Multi-view Stereo Reconstruction of Dense Shape and Complex Appearance

Hailin Jin[†*] Stefano Soatto[‡] Anthony J. Yezzi[§]

† Office of Technology, Adobe Systems Incorporated, 345 Park Avenue, San Jose, CA 95110

Tel: (408) 536-2762, email: hljin@adobe.com

‡ Computer Science Department, University of California, Los Angeles, CA 90095

Tel: (310) 825-4840, Fax: (310) 794-5056, email: soatto@cs.ucla.edu

§ School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, GA 30332

Tel: (404) 385-1017, Fax: (404) 894-4641, email: ayezzi@ece.gatech.edu

Keywords: image-based modeling and rendering, multi-view stereo, non-Lambertian reflection, 3-D shape reconstruction, variational methods, level set methods, appearance models.

[*]This research was performed while Hailin Jin was with Computer Science Department, University of California at Los Angeles.

Abstract

We address the problem of estimating the three-dimensional shape and complex appearance of a scene from a calibrated set of views under fixed illumination. Our approach relies on a rank condition that must be satisfied when the scene exhibits "specular + diffuse" reflectance characteristics. This constraint is used to define a cost functional for the discrepancy between the measured images and those generated by the estimate of the scene, rather than attempting to match image-to-image directly. Minimizing such a functional yields the optimal estimate of the shape of the scene, represented by a dense surface, as well as its radiance, represented by four functions defined on such a surface. These can be used to generate novel views that capture the non-Lambertian appearance of the scene.

1 Introduction

Multi-frame stereo consists of reconstructing the three-dimensional (3-D) shape of a scene from a collection of images taken from different vantage points. This is one of the classical problems of computer vision, where significant progress has been made in the last decade. In the early days of stereo, it was common to decompose the problem into two steps: establishing correspondence between points in different views, and then triangulating their position in space. Points in different images are said to be in correspondence when they are images of the same physical point in space via perspective projection. Once correspondence is established, the position of the points as well as the relative pose of the cameras can be determined using well-established procedures that are now the subject of textbooks [7, 10, 22].

Unfortunately the first step, establishing correspondence, is far less amenable to a clean and simple solution. First of all, point correspondence can only be reliably established for a very small subset of the scene. For instance, given a scene that contains a white wall, we cannot say which point on one image of the wall corresponds to in another image, since the local appearance is the same for every neighborhood of a point. Therefore, after establishing correspondence and reconstructing the 3-D position of relatively few "feature" points,[1] one would have to "densify" the reconstruction by filling in points that cannot be

[1] Even a few thousand feature points are far fewer than the millions of pixels in an image

matched from image to image.

Second, and more important, correspondence cannot be established by just comparing local image statistics unless the scene has the property that its appearance does not change with the viewpoint. Materials that exhibit this property are called Lambertian, or diffuse, and they include matte surfaces such as chalk, rough stone and certain fabrics[2]. However, most of the materials that populate our daily scenes such as plastic, polished stone, skin, glass, metal, etc. do not enjoy this property. Indeed, one can make an object that deviates severely from the ideal Lambertian model, for instance a car, appear arbitrarily different from image to image by changing the viewpoint and the illumination.

In this paper the first issue is addressed at the outset by modeling the shape of the scene as a collection of smooth surfaces: Like many recent works in multi-view stereo, we do not seek to establish correspondence among a sparse set of feature points and then fill in the rest. Rather, we start with a generic surface, say a large sphere or a smoothed cube, and evolve it, possibly via changes of topology, to best approximate the shape of the scene. We do so by numerically integrating systems of partial differential equations using the level set method. The second issue is addressed by bypassing the direct comparison of local image intensity, and instead comparing all images to the underlying model of the scene, which necessarily includes the current estimate of its shape *as well as its radiance.* Our model of the radiance is not in an explicit functional form; instead, it accounts for deviations from Lambertian reflection through a constraint on the rank of the radiance tensor field, which we will define shortly[3].

The result is an algorithm that takes as input a sequence of images of a scene with complex appearance, such as those in Figure 1 and, with no intermediate steps, returns an estimate of its shape, described by one or more "dense" surfaces, and an estimate of its appearance, described by the radiance tensor field. Such a description can be used to render the scene from novel viewpoints, assuming a static illumination, in ways that preserve the complex appearance of the original scene.

[2]Most feature correspondence algorithms implicitly assume that the scene is "almost" Lambertian, in the sense that the deviation from an ideal Lambertian model is small, not modeled explicitly, and instead lumped together with other factors as "noise."

[3]Incidentally, as shown in [28], the distinction of comparing all images to an underlying model, as opposed to matching image-to-image, is relevant only in the presence of non-Lambertian scenes, or other constraints on the diffuse albedo, as we will discuss shortly and as shown in [28].

Since a general scene cannot be reconstructed under varying and unknown illumination, we must make assumptions about the imaging process. Specifically, we assume that illumination is fixed but otherwise arbitrary, except for being "far enough" from the scene in a way that we will make precise in Section 2.1. Furthermore, we assume that the scene is a collection of smooth or piecewise smooth surfaces, and that its reflectance can be modeled by the linear combination of an ideal Lambertian component and a specular component, or what is known in computer graphics as a "diffuse + specular" reflectance model.

In the next subsection we will briefly review the state of the art as it relates to our contribution. Before we formally introduce the quantities at play in Section 2.1 we use the terms "photometry," "radiance," "reflectance" and "appearance" interchangeably, and similarly for "shape," "structure" and "geometry."

1.1 Relation to prior work

In order for any 3-D reconstruction to be possible, some assumption must be made on the photometry of the scene[4]. The most common assumption is that the light is fixed and the scene is *Lambertian*, i.e., the energy radiated from any point in the scene does not depend on the outgoing direction, so that correspondence can be easily established by comparing individual images. Indeed, as we have shown in [28], under the assumption of Lambertian reflection and in the absence of any additional information or constraint on the diffuse albedo, there is no difference between comparing all images to an underlying model of the scene as opposed to matching image-to-image directly. The situation is quite different, as we discuss in [28], when the scene deviates from ideal Lambertian reflection. In this case, reflection is most often described by an explicit model, a bidirectional reflectance distribution function (BRDF), chosen among a parametric class derived by physical or empirical considerations[5].

