape and spatially varying BRDF using uncalibrated photometric stereo. Hertzmann and Seitz (2003) and Treuille et al. (2004) use several example spheres imaged together with the object to recover the object shape. Each example sphere serves to extract the reflectance map for a specific material. Their approach works only in a single view and can deal with multiple non-lambertian materials as well as unknown lighting; it however requires placement of calibration objects in the scene and change of lighting.

The approach we present in this paper is an extension of shape from shading methods to handle specular surface reflectance, uncontrolled illumination, and multiple views. We focus on single-material objects, and assume that light sources are distant. We also assume there is no shadow or inter-reflection effects on the object. Our approach requires neither the knowledge of light sources nor the light calibration tools. In fact, the object itself serves as the calibration source. By imposing a global lighting constraint, we can recover the 3D shape of the object, as well as a View-Independent Reflectance Map (VIRM), which is a representation of the joint effect of the surface reflectance and environment illumination. This information allows us to predict the appearance of the object not only from novel viewpoints, but also under transformed illumination.

This paper is organized as follows: Section 2 formulates the problem in an energy minimization framework. Section 3 introduces the View Independent Reflectance Map, which is used as the reflection model for our approach. Section 4 presents details of our estimation algorithms. Experimental results on both synthetic and real data are given in Section 5. Section 6 presents conclusions and future work.

2. Problem Formulation

Our objective is to reconstruct the 3D shape and surface reflectance of an object from multiple images taken from different viewpoints, given the intrinsic and extrinsic camera parameters of each image and that the object is made of a single material. This problem can be formulated in an energy minimization framework. The energy function E can be written as a weighted sum of the image/data term E_{image} and the regularization term E_{reg}, as in the following equation:

$$E = E_{image} + w_{reg} * E_{reg} \tag{1}$$

We represent the object shape as a triangular mesh, which consists of a set of vertices and their connectivities. Given the connectivity, the mesh can be deformed by changing its vertex positions, which we will denote as the shape parameter vector V. The BRDF of the surface is denoted by $\rho(\theta_i, \phi_i, \theta_o, \phi_o)$, where (θ_i, ϕ_i) and (θ_o, ϕ_o) are the pairs of the polar and azimuthal angles of the distant light direction and viewing direction in the local surface coordinates, respectively. Consider a patch P around the point of interest, small enough so that it can be well approximated by a flat surface triangle. The brightness of the patch $R(\theta_o, \phi_o)$, when viewed from a certain direction, can be computed by multiplying the BRDF with the foreshortened lighting distribution $L(\theta_i, \phi_i)$ and integrating the product over the upper hemisphere of the patch:

$$R(\theta_o, \phi_o) = \int_0^{\frac{\pi}{2}} \int_0^{2\pi} \rho(\theta_i, \phi_i, \theta_o, \phi_o) L(\theta_i, \phi_i) \times \cos\theta_i \sin\theta_i d\theta_i d\phi_i \quad (2)$$

Given the shape V, BRDF model ρ and lighting L, we can synthesize the images of the object using (2) as the following. Let $\pi_j : \mathbf{R}^3 \to \mathbf{R}^2$ denote the perspective projection that maps the 3D world coordinates onto a 2D image plane corresponding to the jth image $I^{(j)}$. For each triangle P visible in $I^{(j)}$, let $O_j(I^{(j)}) = \pi_j(P)$ be the projection of P onto $I^{(j)}$ (Fig. 1). We can compute the brightness of O_j in the synthesized image using (2). Our image energy term can then be written as the difference between the input images and the images synthesized using the model parameters, V, ρ and L:

$$\begin{aligned} E_{image}(V, \rho, L) &= \sum_j D\big(I^{(j)}_{syn}, I^{(j)}_{input}\big) \\ &= \sum_j \sum_{patch\ P\ visible\ in\ j} \big\{d\big[O_j\big(I^{(j)}_{syn}\big), O_j\big(I^{(j)}_{input}\big)\big]\big\}^2 \end{aligned} \quad (3)$$

where $D(I^{(j)}_{syn}, I^{(j)}_{input})$ is the difference between a synthesized image and the input image, $d(\cdot, \cdot)$ is the analogous difference between image patches, and $O_j(I^{(j)})$ is the set of pixels covered by patch O_j in image $I^{(j)}$. We choose the following $d(\cdot, \cdot)$:

$$\begin{aligned} d\big[O_j\big(I^{(j)}_{syn}\big), O_j\big(I^{(j)}_{input}\big)\big] \\ = \{R_j(P) - \text{mean}\big[O_j\big(I^{(j)}_{input}\big)\big]\} \cdot n(O_j) \end{aligned} \quad (4)$$

where $R_j(P)$ is the brightness of P in $I^{(j)}$ computed from (2), $mean[\cdot]$ is the average pixel value in the patch, and $n(\cdot)$ is the number of pixels in the patch.

The regularization term E_{reg} is used to impose a general smoothness constraint on the object shape to reduce the effect of noise. The actual choice of the smoothness regularization will be discussed later in Section 4. Our goal now is to find the optimal parameters that minimize the total energy:

$$\langle V, \rho, L\rangle = \arg\min_{V,\rho,L} [E_{image}(V, \rho, L) + w_{reg} * E_{reg}(V)] \quad (5)$$

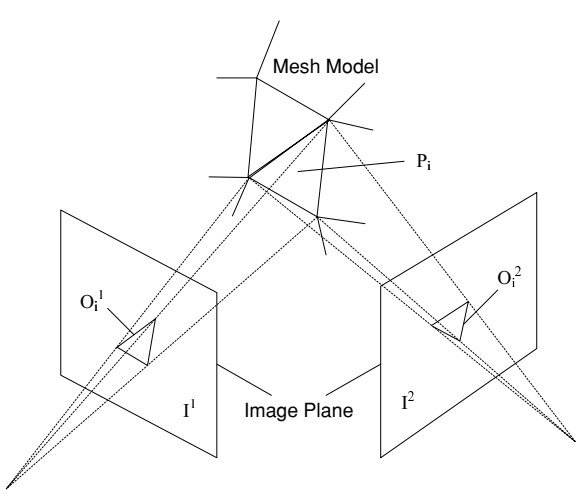


Figure 1. Synthesizing the appearance of the object from different viewpoints by projecting the triangles on the object mesh to different image planes. P_i is a triangle patch on the object mesh, O_{i^1} and O_{i^2} are two projected triangles of P_i to the image plane I^1 and I^2.

3. View Independent Reflectance Map

Reflectance map is used in shape from shading methods to represent the mapping between surface normal and the brightness value viewed from a certain direction. It avoids the separate estimation of the lighting and BRDF, yet contains enough information to recover shape from shaded images. For non-lambertian objects, however, the reflectance map is viewpoint dependent. If multiple reflectance maps are used for different viewpoints, they will inevitably increase the redundancy and the risk of over-fitting. On the other hand, Ramamoorthi and Hanrahan (2001) point out that given the surface shape, there is an inherent ambiguity when one tries to fully recover the BRDF ρ and lighting L. A blurred

light source and a sharp BRDF lead to the same results as those from a sharp light source and a low-pass BRDF. A good intermediate representation that avoids the above two problems is the reflection map (Miller and Hoffman, 1984). Although the reflection map is originally proposed to render specular objects from arbitrary viewpoints given the object geometry and illumination, we show in this paper that the same representation (with some modifications) can also be used to assist the estimation of the object geometry. To distinguish this representation from the reflectance map commonly used in previous shape from shading research, we refer to it as the View-Independent Reflectance Map (VIRM), which emphasizes that it is a viewpoint independent representation of the non-lambertian surface reflection.

