Shape and View Independent Reflectance Map from Multiple Views [*]

Tianli Yu, Ning Xu, and Narendra Ahuja

Beckman Institute & Electrical and Computer Engineering Department
University of Illinois at Urbana-Champaign, Urbana IL 61801, USA
{tianli, ningxu, ahuja}@vision.ai.uiuc.edu

Abstract. We consider the problem of estimating the 3D shape and reflectance properties of an object made of a single material from a calibrated set of multiple views. To model reflectance, we propose a View Independent Reflectance Map (VIRM) and derive it from Torrance-Sparrow BRDF model. Reflectance estimation then amounts to estimating VIRM parameters. We represent object shape using surface triangulation. We pose the estimation problem as one of minimizing cost of matching input images, and the images synthesized using shape and reflectance estimates. We show that by enforcing a constant value of VIRM as a global constraint, we can minimize the matching cost function by iterating between VIRM and shape estimation. Experiment results on both synthetic and real objects show that our algorithm is effective in recovering the 3D shape as well as non-lambertian reflectance information. Our algorithm does not require that light sources be known or calibrated using special objects, thus making it more flexible than other photometric stereo or shape from shading methods. The estimated VIRM can be used to synthesize views of other objects.

1 Introduction

Many multiple-view algorithms have been proposed over the years for 3D reconstruction. These algorithms can be generally classified into image centered or object/scene centered. Image centered algorithms [1] first search for pixel correspondences followed by triangulation. Object/scene centered approaches are another category that has been explored recently [2, 3]. A model of the object or scene is built and a consistency function is defined over the input images; maximizing the function achieves a 3D model that is most consistent with all the input views. In each approach, objects are frequently assumed to have Lambertian reflectance to facilitate finding correspondences. One exception is the radiance tensor field introduced by Jin, et al [3]. They propose a rank constraint of radiance tensor to recover the 3D shape. This is essentially a local reflectance

constraint to model both lambertian and non-lambertian objects. However, constructing the radiance tensor requires that every scene point be seen by a substantial number of cameras. In addition, the estimates obtained by most of these algorithms are confined to individual pixels and they usually cannot recover fine details of the shape, e.g., those encoded by shading.

Shape from shading algorithms, on the other hand, have the potential to recover greater details about surface shape, e.g., surface normal changes from image shading. However, shape from shading algorithms are usually developed for constrained environments, such as single material objects, lambertian reflectance, single viewpoint, known or very simple light source, orthographic projection, and absence of shadows and interreflections. Zhang, et al.[4] present a recent survey of shape from shading methods. Samaras, et al.[5] propose to incorporate shape from shading method into multiple-view reconstruction. They consider lambertian objects and recover piece-wise constant albedo as well as surface shape. In their method, specularities are detected and removed. Although, for lambertian objects complex lighting can be well modeled locally using a single point light source, this is not the case for specular objects. Hertzmann and Seitz [6] use a calibration sphere together with the object to obtain a reflectance map that can be used to recover the shape. Their approach works with 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 object centered and extends the work on shape from shading to allow non-lambertian surface reflectance, uncontrolled lighting, and the use of multiple views. We focus on single material objects, and assume that light sources are distant and there are no shadows or interreflection effects. Our approach does not require the knowledge of light sources or light calibration tools. In fact, the object itself serves as the calibration source. We show that 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 allows us to render from any view point the same or any other object, made of the same material and under the same lighting.

This paper is organized as follows: Section 2 formulates the problem as a minimization problem. Section 3 derives the VIRM. Section 4 presents our estimation algorithms. Experimental results on both synthetic and real data are given in Section 5. Section 6 presents conclusions and extensions.

2   Problem Formulation

Our objective is to reconstruct the 3D shape and reflectance information from multiple images of an object, given images of the object from different viewpoints, the intrinsic and extrinsic camera parameters for each image, and the knowledge that the object is made of a single material. This problem can be posed as that of minimization of the differences between the input images and the images synthesized using the underlying shape, lighting and BRDF model.