Most often, however, deviations from Lambertian reflection are modeled as "noise" or "outliers" and either minimized by choice of a suitable cost functional (such as photo-consistency [21]), or rejected

[4]It is straightforward to show that if a scene has arbitrary reflectance properties and one can change the light distribution from frame to frame, correspondence can*not* be established [32].

[5]For instance, [23] and [33] exploit the reciprocity condition of the BRDF to perform reconstruction using a particular imaging setup where multiple images are obtained by swapping a point light source and the camera. We do not impose constraints on the viewpoint, and do not restrict the illumination to be a point source. Indeed, we do not model illumination explicitly in our approach.

Figure 1: *Scenes with strong specularities are a challenge to algorithms relying on image-to-image matching.*

using robust statistical methods. For instance, one can select candidates for correspondence in each image by looking at image statistics integrated over a region, compute the cross-correlation or other score among putative correspondences, and then test whether they are consistent with a common epipolar geometry. This works well when the scene is composed mostly of matte surfaces with few specular highlights. However, for objects that are *shiny* and concentrated light distributions (see Figure1), this approach shows limitations. Alternatively, one can set up a global cost functional obtained by integrating on the entire scene a local consistency measure (e.g. normalized cross-correlation) computed on the images, and minimize it with respect to the unknown shape using variational techniques, an approach pioneered by [8] in stereo reconstruction. Our approach is based on a similar philosophy, and we also use level set methods [27] to numerically solve the variational problem. However, while Faugeras and Keriven estimate geometry alone, we estimate both geometry and photometry (radiance) and forego the Lambertian assumption that is latent in the cost functional used in [8]. [16] have modified the cost functional to minimize the effects of isolated specularities.

This work also relates to a series of works where the same computational framework is used in estimating the shape and radiance for scenes of increasing complexity: from constant diffuse albedo [31] to smooth diffuse albedo [18], to piecewise constant and piecewise smooth diffuse albedo [17], to arbitrary diffuse and constant specular albedo, to arbitrary diffuse and specular albedos.

In addressing non-Lambertian reflection, our work relates to several studies on specular reflections in stereo matching and reconstruction. [1] consider the likelihood of correct stereo matching by analyzing the relationship between stereo vergence and surface roughness, and propose a trinocular system where only two images are used at a time in the computation of depth at a point. [2, 3] excise specularities as a pre-processing step, similarly to [26], while [24] do so using polarized filters.

Non-Lambertian reflection has also been addressed in the context of photometric stereo, for instance by [13]. In this case, the viewpoint is fixed while the illumination changes. In this respect, this is quite different from our approach, that is more in line with traditional multi-view stereo in assuming that the viewpoint, and not the illumination, moves. Other approaches [11] compare the observed images with that of objects with known shape to obtain surface normals and hence shape.

This work also relates to the general problem of estimating reflectance properties as well as shape from sequences of images. For instance [32] use known shape to estimate global illumination [32]; in light field rendering [5, 9, 25] there is no explicit reconstruction of shape, and the radiance tensor, extended to the volume, is sampled directly. Indeed, the rank constraint on the radiance tensor field is often used in light field rendering, albeit not for inferring properties of the scene but, rather, for computational efficiency.

This work addresses the problem of multi-view stereo with fixed illumination and arbitrarily changing viewpoint. To the best of our knowledge, we are the first to propose a multi-view stereo algorithm that can provide an estimate of both dense shape and non-Lambertian reflection. Our algorithm is based on a constraint on the rank of the radiance tensor field (Section 2.1), which we show to imply (and hence be more general than) a "diffuse + specular" reflection models commonly used in Computer Graphics (Proposition 1).

2 Local modeling of radiance and image discrepancy

In this section we introduce the model of photometry, based on the *radiance tensor field*, and the measure of discrepancy between *model and images* that is the basis of our approach.

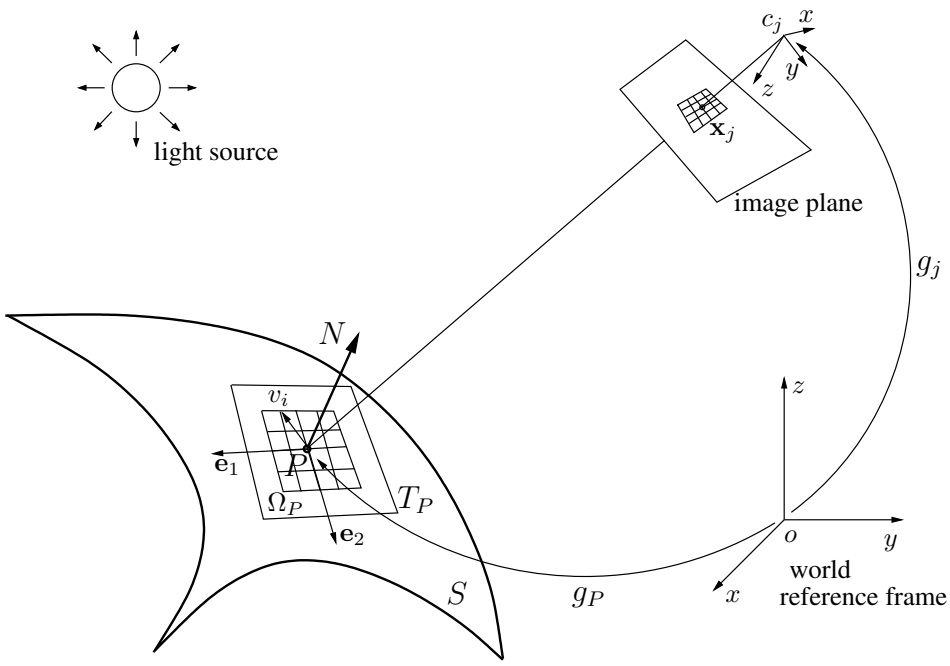

Figure 2: *The local coordinate frame on the tangent plane, the discretization of the local neighborhood, and the projection onto an image.*

2.1 The radiance tensor field

Let S be a (smooth) surface embedded in $\mathbb{R}^3$ and P be the generic point on S, with coordinates $\mathbf{X} = [X_1, X_2, X_3] \in \mathbb{R}^3$ with respect to a fixed world reference frame. We denote with $T_P S$ the tangent plane to the surface at the point P. The generic vector in the tangent plane (embedded in Euclidean space) has coordinates $v \in \mathbb{R}^3$. Let an ideal perspective camera be characterized by a Euclidean reference frame $g \in SE(3)$, that describes the change of coordinates between the world reference frame and the frame attached to the optical center of the camera, represented by a rotation matrix and a translation vector[6]. Therefore, if $\pi : \mathbb{R}^3 \longrightarrow \mathbb{R}^2$ denotes the canonical perspective projection[7], the point P projects onto each image in the coordinates $\mathbf{x} = \pi(gP)$.