3.1. *The VIRM Model*

According to Miller and Hoffman (1984), the surface reflection of a specular object is a sum of different components represented as spherical reflection maps. For an object with both diffuse and specular reflectance, the brightness of a surface patch with normal $\vec{n}$ and viewing direction $\vec{e}$ can then be expressed as:

$$R(\vec{n}, \vec{e}) = R_d(\vec{n}) + R_s(\vec{r}), \tag{6}$$

where $R_d(\vec{n})$ is the diffuse component—a 2D spherical function of the surface normal. $R_s(\vec{r})$ is the specular component. It is a 2D spherical function of the reflected viewing direction $\vec{r}$ (Fig. 2). Here we drop the linear weights on the diffuse and specular components since we consider only objects with homogeneous reflectance where the weights can be absorbed into R_d and R_s.

The method proposed in Miller and Hoffman (1984) requries the two spherical functions R_d and R_s to be the filtered versions of the same environment illumination. This dependency between the two components will require adding nonlinear constraints and will increase the complexity of the estimation problem. In our approach, we treat R_d and R_s as two independent components of the VIRM model. This assumption allows us to pose the estimation of VIRM given the observations and the surface geometry as a simple linear least squares problem. In this paper, we represent both R_d and R_s using spherical harmonics (Weisstein, 2005). The spherical harmonic bases Y_{lm} are given by:

$$N_{lm} = \sqrt{\frac{2l+1}{4\pi}\frac{(l-m)!}{(l+m)!}} \tag{7}$$

$$Y_{lm}(\theta, \phi) = N_{lm} P_l^m(\cos\theta) e^{Im\phi}, \tag{8}$$

where $P_l^m(z)$ is an associated Legendre polynomial, $l \geq 0$ and $-l \leq m \leq l$. R_d and R_s can be written as two truncated series:

$$R_d(\theta, \phi) = \sum_{l=0}^{N_d} \sum_{m=-l}^{l} R_d^{lm} Y_{lm}(\theta, \phi), \tag{9}$$

$$R_s(\theta, \phi) = \sum_{l=0}^{N_s} \sum_{m=-l}^{l} R_s^{lm} Y_{lm}(\theta, \phi), \tag{10}$$

where R_d^{lm} and R_s^{lm} are the coefficients of the spherical harmonic bases. Spherical harmonics representation allows us to control the level-of-detail in the VIRM estimation by choosing the maximal order N_d and N_s. The number of coefficients to be estimated for an order N approximation is $(N+1)^2$. Given the observed brightness, estimating these coefficients is a linear problem.

3.2. *The Relationship between VIRM and other Reflectance Models*

The VIRM model is an approximation to many parametric BRDF models (e.g. Torrance-Sparrow

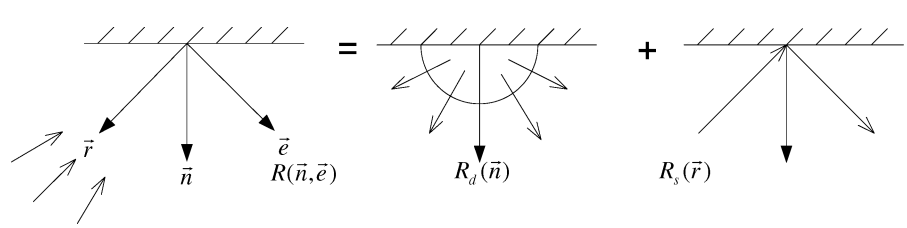


Figure 2. The View Independent Reflectance Map model. The surface reflection $R(\vec{n}, \vec{e})$ viewed from direction $\vec{e}$ under a fixed illumination is modeled as the sum of two independent 2D spherical functions. $R_d(\vec{n})$ is the diffuse component of VIRM which only depends on the surface normal $\vec{n}$. $R_s(\vec{r})$ is the specular component of VIRM which depends on the reflected viewing direction $\vec{r}$.

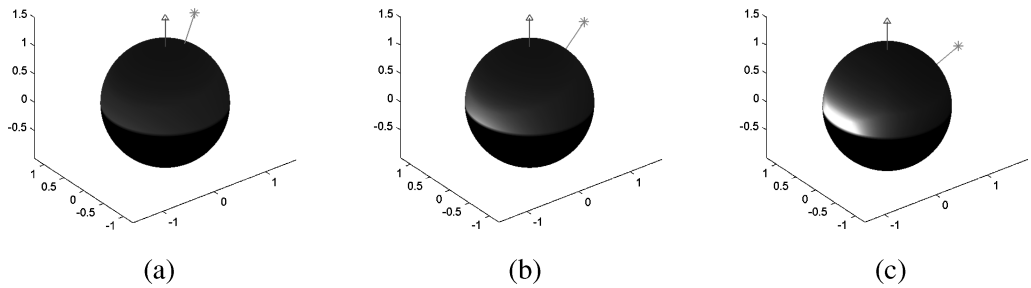


Figure 3. Plot of the $F(\mu, \vec{n}, \vec{l}, \vec{e})G(\vec{n}, \vec{l}, \vec{e})/4(\vec{e} \cdot \vec{n})$ in the Torrance-Sparrow BRDF model with respect to the viewing directions (upper hemisphere) under different illumination directions (blue $\triangle$ is the normal direction $\vec{n}$ and red $*$ represents the illumination direction $\vec{l}$). Intensity value is proportional to the function value. (a) Light angle $\cos^{-1}(\vec{l} \cdot \vec{n}) = \pi/8$ (b) Light angle $\cos^{-1}(\vec{l} \cdot \vec{n}) = \pi/4$ (c) Light angle $\cos^{-1}(\vec{l} \cdot \vec{n}) = 3\pi/8$. For most of the upper hemishpere $F(\mu, \vec{n}, \vec{l}, \vec{e})G(\vec{n}, \vec{l}, \vec{e})/4(\vec{e} \cdot \vec{n})$ remains constant except for very large viewing angles.

model (1967), isotropic Ward model (1992) and Phong model (1975)) proposed for objects with specular highlights. As an example, we show the connection of VIRM to these models by simplifying the physics based Torrance-Sparrow BRDF model. According to the Torrance-Sparrow model, the BRDF of a material can be written as:

$$\begin{aligned}\rho(\theta_i, \phi_i, \theta_o, \phi_o) &= \rho(\vec{n}, \vec{l}, \vec{e}) \\ &= K_d + K_s \frac{F(\mu, \vec{n}, \vec{l}, \vec{e})G(\vec{n}, \vec{l}, \vec{e})D(\sigma, \vec{n}, \vec{l}, \vec{e})}{4(\vec{l} \cdot \vec{n})(\vec{e} \cdot \vec{n})}, \quad (11)\end{aligned}$$

where $\vec{n}, \vec{l}$ and $\vec{e}$ are surface normal, light direction and viewing direction vectors, respectively. $F(\mu, \vec{n}, \vec{l}, \vec{e})$ is the Fresnel term, related to the material's index of refraction μ, $G(\vec{n}, \vec{l}, \vec{e})$ is the geometric attenuation term, and $D(\sigma, \vec{n}, \vec{l}, \vec{e})$ is the microfacet normal distribution function. The brightness value of a patch illuminated by a directional source L is given by

$$\begin{aligned}&R(\vec{n}, \vec{e}, \vec{L}, \rho) \\ &= |\vec{L}| \cdot \left[K_d(\vec{l} \cdot \vec{n}) + K_s \frac{F(\mu, \vec{n}, \vec{l}, \vec{e})G(\vec{n}, \vec{l}, \vec{e})D(\sigma, \vec{n}, \vec{l}, \vec{e})}{4(\vec{e} \cdot \vec{n})} \right], \quad (12)\end{aligned}$$

where $\vec{L} = |\vec{L}| \cdot \vec{l}$ is the light vector for the directional source.