Suppose the surface of the object is $S(V)$, where V is the parameter vector of the object shape. The BRDF of the surface is denoted as $\rho(\theta_i', \phi_i', \theta_o', \phi_o')$, where $(\theta_i', \phi_i'), (\theta_o', \phi_o')$ are polar and azimuthal angles of the distant light direction and viewing direction in the local surface coordinates. Consider a patch P on S, small enough so that the surface normal remains nearly constant over the entire patch. The brightness of the patch when viewed from a certain direction $R(\theta_o', \phi_o')$ 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, as in (1):

$$R(\theta_o', \phi_o') = \int \int \rho(\theta_i', \phi_i', \theta_o', \phi_o') L(\theta_i', \phi_i') \cos \theta_i' \sin \theta_i' \mathrm{d}\theta_i' \mathrm{d}\phi_i' \qquad (1)$$
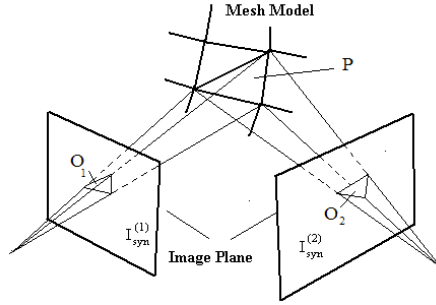

Fig. 1. Project a patch on the surface onto image planes

Given the shape $S(V)$, BRDF model ρ and lighting L, we can synthesize the images of the object using (1) as follows. Let $\pi_j : \mathbf{R}^3 \rightarrow \mathbf{R}^2$ denote the perspective projection that maps the 3D world coordinates onto a 2D image plane corresponding to the jth view. For each P, let $O_j = \pi_j(P)$ be the projection of P onto jth input image (Fig.1). If P is visible in jth view, then we can compute the intensity value of O_j in the synthesized jth view using (1). Our goal is to estimate the model parameters, V, ρ and L, that minimize the difference between the input images and these synthesized images:

$$\langle V, \rho, L \rangle = \arg \min F_{matching}(I_{syn}, I_{input}) \qquad (2)$$

where $F_{matching}$ denotes the matching cost function between input images and synthesized images. It is defined as the sum of all intensity differences between the corresponding patches in the input and synthesized images:

$$F_{matching}(I_{syn}, I_{input}) = \sum_j D(I_{syn}^{(j)}, I_{input}^{(j)})$$

$$= \sum_{for all P on S} \sum_{P is visible in j} \{d[O_j(I_{syn}^{(j)}), O_j(I_{input}^{(j)})]\}^2 \quad (3)$$

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

$$d[O_j(I_{syn}^{(j)}), O_j(I_{input}^{(j)})] = \{R_j(P) - \text{mean}[O_j(I_{input}^{(j)})]\} \cdot n(O_j) \qquad (4)$$

where $R_j(P)$ is the reflectance of P in jth image computed from (1), $mean[\cdot]$ is the average pixel value in the patch, and $n(\cdot)$ is the number of pixels in the patch.

3   View Independent Reflectance Map

Reflectance map is used in shape from shading research to give 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. However, reflectance map is viewpoint dependent, which makes its use inconvenient for multiple-view algorithms. Ramamoorthi and Hanrahan [8] point out that given a 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 a sharp light source and a low-pass BRDF. We use this property to model the specular light reflected by a BRDF as the same light passing through a circular symmetric low-pass filter and then reflected by a perfect mirror. Based on this idea, we introduce the notion of View-Independent Reflectance Map (VIRM) which we use to represent the combined effects of lighting L and BRDF ρ independent of the viewpoints. In this section we show that we can derive VIRM by separating the diffuse and specular parts of reflectance.

As mentioned in Section 2, the brightness value of a surface point can be computed from (1). Specifically, we can use the Torrance-Sparrow microfacet model [7] as the BRDF model and simplify it to derive our VIRM. According to the model, the BRDF of a material can be written as:

$$\rho(\theta_i', \phi_i', \theta_o', \phi_o') = \rho(\boldsymbol{n}, \boldsymbol{l}, \boldsymbol{e}) = K_d + K_s \frac{F(\mu, \boldsymbol{n}, \boldsymbol{l}, \boldsymbol{e})G(\boldsymbol{n}, \boldsymbol{l}, \boldsymbol{e})D(\sigma, \boldsymbol{n}, \boldsymbol{l}, \boldsymbol{e})}{4(\boldsymbol{l} \cdot \boldsymbol{n})(\boldsymbol{e} \cdot \boldsymbol{n})} \qquad (5)$$

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

$$R(\boldsymbol{n}, \boldsymbol{e}, \boldsymbol{L}, \rho) = |\boldsymbol{L}| \cdot [K_d(\boldsymbol{l} \cdot \boldsymbol{n}) + K_s \frac{F(\mu, \boldsymbol{n}, \boldsymbol{l}, \boldsymbol{e})G(\boldsymbol{n}, \boldsymbol{l}, \boldsymbol{e})D(\sigma, \boldsymbol{n}, \boldsymbol{l}, \boldsymbol{e})}{4(\boldsymbol{e} \cdot \boldsymbol{n})}] \qquad (6)$$