For each point $P \in S$ we consider a discretization of a small neighborhood $\Omega_P \subset T_P S$ around it. This discretization is usually done with a tessellation of $T_P S$, which we represent via the vec-

[6] g acts on a point P with coordinates $\mathbf{X}$ via gP, which has coordinates $R\mathbf{X} + T$ where $R \in SO(3)$ is an orthonormal matrix with positive determinant and $T \in \mathbb{R}^3$. The push-forward action of g on vectors $v \in T\mathbb{R}^3$ with coordinates $\mathbf{V}$ is given by $g_* v$, which has coordinates $R\mathbf{V}$.

[7] $\pi(\mathbf{X}) = [X_1/X_3,\ X_2/X_3]$.

tors $v_1, v_2, \ldots, v_m$, where m is the number of points in Ω_P, as shown in Figure 2. We assume to be able to measure the amount of light leaving these points toward a discrete number n of camera poses, $g_1, g_2, \ldots, g_n$. Therefore, to each point P we can associate an array of $m \times n$ *ideal* measurements, one column for each camera view and one row for each point in Ω_P, as

$$R(P) = \begin{bmatrix} \rho(v_1, g_1) & \ldots & \rho(v_1, g_n) \\ \vdots & \ddots & \vdots \\ \rho(v_m, g_1) & \ldots & \rho(v_m, g_n) \end{bmatrix} \tag{1}$$

where $\rho(v_i, g_j)$ can be thought of as an approximate measurement of the radiance of the surface at a point. Notice that $R_{ij} \doteq \rho(v_i, g_j)$ relates to the *ideal* image I_j with an explicit dependence on P via the irradiance equation [12, page 208], assuming a pin-hole projection:

$$R_{ij} = I_j(\pi(g_j(P + v_i))) \quad \forall\, v_i \in \Omega_P \tag{2}$$

for all $j = 1, 2, \ldots, n$. The map $S \to \mathbb{R}^{m \times n};\ P \mapsto R(P)$ thus defines a tensor field on S, $R(\cdot)$ which, for any fixed P, is an $m \times n$ matrix, called the *radiance tensor*, or simply "radiance". In practice, the images I_j are measured only up to noise, so what is available is

$$\tilde{I}_j(\mathbf{x}) = I_j(\mathbf{x}) + w_j(\mathbf{x}); \quad \tilde{R}_{ij} = R_{ij} + w_{ij} \tag{3}$$

where $w_j(\mathbf{x})$ measures the discrepancy of the data from the model and can be considered as the realization of a random process (and therefore assumed to have a distribution associated to it), or simply as an unknown matrix whose norm we wish to minimize. We call $\tilde{R}$ the measured radiance tensor field obtained by substituting the noisy images $\tilde{I}$ in equation (2).

In general, the radiance tensor depends on the material properties of the surface and the lighting

conditions. For instance, for the simplest case of Lambertian reflection,

$$R(P) = R_1(P) \cdot \mathbf{1}_n^T \tag{4}$$

where $R_j(P)$ denotes the j-th column of $R(P)$ and $\mathbf{1}_n$ denotes an n-dimensional vector with all the elements equal to 1. It is because, by the Lambertian assumption, the radiance is independent of the viewpoint, and therefore all the columns of $R(P)$ are identical. In fact, we can replace $R_1(P)$ in equation (4) with any other column of $R(P)$. For more complex materials, $R(P)$ has more structure but is, in general, not arbitrary. Proposition 1 shows that for ideal surfaces that obey a "diffuse+specular" reflection model, the (point-wise) rank of the radiance tensor is two. In order to set up the notation to state the proposition, we choose a reference frame $\langle \mathbf{e}_1, \mathbf{e}_2 \rangle$ for the tangent plane $T_P S$ with the origin at P:$\langle \mathbf{e}_1, \mathbf{e}_1 \rangle = 1, \langle \mathbf{e}_2, \mathbf{e}_2 \rangle = 1, \langle \mathbf{e}_1, \mathbf{e}_2 \rangle = 0$. Let $N_P S$ be the outward unit normal to S at P, so that $\mathbf{e}_1 \times \mathbf{e}_2 = N_P S$. Then $\langle \mathbf{e}_1, \mathbf{e}_2, N \rangle$ forms a Euclidean reference frame for $\mathbb{R}^3$ around P, where we have indicated the normal vector with N as a short-hand for $N_P S$. We denote with g_P the change of coordinates between the world reference frame and $\langle \mathbf{e}_1, \mathbf{e}_2, N \rangle$ (see Figure 2). We can parameterize each unit vector λ in the upper hemisphere at P, H_P^2, with polar coordinates $(\theta_\lambda, \phi_\lambda) \in [0, \pi/2] \times [0, 2\pi]$, i.e., θ_λ is the angle between λ and N and ϕ_λ is the angle between λ and $\mathbf{e}_1$, for all $\lambda \in H_P^2$.

The interaction of light with the surface S can be expressed, for most materials that we are going to deal with, by the *bidirectional reflectance distribution function* (BRDF[8]). This is a function of two directions in H_P^2, the incident direction λ_i, parameterized by (θ_i, ϕ_i) and the reflected direction λ_o, parameterized by (θ_o, ϕ_o), as well as the wavelength and polarization of the incident radiation, which we will ignore (see Figure 3). Ward's (anisotropic) elliptical Gaussian model [30] approximates the BRDF β with a linear combination of a diffuse term and a specular term:

$$\beta(\theta_i, \phi_i, \theta_o, \phi_o) = \frac{\rho_d}{\pi} + \frac{\rho_s \exp(-\tan^2 \delta(\cos^2 \gamma/\alpha_x^2 + \sin^2 \gamma/\alpha_y^2))}{4\pi\alpha_x\alpha_y\sqrt{\cos\theta_i \cos\theta_o}} \tag{5}$$

[8]The BRDF is a simplified description of the radiometry of purely reflective (ideal) materials that yields an approximation of the radiance commonly used in computer graphics. It measures the ratio between the reflected energy along the direction (θ_o, ϕ_o) due to the energy coming from the direction (θ_i, ϕ_i) and the incoming energy.