For simplicity we assume $F(\mu, \vec{n}, \vec{l}, \vec{e})G(\vec{n}, \vec{l}, \vec{e})/4(\vec{e} \cdot \vec{n})$ to be constant and absorb them into K_s. This approximation is valid when the viewing angle is quite different from $\pi/2$. Figure 3 shows the $F(\mu, \vec{n}, \vec{l}, \vec{e})G(\vec{n}, \vec{l}, \vec{e})/4(\vec{e} \cdot \vec{n})$ as a function of the viewing direction for different illumination directions according to the Torrance-Sparrow model. We can see that the value remains near constant over the viewing hemisphere, except for the area where the viewing angle is close to $\pi/2$.

Now let us consider the microfacet normal distribution function. A simple form of D is:

$$D(\sigma, \vec{n}, \vec{l}, \vec{e}) = \frac{1}{\pi\sigma^2} \exp\left(-\left(\frac{\theta_h}{\sigma}\right)^2 \right), \quad (13)$$

where $\cos\theta_h = \vec{n} \cdot \vec{h}$, and $\vec{h}$ is the vector half way between $\vec{l}$ and $\vec{e}$ (Fig. 4) and σ is the variance of the microfacet normals. Let us take the mirror image of viewing direction $\vec{e}$ with respect to the surface normal $\vec{n}$ and denote it as the reflection vector $\vec{r}$. If the light direction $\vec{l}$ is co-planar with the surface normal $\vec{n}$ and viewing direction $\vec{e}$, we will have:

$$\theta_{rl} = 2\theta_h, \quad (14)$$

where θ_{rl} is the angle between the reflection vector $\vec{r}$ and the light direction vector $\vec{l}$. Substituting (14) into

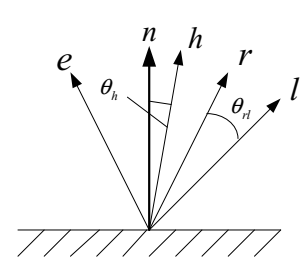


Figure 4. The angular relation between the viewing direction $\vec{e}$, the surface normal $\vec{n}$, the reflected viewing direction $\vec{r}$ and the mid-vector $\vec{h}$. If $\vec{e}$, $\vec{n}$ and $\vec{l}$ are co-plane, then angle θ_{rl} between $\vec{r}$ and $\vec{l}$ is twice of the angle θ_h.

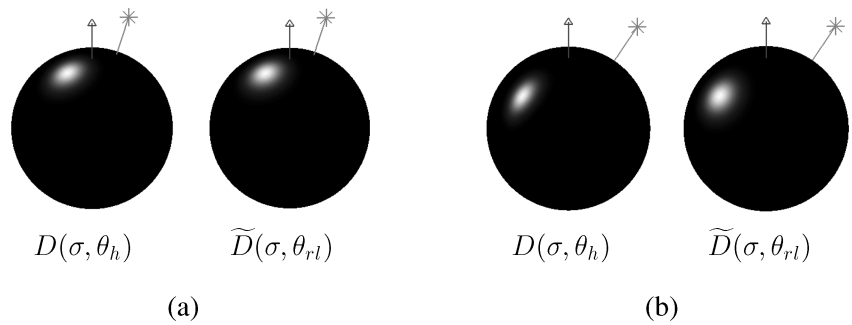


Figure 5. Comparison of the micro-facet distribution function $D(\sigma, \vec{n}, \vec{l}, \vec{e})$ and its approximation $\widetilde{D}(\sigma, \theta_{rl})$ for different light directions. Function values are plotted with respect to the viewing direction on the unit sphere. Blue $\triangle$ is the normal direction $\vec{n}$ and red $*$ is the light direction $\vec{l}$. (a) Light angle $cos^{-1}(\vec{l} \cdot \vec{n}) = \pi/8$ (b) Light angle $cos^{-1}(\vec{l} \cdot \vec{n}) = \pi/4$. The deviation from D increases with the increase of light angle.

(13), and denoting it as $\widetilde{D}$, we get:

$$\widetilde{D}(\sigma, \theta_{rl}) = \frac{1}{\pi\sigma^2} \exp\left(-\left(\frac{\theta_{rl}}{2\sigma}\right)^2\right) \tag{15}$$

Generally, D is not symmetric around $\vec{r}$. So strictly speaking, $\widetilde{D} \neq D$ when $\vec{l}$ deviates from the plane determined by $\vec{e}$ and $\vec{n}$. However, Ramamoorthi and Hanrahan (2001) point out that when viewing angle (angle between $\vec{e}$ and $\vec{n}$) is small, assuming D is symmetric around $\vec{r}$ is a good approximation. Under this assumption, we can use $\widetilde{D}(\sigma, \theta_{rl})$ to approximate $D(\sigma, \theta_h)$. Now the brightness value in (12) is

$$R(\vec{n}, \vec{e}, \vec{L}, \rho) = |\vec{L}| \cdot K_d(\vec{l} \cdot \vec{n}) + K_s|\vec{L}|\widetilde{D}(\sigma, \theta_{rl}) \tag{16}$$

Figure 5 shows the distribution of function $D(\sigma, \theta_h)$ and $\widetilde{D}(\sigma, \theta_{rl})$ with respect to the viewing vector for different light directions. As the angle between light vector and surface normal vector θ_{nl} becomes larger, the bright area (where the reflected light is clearly visible) of $D(\sigma, \theta_h)$ becomes more elliptical while for $\widetilde{D}(\sigma, \theta_{rl})$ it remains circular.

If all the patches have the same material and all the surface patches are illuminated under the same lighting, the diffuse term (16) depends only on the surface normal $\vec{n}$, and the specular term depends only on θ_{rl}. Furthermore, we can merge $|\vec{L}|$ and $\widetilde{D}(\sigma, \theta_{rl})$ in the specular term and view it as the result of filtering the single directional light source with a circular symmetric function $\widetilde{D}$. Since the light source is fixed, the merged term depends only on $\vec{r}$. Therefore (16) can be written as:

$$R(\vec{n}, \vec{e}, \vec{L}, \rho) = R_d(\vec{n}) + R_s(\vec{r}) \tag{17}$$

This is exactly the VIRM representation for the joint effect of illumination and BRDF. This approximation is derived under single directional light source assumption, but it can be extended to the case of multiple directional light sources since both distant illumination model and the circular symmetric filtering are linear operations. The VIRM model uses the property that the specular part of the BRDF can be approximated by a function that has only the reflection angle θ_{rl} as its variable. Note this is also true for the Phong reflectance model (Phong, 1975). Therefore, the VIRM model is also compatible with the Phong model.