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

For simplicity we assume F and G to be constant and absorb them into K_s. Now let us consider the microfacet normal distribution function. A simple form of D is

$$D(\sigma, \boldsymbol{n}, \boldsymbol{l}, \boldsymbol{e}) = \frac{1}{\pi\sigma^2} \exp\left(-\left(\frac{\theta_h}{\sigma}\right)^2\right), \quad \cos\theta_h = \boldsymbol{n} \cdot \boldsymbol{h} \tag{7}$$

where $\boldsymbol{h}$ is the mid-vector between $\boldsymbol{l}$ and $\boldsymbol{e}$ (Fig. 2) and σ is the variance of the microfacet normals. Let us take the mirror image of viewing direction $\boldsymbol{e}$ with respect to the surface normal and denote it as the reflection vector $\boldsymbol{r}$, as in Fig. 2. If the light direction $\boldsymbol{l}$ is co-plane with the surface normal $\boldsymbol{n}$ and viewing direction $\boldsymbol{e}$, we will have

$$\theta_{rl} = 2\theta_h \tag{8}$$

where θ_{rl} is the angle between the reflection vector $\boldsymbol{r}$ and the light direction vector $\boldsymbol{l}$ (Fig. 2). Substituting (8) into (7), and denoting it as $\widetilde{D}$, we get:

$$\widetilde{D}(\sigma, \theta_{rl}) = \frac{1}{\pi\sigma^2} \exp\left(-\left(\frac{\theta_{rl}}{2\sigma}\right)^2\right) \tag{9}$$
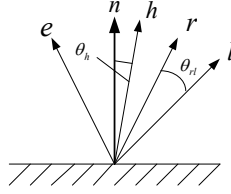

Fig. 2. Reflection vector $\boldsymbol{r}$ and mid-vector $\boldsymbol{h}$

Generally, D is not symmetric around $\boldsymbol{r}$. So strictly speaking, $\widetilde{D} \neq D$ when $\boldsymbol{l}$ deviates from the plane determined by $\boldsymbol{e}$ and $\boldsymbol{n}$. However, Ramamoorthi and Hanrahan [8] point out that when viewing angle is small, assuming D is symmetric around $\boldsymbol{r}$ is a good approximation. Under this assumption, we can use $\widetilde{D}(\sigma, \theta_{rl})$, which is a function of σ and θ_{rl} to approximate D. Now the reflectance value in (6) is

$$R(\boldsymbol{n}, \boldsymbol{e}, \boldsymbol{L}, \rho) = |\boldsymbol{L}| \cdot K_d(\boldsymbol{l} \cdot \boldsymbol{n}) + K_s \frac{|\boldsymbol{L}|\widetilde{D}(\sigma, \theta_{rl})}{4(\boldsymbol{e} \cdot \boldsymbol{n})} \tag{10}$$

In (10) the first term is the diffuse part, and the second term is the specular part. If all the patches have the same material and the lighting is constant with respect to the world coordinate system (e.g. all the surface patches are illuminated under the same lighting), the diffuse term depends only on the surface normal $\boldsymbol{n}$, and the specular term depends only on θ_{rl} and the viewing angle $\boldsymbol{e} \cdot \boldsymbol{n}$. Furthermore, in the specular term, we can merge $|\boldsymbol{L}|$ and $\widetilde{D}(\sigma, \theta_{rl})$ together 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 $\boldsymbol{r}$ and we denote it as $R_s(\boldsymbol{r})$. Similarly, the first term on the right side of (10) depends only on $\boldsymbol{n}$ and is denoted as $R_d(\boldsymbol{n})$. So (10) becomes:

$$R(\boldsymbol{n}, \boldsymbol{e}, \boldsymbol{L}, \rho) = R_d(\boldsymbol{n}) + \frac{R_s(\boldsymbol{r})}{\boldsymbol{n} \cdot \boldsymbol{e}} \tag{11}$$

Meanwhile, since $\boldsymbol{r}$ is the mirror vector of $\boldsymbol{e}$, the right side of equation (11) only depends on $\boldsymbol{e}$ and $\boldsymbol{n}$. Equation (11) gives a very compact way to represent the reflectance of a surface patch under fixed lighting. It is just a linear combination of two components, the diffuse part and the specular part, and each can be represented as a 2D function (since $\boldsymbol{n}$ and $\boldsymbol{r}$ are both unit vectors). The approximation is derived under single directional light source assumption, but it can be extended to the cases of multiple directional light sources since both distant illumination model and the circular symmetric filtering are linear operations.

The simplified model in (11) implies that if we can estimate the diffuse and specular distributions R_d and R_s, we can compute the reflectance of any point given its surface normal and viewing direction. We call R_d and R_s the diffuse and specular components of the VIRM. They serve the same roles as reflectance map in single view shape from shading.