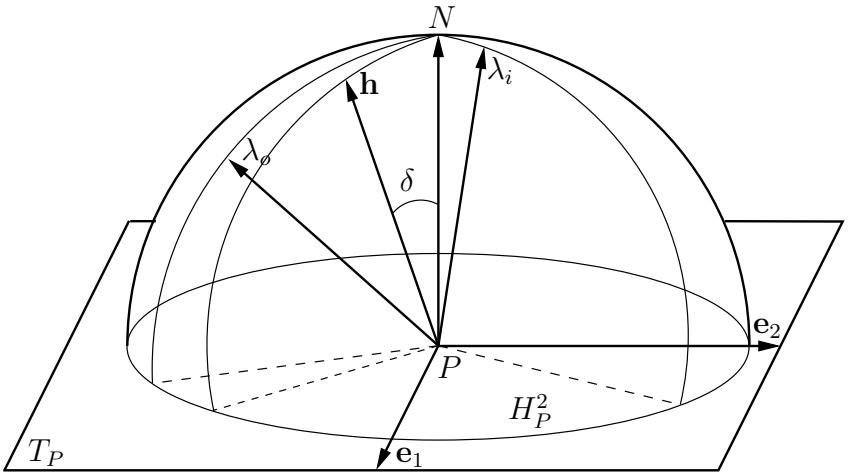

Figure 3: *Illustration of the light interaction with the surface.* H_P^2 *is the unit hemisphere at a point* P. $\lambda_i \in H_P^2$ *is the incoming light direction and* $\lambda_o \in H_P^2$ *is the outgoing light direction.* $\mathbf{h} \in H_P^2$ *is the halfway vector between* λ_i *and* λ_o. δ *is the angle between* $\mathbf{h}$ *and* N. *Surface reflectance is described by the bidirectional reflectance distribution function* β *that measures the ratio between the reflected energy along the direction* λ_o *due to the energy coming from the direction* λ_i *and the incoming energy.*

where ρ_d is the diffuse reflectance coefficient (commonly referred to as the *albedo*) and ρ_s is the specular reflectance coefficient; α_x and α_y are the standard deviations of the microscopic surface slope (surface roughness) in the direction of $\mathbf{e}_1$ and $\mathbf{e}_2$ respectively. The roughness coefficients are related to the properties of the material and we will consider them to be constant in a neighborhood of P. We note that constant surface roughness in a neighborhood does not imply that either diffuse reflectance coefficient (albedo) or specular reflectance coefficient is constant in that neighborhood. Let $\mathbf{h}$ be the halfway vector between the directions λ_i and λ_o: $\mathbf{h} \doteq \frac{\lambda_i+\lambda_o}{\|\lambda_i+\lambda_o\|}$; (δ, γ) are the polar coordinates for $\mathbf{h}$ and are therefore functions of $(\theta_i, \phi_i, \theta_o, \phi_o)$. The radiance in the direction determined by the point $\mathbf{x}_j = \pi(g_j P)$ in the j-th camera view is given by integrating the BRDF against the light distribution L in all directions (θ_i, ϕ_i):

$$\rho(0, g_j) = \int_0^{2\pi}\int_0^{\pi/2} \beta(\theta_i, \phi_i, \theta_o, \phi_o) L(\theta_i, \phi_i) \cos\theta_i \sin\theta_i d\theta_i d\phi_i \tag{6}$$

where the direction from P to c_j, the j-th camera center, in the frame of the point P, i.e., $g_{P_*}^{-1}\left(\frac{c_j - P}{\|c_j - P\|}\right)$ (see Footnote 6), is represented in polar coordinates as (θ_o, ϕ_o).

Proposition 1 (radiance tensor rank)**.** *Let* S *be made of a material that obeys the reflectance model* (5).

Furthermore, consider a surface patch $\Omega_P \subset T_P S$ that is small compared to the distance of P from the light sources and from the cameras. Then, if $R(P)$ is computed for $v_i \in \Omega_P$ as in equation (1)*, we have that* $\forall P \in S$

$$\textit{rank}(R(P)) \leq 2. \tag{7}$$

Proof. To facilitate computing the radiance $\rho(v_i, g_j)$ for each $v_i \in \Omega_P \subset T_P S; i = 1, 2, \ldots, m$, in the direction of the origin of the reference frame of camera $j = 1, 2, \ldots, n$, we will denote with $\tilde{g}_j(v_i)$ the direction $g_{(P+v_i)_*}^{-1}\left(\frac{c_j-(P+v_i)}{\|c_j-(P+v_i)\|}\right)$ from $P + v_i$ to c_j in the frame at the point $P + v_i : \langle \mathbf{e}_1(v_i), \mathbf{e}_2(v_i), N\rangle$. Since $T_P S$ is a plane, we can choose $\langle \mathbf{e}_1(v_i), \mathbf{e}_2(v_i), N\rangle$ to coincide with the reference frame at P : $\langle \mathbf{e}_1, \mathbf{e}_2, N\rangle$. Under the assumption that Ω_P is small, we can approximate $\tilde{g}_j(v_i)$ with $\tilde{g}_j(0)$[9]. Again, (θ_o, ϕ_o) are the polar coordinates of $\tilde{g}_j(0)$. Under the same assumption, we can also approximate the incoming light distribution at the point $P + v_i$ with $L(\theta_i, \phi_i)$. If we denote with $\rho(v_i|w)$ the radiance of point v_i along the direction λ, by equation (6), the radiance in the direction toward c_j is given by

$$\begin{aligned}
& \rho(v_i, g_j) = \rho\left(v_i|\tilde{g}_j(v_i)\right) \approxeq \rho\left(v_i|\tilde{g}_j(0)\right) \\
= & \int \beta(v_i, \theta_i, \phi_i, \theta_o, \phi_o) L(\theta_i, \phi_i) \cos\theta_i \sin\theta_i d\theta_i d\phi_i \\
= & \int_0^{2\pi}\int_0^{\pi/2} \frac{\rho_d(v_i)}{\pi} L(\theta_i, \phi_i) \cos\theta_i \sin\theta_i d\theta_i d\phi_i \\
& + \int \frac{\rho_s(v_i) \exp(-\tan^2\delta(\cos^2\gamma/\alpha_x^2 + \sin^2\gamma/\alpha_y^2))}{4\pi\alpha_x\alpha_y\sqrt{\cos\theta_i \cos\theta_o}} L(\theta_i, \phi_i) \cos\theta_i \sin\theta_i d\theta_i d\phi_i \\
= & \rho_d(v_i) s_0 + \rho_s(v_i) s_1(g_j)
\end{aligned}$$

where

$$\begin{aligned}
s_0 &\doteq \int_0^{2\pi}\int_0^{\pi/2} \frac{1}{\pi} L(\theta_i, \phi_i) \cos\theta_i \sin\theta_i d\theta_i d\phi_i \\
s_1(g_j) &\doteq s_1(\theta_o, \phi_o) = \int_0^{2\pi}\int_0^{\pi/2} \frac{\exp(-\tan^2\delta(\cos^2\gamma/\alpha_x^2 + \sin^2\gamma/\alpha_y^2))}{4\pi\alpha_x\alpha_y\sqrt{\cos\theta_i \cos\theta_o}} L(\theta_i, \phi_i) \cos\theta_i \sin\theta_i d\theta_i d\phi_i.
\end{aligned}$$