3.3. *VIRM based Shape Reconstruction*

VIRM serves the same role in our reconstruction algorithm as the reflectance map in single view shape from shading. If we assume that all the surface patches of the object have the same BRDF and the lighting remains constant, then the VIRM model is constant for all the patches and viewing directions. This gives a global constraint over all the surface patches and input views. By using VIRM as the reflectance model, we can write our optimization target in (5) as:

$$\langle V, R_d, R_s \rangle = \arg\min_{V, R_d, R_s} [E_{image}(V, R_d, R_s) + w_{reg} * E_{reg}(V)] \tag{18}$$

However, we should point out that when there are local variations of lighting caused by non-distant light sources, self-shadows or inter-reflections, VIRM will not necessarily be constant. Also, in the process of reducing the Torrance-Sparrow model to VIRM, we

make the assumption that $FG/(\vec{e}\cdot\vec{n})$ in (12) is constant and D can be approximated by $\widetilde{D}$, both of which require the viewing angle to be sufficiently different from $\pi/2$ to keep the approximation valid.

4. Algorithm and Implementation

In this section we present the various aspects of the algorithm we have used to implement the approach described in Sections 2 and 3.

4.1. Overall Algorithm

Equation (18) defines a nonlinear optimization problem with a large number of parameters to be chosen. However, it is easy to find out that the VIRM parameters are linearly constrained. Since E_{reg} depends only on the object shape, if all the shape parameters are fixed, estimating the optimal VIRM is just a linear least squares problem. Because of this, we choose to optimize the shape and VIRM parameters separately and interleave these optimization processes, as illustrated in Fig. 6.

The inputs to our algorithm are the object images taken from different viewpoints and the corresponding camera parameters. A coarse visual hull (represented as a triangular mesh) is computed from the silhouettes of the object and used as the initial shape for the first VIRM optimization. During the VIRM optimization, we fix all the vertices of the mesh and find an optimal VIRM that minimizes the total energy in (18). During the shape optimization, we fix the VIRM parameters and optimize the mesh vertex positions that minimize the total energy. The iteration is terminated when the average vertex position change after the shape optimization is smaller than a preset threshold.

4.2. VIRM Optimization

When shape parameters are fixed, optimizing (18) to find VIRM is equivalent to solving a set of linear equations in least squares sense. Each visible triangle patch in a single view gives a linear equation in $R_d(\vec{n})$ and $R_s(\vec{r})$, which can be converted into a linear equation in coefficients R_d^{lm} and R_s^{lm} by using Eqs. (9) and (10). We filter out patches that have large viewing angles ($>70^\circ$ in our experiments) to avoid poor constraints being used in estimating VIRM. Because the number of patches is usually greater than the number of coefficients, we still obtain an over-determined system for solving the VIRM coefficients.

4.3. Shape Optimization

We represent the object surface using a triangular mesh, where each triangle serves the role of patch P in (4), and the 3D positions of all the vertices in the mesh are the shape parameters V. Shape optimization in (18) for a fixed VIRM is a non-linear least squares problem. We solve it using the trust region reflection newton method implemented in Matlab (2004). The gradient of the energy function with respect to each vertices are computed usingfinite differencing. The visibility

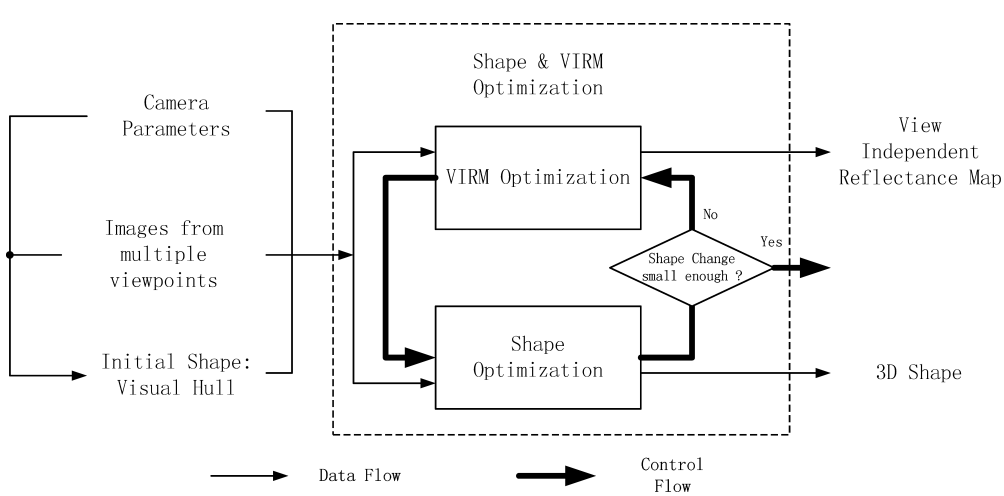


Figure 6. Flow chart of the iterative optimization algorithm for the shape and VIRM parameters.

of each triangle is determined by projecting the current mesh to input viewpoints while removing the hidden surfaces. Similar to the VIRM estimation, patches that have large viewing angles ($>70°$) are not used in computing E_{image} in (3). Due to the continuous shape deformation during the optimization, the elimination of the oblique triangles might cause the changing of optimization target and possible divergence of the algorithm. We currently do not have a theoretical proof of stability of the optimization algorithm, but in practice, the overall system is highly constrained and converges in a multi-camera configuration.[1]

We choose the total surface area of the mesh as the shape regularization energy:

$$E_{reg}(V) = \sum_i A(F_i) \tag{19}$$

where $A(F_i)$ is the area of the ith triangle in the mesh. This smoothness term will favor surfaces with smaller overall areas to reduce the effect of noise.

Since each vertex on the mesh model has three degrees of freedom, the number of parameters that represent the shape is three times the number of vertices. To reduce the number of parameters, we impose a restriction that each vertex can only move along the vertex normal direction.[2] In addition to reducing the number of shape parameters, this restriction also prevents vertices from clustering together during the optimization. At each iteration, the visibilities and vertex normals are updated according to the current estimate of the shape.

Our algorithm requires a fairly good initial shape estimate, since the trust region reflective newton method only finds a local minimum. The need to solve iteratively the VIRM further limits the size of the attraction region of the optimization algorithm. This is in contrast to the method proposed in Jin et al. (2004) where a cube shape initialization is enough. This can be explained by both the modeling and algorithmic differences. First, we use a more complex spherical harmonics illumination representation while Jin et al. (2004) uses an ambient and positive/negative point sources model. Second, we consider both diffuse and specular component instead of just the diffuse reflectance in Jin et al. (2004). Third, we do not use a constant radiance background model. Finally, the choices of different optimization methods (level set evolution v.s. gradient based method over the mesh) also contribute to the difference.

To ensure a close enough initialization, we choose the visual hull computed from silhouette cone intersection as the initial shape of our algorithm.[3] The mesh topology and connectivity of the initial mesh are fixed in the entire shape optimization process in our experiments. However, if necessary, they can also be updated by using re-meshing operations (e.g. converting the mesh into a signed distance function and extracting the zero isosurface). The visual hull is also used as an outer bound of the shape being estimated.

4.4. Multi-scale Processing

To avoid local minima and for computational efficiency, we use multi-scale processing in the optimization. We first optimize the shape parameters using a coarse triangular mesh and a low order spherical harmonics for VIRM. Then we use $\sqrt{3}$-subdivision scheme (Kobbelt, 2000) to subdivide the triangles to obtain a finer mesh and also increase the maximal order of spherical harmonics in the VIRM.