If we assume that all the surface patches have the same BRDF and the lighting remains constant, then the VIRM is constant for all the patches and viewing directions. This is equivalent to a global constraint over all the surface patches and input views. By using VIRM as our reflectance model, we can write (2) as:

$$\langle V, R_d, R_s \rangle = \arg \min F_{matching}(I_{syn}, I_{input}) \tag{12}$$

However, we should point out that when there are local variations of lighting such as due to a non-distant light sources, self-shadowing or inter-reflection, VIRM will not necessarily be constant. Our derivation of VIRM makes the assumption that F and G in (6) are constant and D can be approximated by $\widetilde{D}$, both assumptions require the viewing angle away from $\pi/2$ to approximates accurately the Torrance-Sparrow model.

4   Algorithm and Implementation

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

4.1   Data Representation

We use a triangular mesh to represent the object surface, 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. VIRM is represented by a collection of samples of the diffuse and specular distribution functions. We choose the longitude-latitude grid to sample the azimuthal and polar components at a fixed angular interval. Function values that are not on the sampling grid are computed using cubic interpolation.

4.2   Iterative Optimization

Equation (12) defines a nonlinear optimization problem with a large number of parameters to be chosen. However, note that the VIRM parameters are only linearly constrained. If we fix all the shape parameters, estimating the optimal VIRM is just a constrained 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. 3.
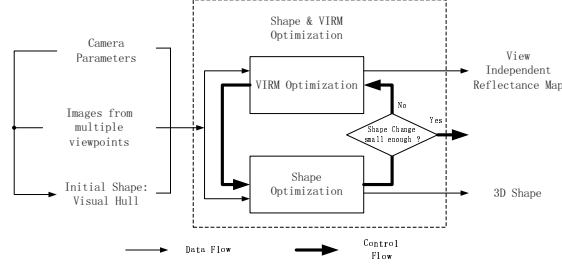

Fig. 3. Flow chart of the iterative optimization algorithm

The inputs to our algorithm are the object images taken from different viewpoints and the corresponding camera parameters. A coarse visual hull is computed from the silhouettes of the object (silhouettes can be obtained by segmentation or background subtraction) and used as the initial shape for the first VIRM optimization. During VIRM optimization, we fix all the shape parameters and find an optimal VIRM that minimizes the matching cost function in (3). During shape optimization, we fix the VIRM parameters and find an optimal set of shape parameters that minimize the matching cost function. The iteration is terminated when the average vertex change after shape optimization is smaller than a preset threshold.

4.3   VIRM and Shape Optimization

When shape parameters are fixed, optimizing (12) to find VIRM is equivalent to solving a set of linear equations in least squares sense. Each visible triangle patch in one view gives a linear equation of $R_d(\boldsymbol{n})$ and $R_s(\boldsymbol{r})$. Because of the discretization, we let the equations constrain the nearest samples of VIRM. We filter out patches that have large viewing angles (> 80 degree in our experiments) to avoid poor constraints being used in estimating VIRM. The optimal solution gives estimates of all values of $R_d(\boldsymbol{n})$ and $R_s(\boldsymbol{r})$ on the sample grid. Some samples on the VIRM grid may not have any constraint; we obtain their values by interpolation.

Shape optimization in (12) for a fixed VIRM is a non-linear least squares problem. Again, for the same reason, patches that are tilted away from the camera are not used in computing $F_{matching}$ in (3). This won't create many

unconstrained patches though, since in a multi-camera configuration every patch must has some cameras facing toward it. We solve the optimization using the large scale optimization method called Trust Region Reflective Newton (TTRN) method [9]. In TTRN method, the matching cost function is viewed as the square of norm of a multi-input multi-output (MIMO) function. Every iteration of the optimization involves the approximate solution of a large linear system using the method of preconditioned conjugate gradients (PCG). TRRN method requires the Jacobian matrix of the MIMO function, and this can be computed using finite difference.

Since each vertex on the mesh model has 3 degree-of-freedom, the number of parameters that represent the shape is 3 times the number of vertices. To reduce the number of parameters, we impose a restriction that each vertex can only move along a specific direction. This direction, called the weighted average normal direction (WAND), is the average of the surface normal vectors over all the triangles sharing the vertex, weighted by the areas of these triangles. In addition to reducing the number of shape parameters, this restriction also prevents vertices from clustering together during optimization. At each iteration, the visibilities and WANDs of vertices are updated according to the current estimate of the shape. Also, the visual hull computed from silhouettes is 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 use a low sampling rate for VIRM. Then we iteratively reduce the triangle size and increase the VIRM sampling rate. Triangles having larger gray level variations at a coarse scale are subdivided into four small triangles to obtain finer scale triangles. They are constructed from 3 new vertices which are the mid-points of three edges of the coarse triangle.