[9]The meaning of approximation goes as follows: $\forall \epsilon > 0$, we can choose the size of Ω_P small such that $\|\tilde{g}_j(v_i) - \tilde{g}_j(0)\| < \epsilon$.

This concludes the proof. □

Remark 1. *ρ_d, s_0, ρ_s and s_1 are all functions of P. We do not assume that either the albedo or the light distribution is the same for every P, but only that the surface roughness is locally constant in Ω_P.*

Remark 2. *Using $\tilde{g}_j(v_i)$ to approximate $\tilde{g}_j(v_0)$ is equivalent to using a scaled orthographic projection for the imaging model in Ω_P. However, when P moves over the surface, the parameters for the scaled orthographic projection are allowed to change. Therefore, we are* not *enforcing a scaled orthographic projection for the entire scene. The imaging model is still the perspective projection we put up at the beginning. In other words, we do* not *assume the overall size of the scene is small with respect to the distances to light sources or cameras.*

The intuition behind this proposition is that, in the limit where the light sources are far, and the patch Ω_P is small, the specular component of the radiance of Ω_P is modulated by a scalar function that depends on the viewpoint [10]. Of course, these conditions are a mathematical idealization. In practice, we verify experimentally that the singular values of $R(P)$ decrease sharply and are negligible beyond the second. In the experimental section (Section 4) we will report how the size of the neighborhood affects the performance of the algorithm and we will also discuss the range of applicability of this constraint on realistic imaging conditions.

Regardless of the actual numerical rank for $R(P)$, a limitation on the rank can be exploited to set up a discrepancy function for stereo reconstruction, as we do in the next section. In view of the claim above, one can then express the radiance tensor as the sum of two rank-one matrices with certain orthogonal properties. The relevance of Proposition 1 will be further discussed in Section 5.

Corollary 1 (local radiance model)**.** *At each point P of an ideal surface S that obeys the conditions of Proposition 1, the radiance tensor field can be represented with four vectors $d_1(v), d_2(v) \in \mathbb{R}^m$ and $s_1(g), s_2(g) \in \mathbb{R}^n$ as:*

$$R(P) = d_1(v)s_1^T(g) + d_2(v)s_2^T(g), \tag{8}$$

[10] Also note that the limit where the area of Ω_P goes to zero does not cause the rank to go to zero because the matrix $R(P)$ becomes smaller, since one can resize the tessellation of the tangent plane so as to keep the number of rows of $R(P)$ constant.

such that

$$\langle d_1(v), d_2(v) \rangle = 0 \quad \textit{and} \quad \langle s_1(g), s_2(g) \rangle = 0. \tag{9}$$

As we have pointed out in Remark 1, the reader should notice that $d_1(v)$, $d_2(v)$, $s_1(g)$ and $s_2(g)$ are all functions of the point P on the surface. Note that s_1 and s_2 depend on the viewing directions, a necessary element in modeling non-Lambertian reflection.

2.2 A discrepancy measure for non-Lambertian scenes

Naturally, due to image noise and deviation from the "diffuse+specular" reflectance model, the *measured* tensor $\tilde{R}(P)$ has rank greater than 2, most often full. The key idea here is to use this rank discrepancy to set up a matching criterion for stereo reconstruction. This is done by setting up an error function between the measured radiance tensor $\tilde{R}(P)$ and the model $R(P)$ at each point P (see equation (3)):

$$\Phi(P) \doteq \|\tilde{R}(P) - d_1(v)s_1^T(g) - d_2(v)s_2^T(g)\|_F^2 \tag{10}$$

where we have chosen the squared Frobenius norm to compare radiance tensors. Clearly $\Phi(P)$ will depend on the coordinates of P. In addition, $\Phi(P)$ will also depend on the normal at P, since v_i lives in $T_P S$: $\Phi(P) = \Phi(\mathbf{X}, N)$. If we define

$$\phi_{ij} = \tilde{R}_{ij} - d_1(v_i)s_1^T(g_j) - d_2(v_i)s_2^T(g_j), \tag{11}$$

where $\tilde{R}_{ij}$ is the (i,j)-th element of $\tilde{R}(P)$, $d_k(v_i)$ and $s_k(g_j)$ are the i-th and j-th components of $d_k(v)$ and $s_k(g)$ respectively for $k = 1, 2$, then the squared Frobenius norm is the sum of the square of each element ϕ_{ij}. The surface S can then be found as the minimizer of the energy $E \doteq \int_S \Phi(P) dA$:

$$\hat{S} \doteq \arg\min_S \int_S \Phi(P) dA \tag{12}$$

where dA is the area measure on S.

As we have noted, since the actual measured tensor $\tilde{R}$ will in general have full rank, we can write it, for each P, using the singular value decomposition (SVD) as

$$\tilde{R}(P) = \sum_{i=1}^{r} \tilde{d}_i(v)\tilde{s}_i^T(g) \tag{13}$$

where r is the rank of $\tilde{R}(P)$[11]. The singular values are sorted in a decreasing order with respect to k. Since, from the rank constraint of Proposition 1, we can choose the basis of R arbitrarily, we can have

$$d_i(v) = \tilde{d}_k(v), \quad \text{and} \quad s_k(g) = \tilde{s}_k(g), k = 1, 2 \tag{14}$$

and $R(P) = \tilde{d}_1(v)\tilde{s}_1^T(g) + \tilde{d}_2(v)\tilde{s}_2^T(g)$. The function Φ can therefore be written as

$$\Phi(P) = \|\tilde{d}_3(v)\tilde{s}_3^T(g) + \tilde{d}_4(v)\tilde{s}_4^T(g) + \cdots + \tilde{d}_r(v)\tilde{s}_r^T(g)\|_F^2. \tag{15}$$

By the properties of the SVD, we have that

$$\langle \tilde{d}_i(v), \tilde{d}_j(v) \rangle = \|\tilde{d}_i(v)\|^2\delta_{ij} \quad \text{and} \quad \langle \tilde{s}_i(g), \tilde{s}_j(g) \rangle = \|\tilde{s}_i(g)\|^2\delta_{ij} \tag{16}$$

where δ_{ij} is the Kronecker delta function, i.e., $\delta_{ij} = 1$, if $i = j$; $\delta_{ij} = 0$, otherwise. This is consistent with Corollary 1.