5. Experiments

We test our algorithm on three data sets. The first one contains a set of synthetic images of a sphere. The experiment on this set is meant to test the effectiveness of the VIRM estimation algorithm. The second data set shows a synthetic buddha statue that has a more complex shape, containing many surface details, and under a complex illumination setting. The third is a real data set of a Van Gogh statue taken under uncontrolled indoor illumination.

5.1. VIRM Validation

We first validate the VIRM model using synthetic images. A set of 24 views of a specular sphere is synthesized and used as the input to the VIRM optimization method described in Section 4.2. We assume the sphere radius is known and want to verify if the VIRM model can reproduce the specular reflectance of the sphere.

In VIRM optimization, we need to choose the maximal order of the spherical harmonics for diffuse and specular components. The order should be high enough to model the frequencies in the actual diffuse and specular components, and yet still low to avoid overfitting. We made these choices based on the visual quality of the reconstructions. We choose the maximal orders that did not introduce extraneous high frequency variations in the reconstructions, which is 5 for diffuse VIRM, and 7 for specular VIRM. Figure 7 shows four of the in-

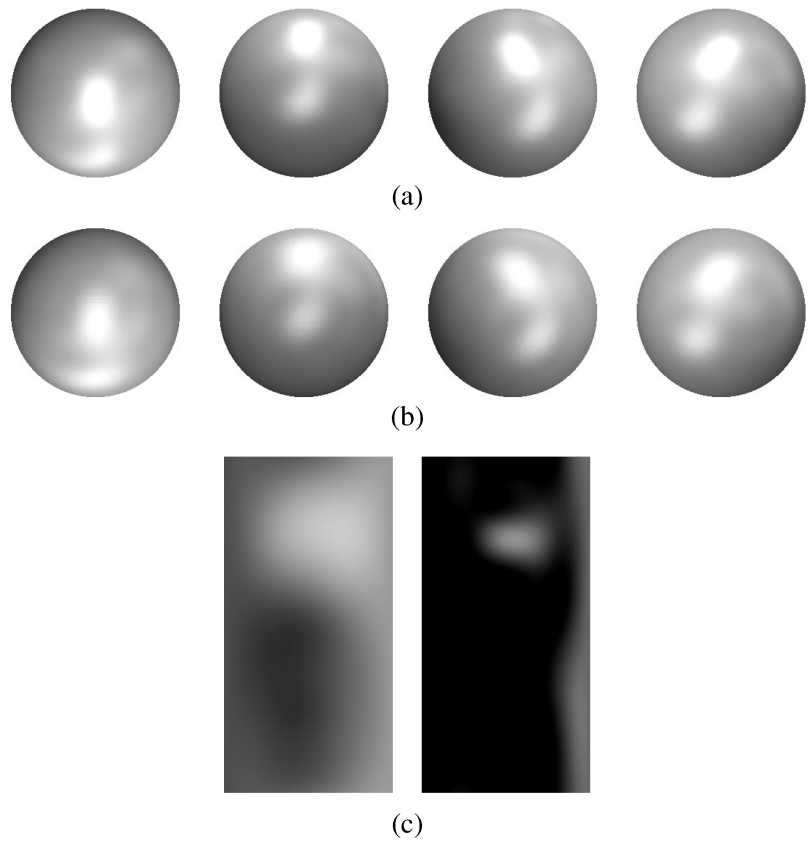


Figure 7. (a) Four of the 24 input sphere images (b) Sphere rendered using estimated VIRM (average absolute image difference over all 24 views is 0.0075 on a scale of [0, 1]) (c) Estimated diffuse (left) and specular (right) component of VIRM along a grid defined by latitude $(0 - \pi)$ and longitude $(0 - 2\pi)$.

put sphere images as well as the corresponding images rendered using estimated VIRM. Visually, the reconstruction matches the originals well except for some highlight regions where image values are saturated, and some areas where viewing angles are large. The average absolute image difference between the input and reconstructed images over the entire set is 0.0075 on a scale of [0, 1]. The reconstructed VIRM is plotted in longitude-latitude format as shown in Fig. 7(c).

5.2. *Buddha Data Set (Synthetic)*

The Buddha data set is a synthetic data set which consists of 24 views of a Buddha sculpture made from a single shiny material. The sculpture is illuminated by 60 directional light sources from a clustered result of the light probe data (Debevec, 2004). Some input images are shown in Fig. 8. We run our algorithm at three different scales. The numbers of triangles at the three scales are around 6300, 19000, and 57000. The maximal order of spherical harmonics for diffuse VIRM is 3 and 7 for specular VIRM. The final reconstructed shape is also shown in Fig. 8, along with the ground truth shape and the initial shape.

By comparing the second row and the fourth row in Fig. 8, we can see that overall the algorithm correctly recovers the surface features of the object. We can find that areas inside Buddha's arms are not well reconstructed. These concave areas have shadows and inter-reflections that are not modeled in VIRM. In addition, the optimization algorithm is more prone to getting stuck in local minima in these areas, which results in several wrinkles that are visible.

To visually evaluate the quality of the VIRM estimates, we render an synthetic sphere under different estimated VIRMs. Figure 9(a, b) shows the sphere rendered under the estimated VIRMs using the initial

Figure 8. First row: 5 of the 24 input images of the Buddha data set. Second row: The ground truth 3D model rendered with a dull material to eliminate specularities, thus making visual evaluation of shape easier. Third row: The initial 3D shape, computed as the intersection of silhouette cones from the input images. Fourth row: The recovered 3D shape using our algorithm.

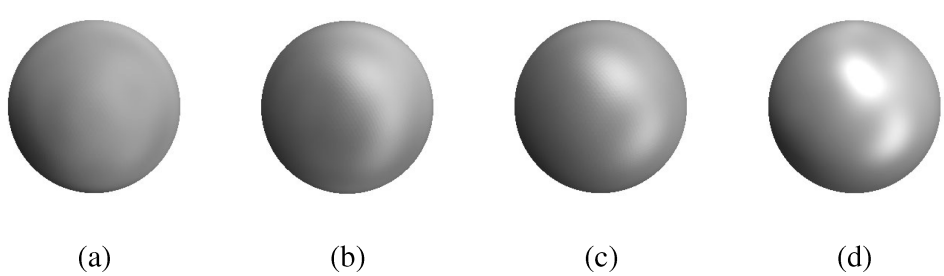


Figure 9. Sphere rendered using different VIRM estimates of the Buddha data set. (a) Estimated VIRM using the initial shape from silhouette cone intersection. (b) Estimated VIRM using the final shape of our algorithm. (c) Estimated VIRM using the ground truth shape. (d) Ground truth sphere image rendered under the same illumination as the buddha.