5   Experiments

5.1   VIRM Validation

We first perform a synthetic experiment to validate our VIRM model. A set of 20 images of a sphere is synthesized and used as the input to VIRM optimization described in 4.3. We assume the sphere radius are known and want to check whether the simplified VIRM model can reproduce the non-lambertian reflectance of the sphere.

Fig. 4 shows four of the input sphere images as well as the corresponding images rendered using reconstructed VIRM. The sampling grid for diffuse VIRM is 18x9, and specular VIRM is 32x16. The result shows that the reconstruction matches the originals well except for some highlights 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 0.017 on a scale of [0,1]. The grid representation of the reconstructed VIRM is shown in Fig. 4(c).
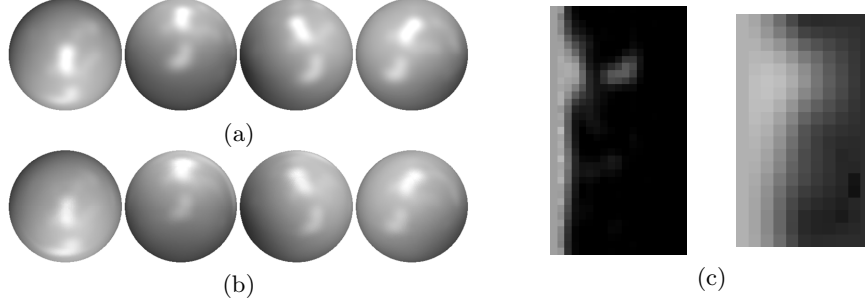

(a)

(b)

(c)

Fig. 4. (a) 4 of the 20 input sphere images (b) Sphere rendered using estimated VIRM (average absolute image difference over all 20 views is 0.017 on a scale of [0,1]) (c) Estimated specular(left) and diffuse(right) component of VIRM along a grid defined by longitude and latitude.

5.2 Buddha Data Set (Synthetic)

The Buddha data set is synthetic and consists of 24 views of a Buddha sculpture made from a single shiny material. The sculpture is illuminated by 60 directional light sources. Some input images are shown in Fig. 5. We run our algorithm at three different scales. The numbers of triangles at each scale are around 6300, 21000, and 50000. The sampling grid for diffuse VIRM is 6x3, 12x6, 18x9, and specular VIRM is 12x6, 24x12, 32x16. The final reconstructed shape is also shown in Fig. 5, compared with the ground truth shape and the initial shape.

By comparing Fig. 5(d-f) and 5(j-l), we can see that Buddha's ears are not well recovered. Thin surface parts are difficult to recover since they do not cause enough image differences to affect the cost function.

To obtain more quantitative measures of the performance of our algorithms, and to seperately evaluate the quality of shape and VIRM based estimates, we compute the range images of the reconstructed shape and the images of a sphere using the estimated VIRM for both the input viewpoints as well as some novel ones. They are compared with ground truth images. The synthesized gray scale image (Fig. 6a), range image (Fig. 6c) and sphere image (Fig. 6e) for one of the novel views are shown in Fig. 6. In Fig. 6(e) the specular highlights on the sphere are not fully recovered. One reason for this is that the surface normal along the surface of the sculpture is not continuous. For example, the shape does not have many surface normals facing downward, so the VIRM estimation is not well constrained in the corresponding direction. Low sample rate of VIRM, noise in the recovered local surface orientation, and other noises such as shadow and inter-reflection that VIRM did not assume also contribute to the reconstruction error.
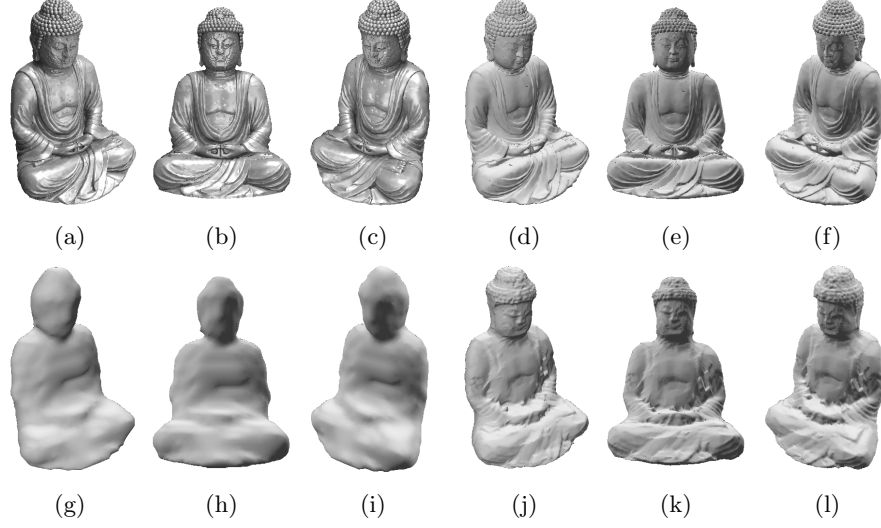
Fig. 5. (a, b, c): Three input images of the data set. (d, e, f): The ground truth 3D model rendered with a dull material to eliminate specularities, which makes visual evaluation of shape easier. (g, h, i): The initial 3D shape computed from silhouettes in the input images. (j, k, l): The recovered 3D shape after optimization.
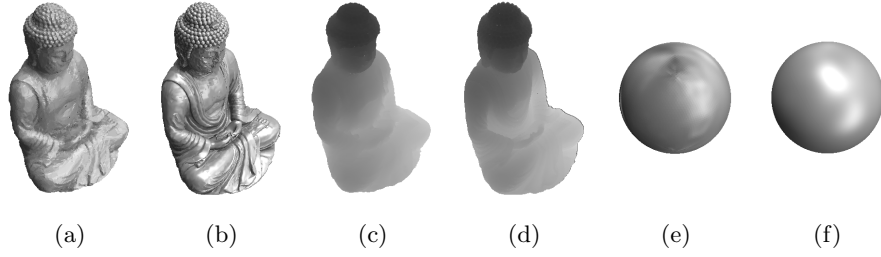