3 Estimation of shape and radiance for non-Lambertian scenes

In this section we present our algorithm to recover the representation of shape and radiance described in the previous section from a collection of images.

[11]The usual SVD yields unit-norm vectors $\tilde{d}_i(v), \tilde{s}_i(g)$ and additional singular values σ_i. In this paper, what we are really interested is the fixed rank approximation of $\tilde{R}(P)$ via SVD. Therefore, once SVD is computed, one can lump σ_i into either $\tilde{d}_i(v)$ or $\tilde{s}_i(g)$ or even divide $\tilde{d}_i(v)$ and multiply $\tilde{s}_i(g)$ by some constant simultaneously without changing the decomposition, since $\tilde{d}_i(v)$ and $\tilde{s}_i(g)$ appear together in a product in the decomposition.

3.1 Shape estimation

Shape, in our context, is described by a representation of the surface S relative to *any* Euclidean reference frame. When S is represented explicitly, one can look for the solution $\hat{S}$ via a local descent along the gradient of the cost (12). The analysis for this type of cost functional, which has $\mathbf{X}$ and N in the integrand, was first done by Faugeras and Keriven in [8] and can be found in [19] (in French). In particular, Faugeras and Keriven derived the Euler-Lagrange equations for the cost functional and then designed a flow based on it to find the optimal shape. However, in their derivation it is not immediate to see whether the resulting flow minimizes the cost functional. In his Ph.D. thesis [14] and [15], Jin et. al. presented an alternative proof which directly minimizes the cost functional and showed that the flow considered by Faugeras and Keriven in [8] is indeed the gradient descent for the cost (12). In this paper, we will present the optimality result and refer the interested reader to [19, 14, 15] for details.

Theorem 1 (optimality condition)**.** *Let $\Phi_{\mathbf{X}}, \Phi_N$ be the first-order derivatives of Φ with respect to $\mathbf{X}$ and N and $\Phi_{\mathbf{X}N}, \Phi_{NN}$ be the second-order derivatives. We assume that Φ_{NN} can be decomposed as: $\Phi_{NN} = \sum_{i=1}^{k} \lambda_i p_i p_i^T$ where $\lambda_i \in \mathbb{R}$ and $p_i \in \mathbb{R}^3$ (note that this decomposition is always possible since Φ_{NN} is real and symmetric). We have that the following partial differential equation is the gradient descent flow for the cost* (12)*:*

$$S_t = \Big(2H\Phi - \langle \Phi_{\mathbf{X}}, N\rangle - 2H\langle \Phi_N, N\rangle - \mathit{trace}(\Phi_{\mathbf{X}N}) + N^T \Phi_{\mathbf{X}N} N + \sum_{i=1}^{k} \lambda_i \mathbf{II}(P_N^{\perp} p_i)\Big) N. \qquad (17)$$

where $P_N^{\perp}$ is the projection from $\mathbb{R}^3$ to $T_P(S)$, i.e., $P_N^{\perp} = I - NN^T$, H is the mean curvature and $\mathbf{II}(v)$ is the second fundamental form of a vector $v \in T_P(S)$.

Note that equation (17) involves second-order derivatives: $\Phi_{\mathbf{X}N}$ and Φ_{NN} and no higher-order derivatives. This should not be surprising because the cost functional involves N in the integrand, which is the first-order variation of the surface S. In practice the following flow based on the first-order derivatives in equation (17) yields similar results to that of the flow (17), while saving a significant amount of

computations.

$$S_t = \big(2H\Phi - \langle \Phi_{\mathbf{X}}, N\rangle - 2H\,\langle \Phi_N, N\rangle\,\big)N. \tag{18}$$

The calculation of the flow above reveals some interesting structure, as major simplification occur after equations (16).

We will prove a stronger result than needed to compute the flow (18). In particular, we will show that even if the modeled rank of $R(P)$ is higher than 2, the resulting flow still takes a simple expression. Let r be the measured rank the radiance tensor $\tilde{R}(P)$. Suppose that the ideal $R(P)$ satisfies a rank constraint of r_0 and thus we take r_0 terms from the SVD of $\tilde{R}(P)$ (equation (13)). Therefore, the function Φ takes the expression

$$\Phi(P) = \|\tilde{d}_{r_0+1}(v)\tilde{s}^T_{r_0+1}(g) + \tilde{d}_{r_0+2}(v)\tilde{s}^T_{r_0+2}(g) + \cdots + \tilde{d}_r(v)\tilde{s}^T_r(g)\|^2_F. \tag{19}$$

Let ϕ_{ij} be the (i,j)-th element of $\Phi(P)$.

Theorem 2 (differentiation of the score)**.** *Let ξ indicate the arguments of Φ, i.e., ξ is one of X_1, X_2, X_3, N_1, N_2, N_3. Then*

$$\dot{\Phi} = \sum_{i,j=1}^{m,n} 2\phi_{ij}\dot{\tilde{R}}_{ij} \tag{20}$$

where the dot indicates differentiation with respect to ξ.