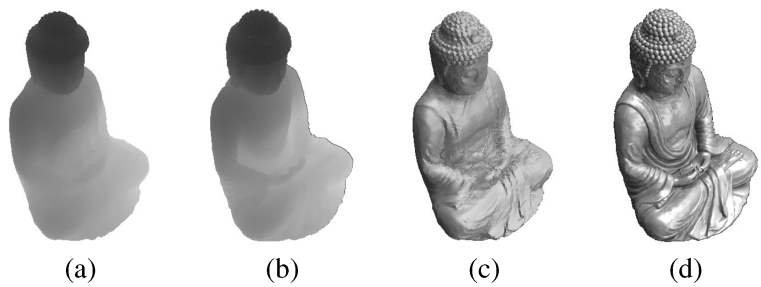


Figure 10. Synthesized images from a novel viewpoint. (a) Range image computed from the estimated shape. (b) Range image computed from the ground truth shape. (c) A gray scale image with estimated VIRM and shape. (d) Ground truth image not used as input.

shape from silhouette and the final shape of our algorithm. With the improvement of the shape estimate, the estimated VIRM shows more specular component. Figure 9(c) is rendered under the VIRM estimated using the ground truth shape. The difference between (b) and (c) is solely due to the shape estimation error. The sphere rendered in the ground truth illumination is also shown in (d) for comparison. We can see even with ground truth shape, the sphere image (c) shows a considerable amount of under-estimation of the specular component comparing with (d). There are several reasons for this underestimate. First, unlike the sphere data set, the buddha data set has self-shadowing and inter-reflection effects that are not modeled in VIRM. Second, even without the self-shadowing and inter-reflection, the VIRM model is still an approximation to the actual surface reflectance. Third, the spherical harmonic representation of VIRM introduces some truncation error. Forth, the intensity saturation of the specular highlights causes the input to the specular VIRM estimation to be smaller than the actual value.

We also synthesize the range images of the estimated shape, and the intensity image using reconstructed shape and VIRM. Figure 10(a, c) show these images from a novel viewpoint, compared with the ground truth images not used as input. To evaluate the performance of our algorithms quantitatively, we use several measures (Fig. 11). We compute the average absolute pixel difference between ground truth and synthesized intensity images. Average Object Image Difference (AOID) and Average Sphere Image Difference (ASID) denote the differences for the synthesized object and sphere images, respectively. AOID reflects the quality of both shape and VIRM estimates, whereas ASID reflects the quality of the VIRM estimate. Ratio of Uncovered Area (RUA) is the percentage of the non-overlapping silhouette areas between the ground truth and synthesized objects. Pixel values in these uncovered areas are not

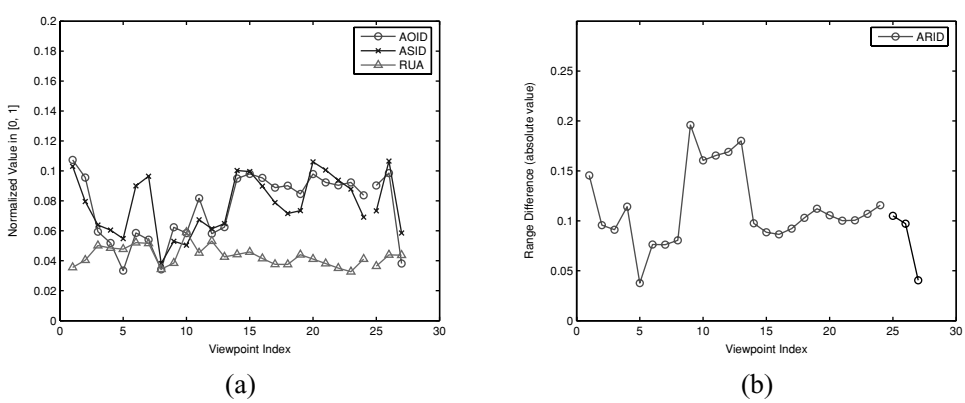


Figure 11. The various performance measures shown for different viewpoints. (a) Average Object Image Difference (AOID), Average Sphere Image Difference (ASID) and Ratio of Uncovered Area (RUA), value normalized to [0, 1]. (b) Average Range Image Difference (ARID, absolute value); the object's bounding box is about $5 \times 5 \times 7$ and distance to camera is 15. For both (a) and (b), data points 1–24 are from input views, and 25–27 are from novel views.

(a) (b) (c) (d) (e)

Figure 12. Synthesized novel view using estimated VIRM (a), novel view with VIRM rotated by 60 degree (c), to be compared with ground truth (b, d). (e) is another object synthesized using the same VIRM as (a).

defined in either synthesized image or ground truth image, so we do not include them in the calculation of image differences. Finally, Average Range Image Difference (ARID) more directly measures the errors in estimated shape by computing average absolute object range difference between synthesized range images from estimated shape and those from the ground truth. In Fig. 11(b), high ARID values are due to views that have self-occluding boundaries. Since the recovered self-occluding boundaries are not fully aligned with the actual boundaries, they cause large differences in the range image.

We also synthesize an image from a novel viewpoint using estimated VIRM (Fig. 12(a)) and another image with VIRM rotated by 60 degrees (Fig. 12(c)). The images are compared with ground truth images in Fig. 12. Another new object rendered using the extracted VIRM is shown in Fig. 12(e).

5.3. Van Gogh Data Set (Real)

The Van Gogh data set is provided by J.-Y. Bouguet and R. Grzeszczuk of Intel. It consists of more than 300 calibrated images of a Van Gogh statue. We selected 21 images taken from different directions. These images are manually segmented to remove the background and the silhouettes are used to compute the initial shape. We segment out the base of the statue since it is made of a different material. Five of the input images are shown in the first row of Fig. 13. We also have the 3D shape obtained by structured light scanning of the statue (Fig. 13 second row).

The minimization is done at two different scales. The numbers of triangles at the two scales arearound 10,000 and 30,000. Since the statue is made of polished metal, which exhibits a typical metal BRDF with almost no diffuse component, we choose a very low maximal order for the diffuse component. The maximal order of spherical harmonics for diffuse VIRM is 2, and for specular VIRM is 7. The reconstructed shape is shown in the fourth row of Fig. 13. Note that calibration errors are present in the reconstruction and they affect both the VIRM and shape estimation.

Figure 14 shows the sphere rendered under different VIRMs estimated from the Van Gogh data set, where (a) and (b) compare the estimated VIRMs using initial shape and the final shape of our algorithm. Figure 14(c) shows the render sphere under the VIRM estimated from the structured light scanned shape. It appears to be more blurry than (b) because the scanned shape lacks many detailed surface normal variations than those recovered by our algorithm. We do not have the ground truth sphere images under the original illumination, but overall, the estimated VIRM has blurry specular component due to the same reason mentioned in Section 5.2. Nevertheless, the plot of the specular component of VIRM still shows the distribution of the illumination, which is composed of four light sources in the upper hemisphere and some reflection from the desktop in the lower hemisphere.

We again use AOID, ARID and RUA defined in Section 5.2 to evaluate the performance of our algorithm. However, since we do not have the ground truth lighting data from the original data set, we cannot compute the ASID. The synthesized gray scale image (Fig. 15(a)) and range image (Fig. 15(c)) for one novel view are also shown. We also synthesize the Buddha statue image (Fig. 15(f)) with the estimated VIRM. Performance measures for all viewpoints are summarized in Fig. 16.