Fig. 6. (a) Synthesized gray scale image with estimated VIRM and shape. (b) Ground truth image. (c) Range image computed from the estimated shape. (d) Range image computed from the ground truth shape. (e) Synthesized sphere with estimated VIRM. (f) Rendered sphere with ground truth material and lighting. All images are from a novel viewpoint.

We evaluate the performance of our algorithms using several measures (Fig. 7). 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 rendered object and sphere images, respectively. AOID reflects the quality of both shape and VIRM estimates, whereas ASID reflects the quality of 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 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) measures more directly the errors in estimated shape by computing average absolute object range difference between synthesized range images from estimated shape and those from ground truth. In Fig. 7(b), images with high ARID values are from views that have occluding boundaries. Since the recovered occluding boundaries are not fully aligned with the actual boundaries, they will create large differences in the range image.
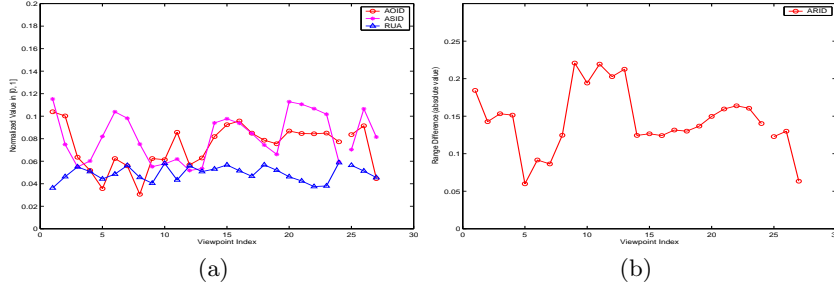

(a)   (b)

Fig. 7. The various performance measures shown for different viewpoints. (a): AOID, ASID and RUA (value normalized to [0,1]) (b): ARID (absolute value); the object's bounding box is about 5x5x7 and distance to camera is 15. For both (a) and (b), datapoints 1-24 are from input views, and 25-27 are from novel views.

We also synthesize an image from a novel viewpoint using estimated VIRM (Fig. 8a) and another image with VIRM rotated by 60 degree (Fig. 8c). The images are compared with ground truth images in Fig. 8. Another object rendered using the same VIRM is shown in Fig. 8(e).

5.3   Van Gogh Data Set(Real)

The Van Gogh data set is by courtesy of J.-Y. Bouguet and R. Grzeszczuk (Intel). It consists of more than 300 calibrated images of a Van Gogh statue. We select 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. Three of the input images are shown in Fig. 9(a-c). We have the reconstruction result from laser scanning of the statue (Fig. 9(d-f)). The
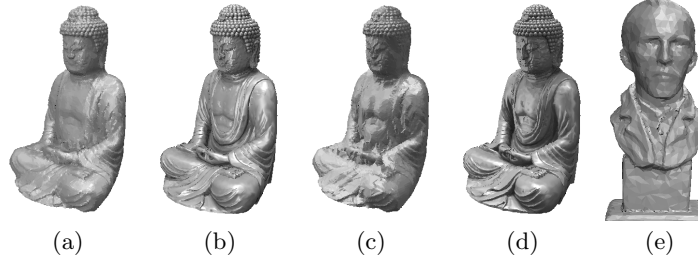
Fig. 8. Synthesized novel view using estimated VIRM (a), novel view with VIRM rotated by 60 degree (c), to be compared with ground truth (b, d). (e) is another object synthesized using the same VIRM as used in (a).

scanned shape is processed by manual mesh cleaning process to make a smooth surface.