Proof. We define

$$\phi^i \doteq \tilde{R}^i - \sum_{k=1}^{r_0} \tilde{d}_k(v_i)\tilde{s}_k(g), \quad i = 1, 2, \ldots, m, \tag{21}$$

$$\phi_j \doteq \tilde{R}_j - \sum_{k=1}^{r_0} \tilde{d}_k(v)\tilde{s}_k(g_j), \quad j = 1, 2, \ldots, n, \tag{22}$$

where $\tilde{R}^i$ is the i-th row of $\tilde{R}$ and $\tilde{R}_j$ is the j-th column of $\tilde{R}$, i.e., $\tilde{R}^i = [\tilde{R}_{i1}, \tilde{R}_{i2}, \ldots, \tilde{R}_{in}]^T$ and

$\tilde{R}_j = [\tilde{R}_{1j}, \tilde{R}_{2j}, \ldots, \tilde{R}_{mj}]^T$. Expanding the derivative we get

$$\begin{aligned}\dot{\Phi} &= \sum_{i,j=1}^{n,m} \dot{\phi}_{ij}^2 = \sum_{i,j=1}^{n,m} 2\phi_{ij}\Big(\dot{\tilde{R}}_{ij} - \sum_{k=1}^{r_0} \dot{\tilde{d}}_k(v_i)\tilde{s}_k(g_j) - \sum_{k=1}^{r_0} \tilde{d}_k(v_i)\dot{\tilde{s}}_k(g_j)\Big) \\ &= \sum_{i,j=1}^{n,m} 2\phi_{ij}\dot{\tilde{R}}_{ij} - \sum_{i=1}^{n}\left\langle \phi^i, \sum_{k=1}^{r_0} \dot{\tilde{d}}_k(v_i)\tilde{s}_k(g)\right\rangle - \sum_{j=1}^{m}\left\langle \phi_j, \sum_{k=1}^{r_0} \dot{\tilde{s}}_k(g_j)\tilde{d}_k(v)\right\rangle.\end{aligned}$$

However, from equations (21) and (22) we see that ϕ^i is in the span of $\tilde{s}_{r_0+1}(g), \tilde{s}_{r_0+2}(g), \ldots, \tilde{s}_r(g)$ and ϕ_j is in the span of $\tilde{d}_{r_0+1}(v), \tilde{d}_{r_0+2}(v), \ldots, \tilde{d}_r(v)$. Therefore, from equations (16), we can conclude the proof by noting that the only term that contributes to the derivative is $\sum_{i,j=1}^{n,m} 2\phi_{ij}\dot{\tilde{R}}_{ij}$. □

As a consequence of the previous result, flow (18) for arbitrary rank in an explicit form read as:

$$S_t = \left(2H\Phi - \sum_{i,j=1}^{m,n} 2\phi_{ij}\left\langle \frac{\partial \tilde{R}_{ij}}{\partial \mathbf{X}} + 2H\frac{\partial \tilde{R}_{ij}}{\partial N}, N\right\rangle\right)N. \tag{23}$$

We implement the flow (23) using level set methods [27]. Naturally, as with most of these variational techniques, one can only hope to achieve convergence to a local extremum of the original cost functional, since the flow is based on the gradient descent principle, and existence and uniqueness results are not available for this class of flows. In the experimental section we will give empirical validation to this approach by testing the flow above on real image data starting from generic initial conditions.

3.2 Radiance estimation

Once the surface $\hat{S}$ has been found, one can use the representation of the radiance to generate images by "radiance-mapping" the tensor $R(P)$ onto the surface S. Naturally, the visualization of S in this case is view-dependent, since different columns of $R(P)$ contribute to the image of the same point P depending on the viewpoint g_i.

The radiance map is provided by the functions $d_1(v)$, $d_2(v)$ $s_1(g)$ and $s_2(g)$, estimated at each point of the surface, P, using the singular value decomposition of the measured radiance tensor $\tilde{R}$, according to Corollary 1 and equation (14). Given a novel vantage point g', the corresponding functions $s_1(g')$ and

$s_2(g')$ can be interpolated from the existing $s_1(g_j)$ and $s_2(g_j)$. One simple way of doing so is to find the three views closest to g', and then to use linear interpolation to obtain $s_i(g')$ from $s_i(g_j)$, for $i = 1, 2$. This technique also allows extrapolating the radiance; as we show in the experimental section, one can notice artifacts when comparing the results to actual images obtained from a novel viewpoint. However, such artifacts are only noticeable by direct comparison. Notice that $d_1(v)$ and $d_2(v)$ do not depend on the viewpoint, and therefore do not need to be interpolated. Since $s_1(g')$ and $s_2(g')$ are linearly interpolated from $s_1(g_j)$ and $s_2(g_j)$, this new radiance component does not increase the rank of the radiance tensor and therefore is consistent with the rank constraint (Proposition 1).

Notice that the images generated from the radiance map are significantly different than those generated by "texture mapping" the images $\tilde{I}$ onto the surface S. In fact, the functions $s_1(g)$ and $s_2(g)$ depend directly on the viewpoint, and therefore when the viewpoint moves, the highlights move on the estimated surface, giving an overall result that is visually comparable with image-based rendering techniques that assume true surface shapes [9, 5].

4 Experiments

In this section we report the experimental results of our algorithm tested on three datasets: "Van Gogh", "Buddha" and "elephant". The first two datasets (shown in Figure 1) are courtesy of Jean-Yves Bouguet and Radek Grzeszczuk (Intel Corp.). The third dataset (shown in the top row of Figure 10) is courtesy of Daniel Wood (University of Washington). The Van Gogh statue is made of polished metal, and is highly specular. There is a total of 339 images in the dataset. Pseudo-ground truth has been generated by laser or shadow scanning followed by mesh polishing (Figure 4). Buddha is a synthetic scene. There is a total of 281 images in the dataset. Ground truth is available (Figure 7). In Figure 4 we show the estimates of shape produced by the algorithm described in Section 3.1, together with the estimates obtained with the algorithm of [16], both compared with pseudo ground truth. In both algorithms, the numerical grids we use are of size $128 \times 128 \times 128$. Our estimate is obviously not as crisp as the ground truth, but it does capture important details on the face. Figure 8 shows the evolution of the estimate of shape. In Figure 9