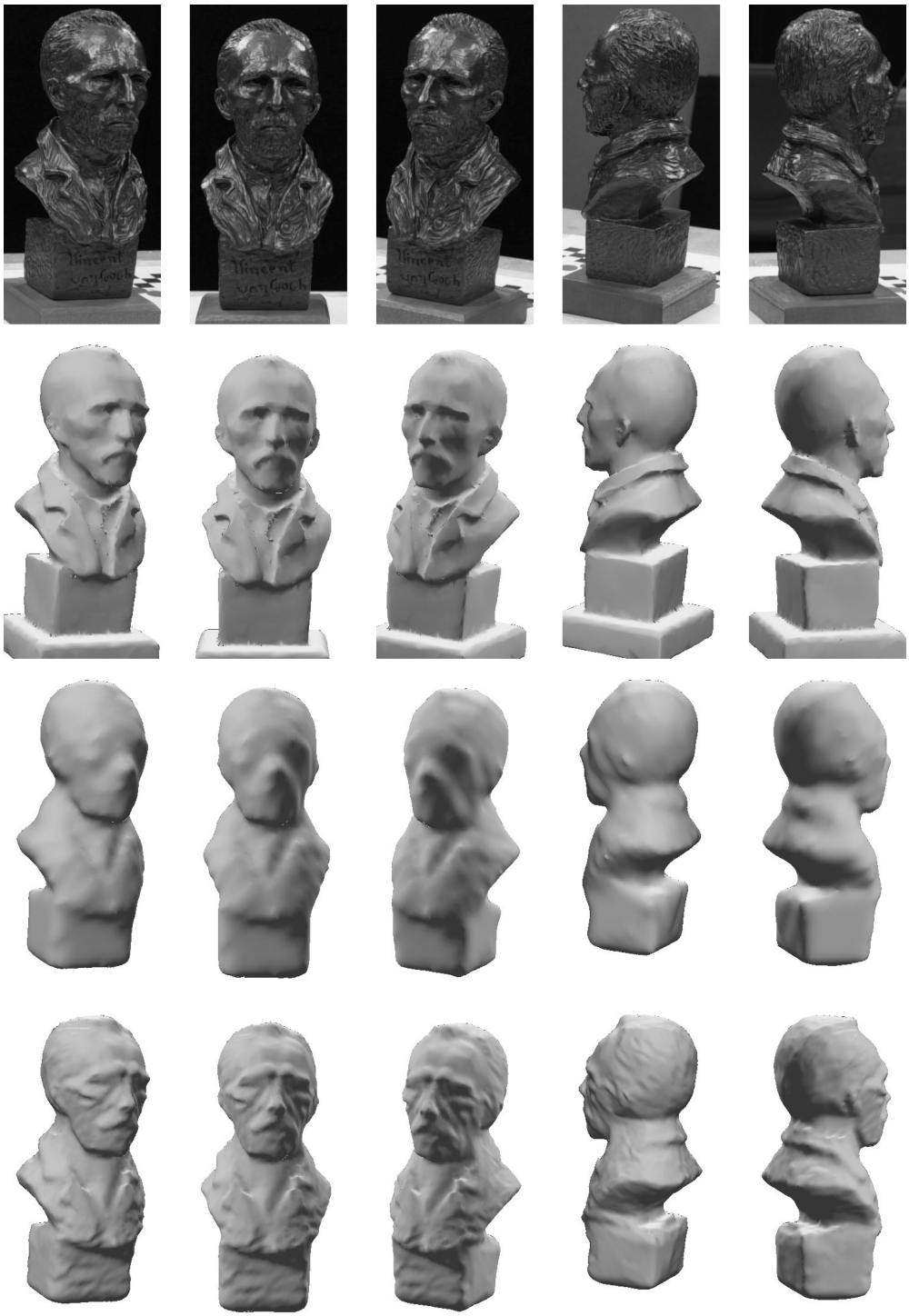
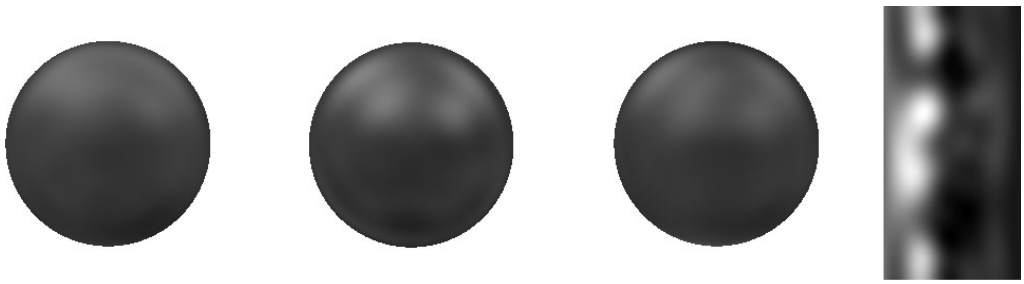
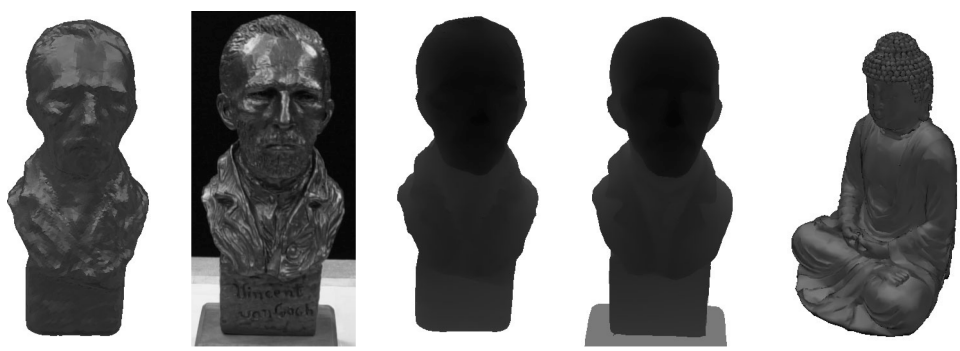


Figure 13. First row: 5 of the 21 input images. Second row: Shape obtained by structured light scanning rendered with a dull material for visual shape comparison. Third row: Initial shape computed from silhouettes of the input images. Fourth row: The recovered 3D shape using our algorithm.



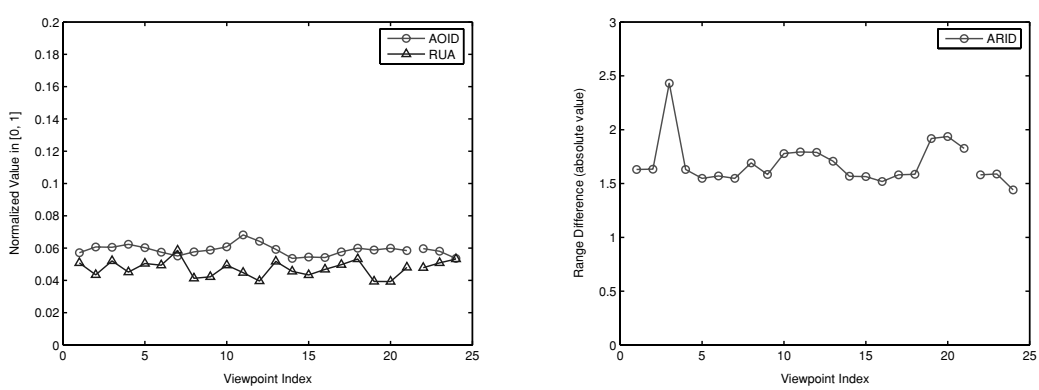
(a) (b) (c) (d)

Figure 14. Sphere rendered using different VIRM estimates of the Van Gogh data set. (a) Estimated VIRM using initial shape from silhouette cone intersection. (b) Estimated VIRM using the final shape of our algorithm. (c) Estimated VIRM using the structured light scanned shape. (d) The specular VIRM plotted in latitude-longitude format, showing the four light sources in the environment (left part) and some reflections of the white desktop (right part). The intensity of (d) is scaled for better visualization.