The minimization is done at two different scales. The numbers of triangles at the two scales are arround 10000 and 40000. Since the statue is made of polished metal, which exhibits a typical metal BRDF with almost no diffuse component, we choose a very low sampling rate for the diffuse part in VIRM. The sample grids at two scales for diffuse VIRM are 6x3 and 6x3, and specular VIRM are 24x12 and 48x24. The reconstructed shape is shown in Fig. 9 (j-l). Note that calibration errors are present in the reconstruction and they affect both the recovered VIRM and the shape.

We again use AOID and ARID defined in Section 5.2 to evaluate the performance of our algorithm. But since we do not have the lighting data from the original data set, we cannot compute the ASID. The synthesized gray scale image (Fig. 10a) and range image (Fig. 10c) for one novel view are shown below. We also synthesize the sphere image (Fig. 10e) and Buddha image (Fig. 10f) with the estimated VIRM. Performance measures for all viewpoints are summarized in Fig. 11 .

This data set is also used in [3]. 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 synthesize images of any shape from any viewpoint.

6   Conclusion and Fugure Works

In this paper we have proposed an algorithm to reconstruct 3D shape and the view independent reflectance map (VIRM) from multiple calibrated images of the object. We pose this problem as that of minimizing of difference between the input images and the synthesized images using estimated 3D shape and VIRM. VIRM is derived from Torrance-Sparrow model, and used as a simplified model for single material reflectance under distant lighting with no self-shadowing and inter-reflections. An iterative method is used to minimize the matching cost function in order to find the optimal shape and VIRM. Our algorithm does
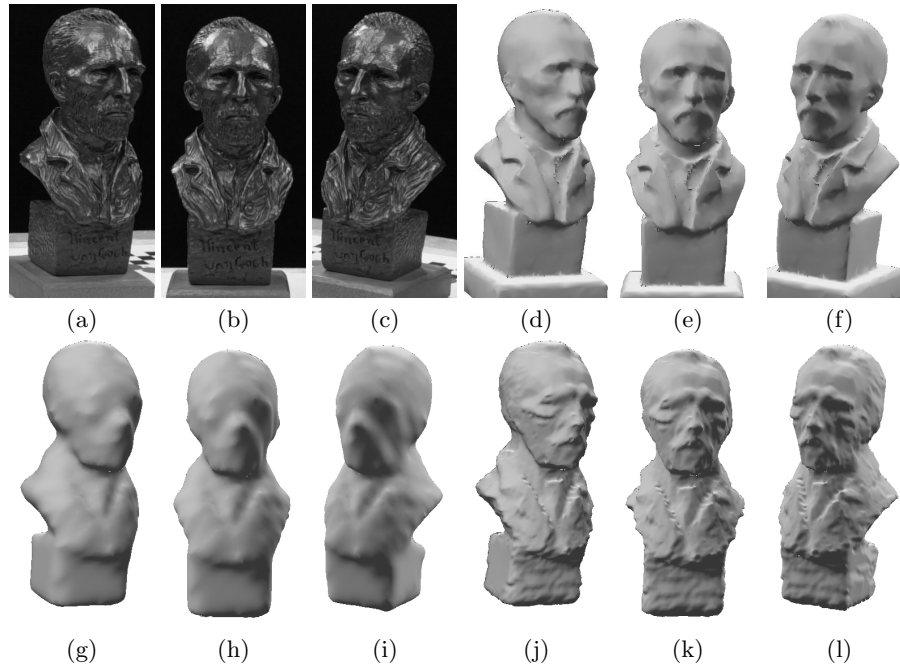
Fig. 9. (a, b, c): Three of the input images. (d, e, f): Shape obtained by laser scanning rendered with a dull material for better shape comparison. (g, h, i):Initial shape of our algorithm computed from silhouettes of the input images. (j, k, l): Reconstructed 3D shape of our algorithm.
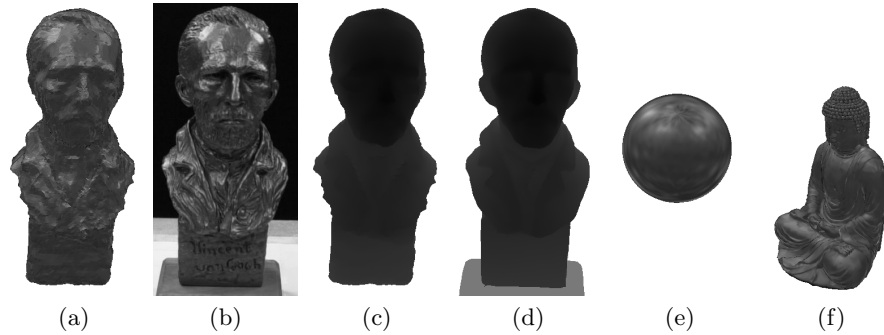

Fig. 10. (a): Reconstructed gray scale image with estimated VIRM and shape from a novel viewpoint (b): Ground truth image from the same viewpoint. (c): Range image computed from estimated shape. (d): Range image obtained from a laser scan. (e): Synthesized sphere with the estimated VIRM. (f): Synthesized Buddha with the estimated VIRM.
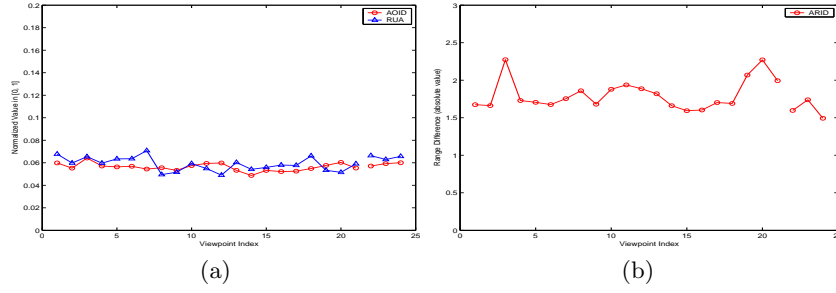
Fig. 11. The various performance measures of Van Gogh data set, shown for different viewpoints (a): AOID and RUA (value normalized to [0, 1]). (b): ARID (absolute value); the bounding box of the object is about 90x80x200, distance to camera is about 950. For both (a) and (b), data points 1-21 are from input images, and 22-24 are from novel viewpoints.

not require the light source to be known, and it can deal with non-lambertian reflectance. Experimental results on both synthetic and real objects show that our algorithm is effective in recovering the 3D shape and the VIRM information.

Our ongoing and planned work includes the following. The estimated VIRM can be used to render other objects with the same material and lighting, or to create animations that are consistent with the original lightings. Alternatively, the material/lighting of the synthesized image can be changed by directly modifying VIRM. Other directions include taking into account the effect of shadowing and inter-reflection and allowing objects with multiple materials.

References

1. R. Koch, M. Pollefeys, L. Van Gool: Multi Viewpoint Stereo from Uncalibrated Video Sequences. Proc. ECCV'98, LNCS, Springer-Verlag, Freiburg, (1998) 55–65
2. S. M. Seitz, C. R. Dyer: Photorealistic Scene Reconstruction by Voxel Coloring. Proc. Computer Vision and Pattern Recognition Conf., (1997) 1067–1073
3. H. Jin, S. Soatto, A. Yezzi: Multi-view Stereo Beyond Lambert. Computer Vision and Pattern Recognition, 2003. Proceedings, Vol. 1, (2003) 171–178
4. Ruo Zhang, Ping-Sing Tsai, James Edwin Cryer, Mubarak Shah: Shape from Shading: A Survey. IEEE Trans. PAMI, Vol 21, No.8, (1999) 690–706
5. D. Samaras, D. Metaxas, P. Fua, Y. G. Leclerc: Variable Albedo Surface Reconstruction from Stereo and Shape from Shading. Computer Vision and Pattern Recognition, 2000. Proceedings. Vol. 1, (2000) 480–487
6. A. Hertzmann, S. Seitz: Shape and Materials by Example: A Photometric Stereo Approach. Computer Vision and Pattern Recognition, 2003. Proceedings. Vol. 1, (2003) 533–540
7. K. E. Torrance, E. M. Sparrow: Theory for off-specular reflection from roughened surfaces. JOSA, 57(9) (1967) 1105–1114
8. R. Ramamoorthi, Pat Hanrahan: A Signal-Processing Framework for Inverse Rendering. Proceedings of the 28th annual conference on Computer graphics and interactive techniques, ACM Press, New York, NY, (2001) 117 – 128
9. Matlab Optimization Toolbox. http://www.mathworks.com/products/optimization/