(a) (b) (c) (d) (e)

Figure 15. (a) Synthesized gray scale image with estimated VIRM and shape from a novel viewpoint. (b) Ground truth image from the same viewpoint. (c) Range image computed from estimated shape. (d) Range image obtained from a laser scan. (e) Synthesized Buddha with the estimated VIRM.



(a) (b)

Figure 16. The various performance measures of the Van Gogh data set, shown for different viewpoints. (a) Average Object Image Difference (AOID) and Ratio of Uncovered Area (RUA), value normalized to [0, 1]. (b) Average Range Image Difference (ARID, absolute value); the bounding box of the object is about $90 \times 80 \times 200$, distance to camera is about 950. For both (a) and (b), data points 1–21 are from input images, and 22–24 are from novel viewpoints.

This data set is also used in Jin et al. (2003). Interested readers can compare the two results. Our major improvements are the recovery of shape details, and since VIRM is estimated, we get a compact reflectance map that can predict the appearance of any shape and from novel viewpoints or under transformed illumination.

6. Conclusions and Future Work

In this paper, we have proposed an algorithm to reconstruct the 3D shape and the View Independent Reflectance Map (VIRM) from multiple calibrated images of an object. We pose the reconstruction task as minimizing the difference between the input images and the synthesized images while regularizing the process by enforcing the surface to be smooth. The VIRM is an approximation to other parametric BRDF models, and is used as a simplified model for single material reflectance under distant lighting with no self-shadowing and no inter-reflections. An iterative method is used to minimize the cost function to find the optimal shape and VIRM. Our algorithm does not require the light source to be known, and it can deal with specular reflectance. Experimental results on both synthetic and real objects show that our algorithm can recover both the 3D shape and the VIRM information.

We plan to investigate further on how the estimated VIRM can be used to render other objects under the same illumination, or to create animations that are consistent with the original lighting. Alternatively, it is possible to change the material/lighting of the synthesized image by directly modifying the VIRM. Other directions include taking into account self-shadowing and inter-reflection, and allowing objects made of multiple materials.

Notes

1. For multiple cameras, patches whose corresponding constraints are removed in one camera can still have constraints in other cameras that have smaller viewing angles. In our experiments, we use more than 20 cameras, making the average ratio of the number of constraints versus the number of vertices about 20:1. Even after removing the oblique triangles this ratio is around 16:1, which means we are still solving an over-determined system.
2. The vertex normal is computed by averaging the surface normals of its neighboring triangles.
3. The object silhouettes can be obtained by using background subtraction methods, interactive or automatic segmentation methods.

References

Bhat, D.N. and Nayar, S.K. 1998. Stereo and specular reflection. *Int. J. Comput. Vision*, 26(2):91–106.

Debevec, P. 2004. *High Dynamic Range Image Processing and Manipulation* [online] Available: http://www.debevec.org/ HDRShop/

Faugeras, O. and Keriven, R. 1998. Variational principles, surface evolution, pdes, level set methods, and the stereo problem. *IEEE Transactions on, Image Processing*, 7(3):336–344.

Fua, P. and Leclerc, Y.G. 1995. Object-centered surface reconstruction: Combining multi-image stereo and shading. *Int. J. Comput. Vision*, 16(1):35–55.

Georghiades, A. 2003. Incorporating the torrance and sparrow model of reflectance in uncalibrated photometric stereo. In *CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on, Computer Vision and Pattern Recognition*, vol. 2, IEEE Computer Society, pp. 816–823.

Hertzmann, A. and Seitz, S. 2003. Shape and materials by example: A photometric stereo approach. In *Proceedings. 2003 IEEE Computer Society Conference on, Computer Vision and Pattern Recognition*, vol. 1, pp. 533–540.

Jin, H., Cremers, D., Yezzi, A., and Soatto, S. 2004. Shedding light on stereoscopic segmentation. In *Proceedings. 2004 IEEE Computer Society Conference on, Computer Vision and Pattern Recognition*, vol. 1, pp. 36–42.

Jin, H., Soatto, S., and Yezzi, A. 2003. Multi-view stereo beyond lambert. In *Proceedings. 2003 IEEE Computer Society Conference on, Computer Vision and Pattern Recognition*, vol. 1, pp. 171–178.

Kobbelt, L. 2000. $\sqrt{3}$-subdivision. In *Proceedings of the 27th Annual Conference on Computer Graphics and Interactive Techniques*, pp. 103–112. ACM Press/Addison-Wesley Publishing Co.

Kutulakos, K.N. and Seitz, S.M. 2000. A theory of shape by space carving. *Int. J. Comput. Vision*, 38(3):199–218.

Lee, K.M. and Kuo, C.-C.J. 1997. Shape from shading with a generalized reflectance map model. *Comput. Vis. Image Underst.*, 67(2):143–160.

Lin, S., Li, Y., Kang, S.B., Tong, X., and Shum, H.-Y. 2002. Diffuse-specular separation and depth recovery from image sequences. In *Proceedings of the 7th European Conference on Computer Vision-Part III*, pp. 210–224. Springer-Verlag.

Matlab Optimization Toolbox, 2004, The MathWorks, Inc. http://www.mathworks.com/access/helpdesk/help/toolbox/optim/.

Miller, G.S. and Hoffman, R. 1984. Illumination and reflection maps: Simulated objects in simulated and real environments. In *Course Notes for Advanced Computer Graphics Animation, SIGGRAPH 84*.

Phong, B.T. 1975. Illumination for computer generated pictures. *Commun. ACM*, 18(6):311–317.

Ramamoorthi, R. and Hanrahan, P. 2001. A signal-processing framework for inverse rendering. In *Proceedings of the 28th Annual Conference on Computer Graphics and Interactive Techniques*, pp. 117–128. ACM Press.

Samaras, D., Metaxas, D., Fua, P., and Leclerc, Y.G. 2000. Variable albedo surface reconstruction from stereo and shape from shading. In *Proceedings. 2000 IEEE Computer Society Conference on, Computer Vision and Pattern Recognition*, vol. 1, pp. 480–487.

Samaras, D. and Metaxas, D.N. 2003. Incorporating illumination constraints in deformable models for shape from shading and

light direction estimation. *IEEE Trans. Pattern Anal. Mach. Intell.*, 25(2):247–264.

Torrance, K.E. and Sparrow, E.M. 1967. Theory for off-specular reflection from roughened surfaces. *Journal of the Optical Society of America*, 57(9):1105–1114.

Treuille, A., Hertzmann, A., and Seitz, S.M. 2004. Example-based stereo with general brdfs. In *Proceedings of the 8th European Conference on Computer Vision-Part II*, pp. 457–469. Springer-Verlag.

Ward, G.J. 1992. Measuring and modeling anisotropic reflection. In *SIGGRAPH '92: Proceedings of the 19th Annual Conference on Computer Graphics and Interactive Techniques*, pp. 265–272, New York, NY.

Weisstein, E.W. 2005. Spherical harmonic. *MathWorld–A Wolfram Web Resource*.

Yu, Y., Debevec, P., Malik, J., and Hawkins, T. 1999. Inverse global illumination: recovering reflectance models of real scenes from photographs. In *Proceedings of the 26th Annual Conference on Computer Graphics and Interactive Techniques*, pp. 215–224. ACM Press/Addison-Wesley Publishing Co.

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

Zickler, T., Belhumeur, P.N., and Kriegman, D.J. 2002. Helmholtz stereopsis: Exploiting reciprocity for surface reconstruction. *Int. J. Comput. Vision*, 49(2/3):215–227.