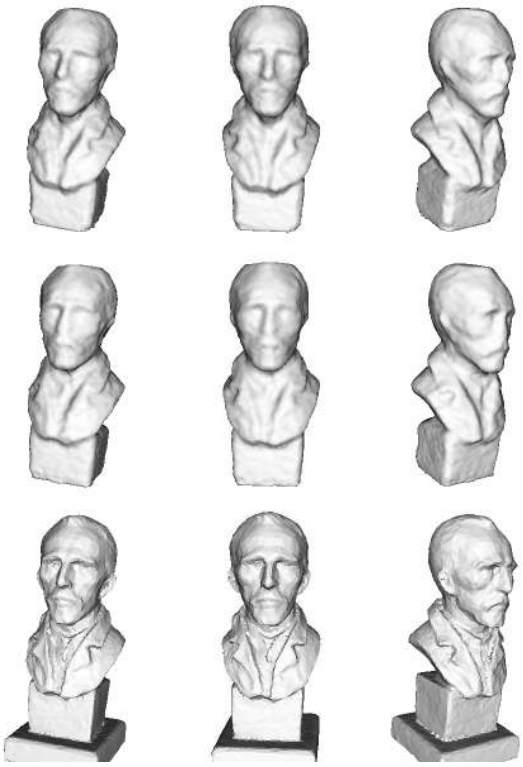

Figure 4: *Estimated shape (top), compared with pseudo-ground truth (bottom), obtained with a 3D laser scanner and mesh polishing. Our results improve those obtained with the algorithm of [16] (middle).*

we show synthetic images generated using the radiance map, as described in Section 3.2. Note that the specularities move with the viewpoint. In Figure 5 we show a few synthetic images compared with the real images from the same vantage point. In Figure 7 we show the estimated shape for the Buddha in Figure 1. The numerical grid size is $128 \times 128 \times 128$. In this case, ground truth is available since the images are synthetic. We also show the results obtained with the algorithm of [16]. In Figure 6 we show images synthesized from the model, compared with corresponding true images. In Figure 8 we show the evolution of shape, and in Figure 9 we show several novel views. In Figure 10 (top row) we show several views of an elephant made of polished marble. There is a total of 397 images in the dataset. The numerical grid size we used is again $128 \times 128 \times 128$. The estimated shape of our algorithms is reported in Figure 11 compared with pseudo-ground truth and that obtained using [16]. In the bottom row of Figure 10 we show images synthesized from the model, whose viewing positions and directions are the same as those in the top row.

Figure 5: *Synthetic images using the estimated radiance tensor (top) compared with the true images taken from the same vantage point. Note that one can actually* read *the text at the base of the bust. This is obtained from the radiance estimate,* not *from texture mapping.*

In Table 1 we summarize the shape error for different approximations of surface reflectance. The error is measured by the ratio between the volume of the symmetric difference between the estimated shape and the true shape (or the pseudo-ground truth) and the volume of the true shape (or of the pseudo-ground truth). Let ψ be the level set function for surface S. Suppose ψ is negative inside S and positive outside S. Then the volume contained by S, defined as S_{in}, can be measured as

$$\text{Vol}(S_{in}) = \int_{\mathbb{R}^3} (1 - \mathcal{H}(\psi)) dx dy dz \tag{24}$$

where $\mathcal{H}(x) : \mathbb{R} \to \{0, 1\}$ is the Heaviside function: $\mathcal{H}(x) = 1$ if $x \geq 0$; $\mathcal{H}(x) = 0$ otherwise. In practice, one can mollify the Heaviside function with a smooth approximation [4]. The volume of the

Reflectance models	Van Gogh	Buddha	elephant
Lambertian (the algorithm presented in [16])	6.9%	5.5%	24.3%
Rank-2 (the proposed algorithm)	5.7%	3.5%	7.3%
Rank-3 (a modified version of the proposed algorithm by using 3 SVD components in approximating $\tilde{R}(P)$ in equation (13))	5.6%	3.4%	7.2%

Table 1: *Shape error comparison chart for different reflectance approximations. The error is measured by the ratio between the volume of the symmetric difference between the estimated shape and the true shape (or the pseudo-ground truth) and the volume of the true shape (or the pseudo-ground truth). We observe that using the proposed rank-2 constraint for the radiance tensor, we can reduce the shape error by a factor of 2 in average, while using higher ranks does not improve the results much.*

symmetric difference between S_{in} and T_{in} can be calculated by:

$$\text{Vol}(S_{in} \Delta T_{in}) = \text{Vol}(S_{in}) + \text{Vol}(T_{in}) - 2\text{Vol}(S_{in} \cap T_{in}). \tag{25}$$

We observe that using the proposed rank-2 constraint for the radiance tensor, we can reduce the shape error by a factor of 2 in average, while using higher ranks does not improve the results as much. In Table 2 we show the degradation of the reconstruction as a function of the size of the patch Ω_P for the Van Gogh dataset. We tested neighborhood sizes from 3×3 to 19×19. We use odd sizes to have the neighborhoods symmetric around the center point. The unit of the neighborhood size is chosen to be corresponding to the actual pixel size in the best view, for instance 5×5 means that the projected neighborhood in the best view occupies an approximate 5×5 region in image pixels. We observe that the proposed algorithm is robust with respect to the neighborhood size in the sense that the reconstruction errors are almost the same from 7×7 to 15×15 neighborhoods. When the neighborhood is too small, the algorithm is sensitive to image noise and therefore has trouble converging. When the neighborhood is too large, the algorithm has trouble capturing sharp features present in the object shape.

Occlusions are handled by computing visibility at each step of the iteration. Therefore, the technique we present is computationally intensive and processing an entire dataset takes several hours. On the other hand, the algorithm requires no manual intervention, no intermediate step, no mesh polishing and no texture mapping after reconstruction. Therefore, its computational cost should be compared to

Size of Ω_P	3×3	5×5	7×7	9×9	11×11	13×13	15×15	17×17	19×19
Shape error	14.8%	7.6%	6.3%	6.0%	5.7%	5.8%	6.1%	6.4%	6.6%

Table 2: *Shape error comparison chart for different sizes of Ω_P for the Vangogh dataset. The error is measured by the ratio between the volume of the symmetric difference between the estimated shape and the pseudo-ground truth and the volume of the pseudo-ground truth. The unit of the neighborhood size is chosen to be corresponding to the actual pixel size in the best view, for instance 5×5 means that the projected neighborhood in the best view will occupy an approximate 5×5 region in image pixels.*

implementing the entire pipeline from images to rendering.

Figure 6: *Synthetic images obtained with the estimated radiance tensor field (top) compared with the true images taken from the same vantage point.*

5 Discussion

We have presented a novel algorithm for estimating dense shape and non-Lambertian photometry from a collection of images. Our algorithm relies on a constraint on the rank of the radiance tensor field, which is derived from the diffuse+specular reflection model commonly used in Computer Graphics, in the sense elucidated in Proposition 1. While one could dismiss the analysis and just introduce the cost function (10) point-blank without detracting from the algorithm proposed (which is validated experimentally), the proposition indicates precisely under what conditions the rank constraint is satisfied, i.e., what the

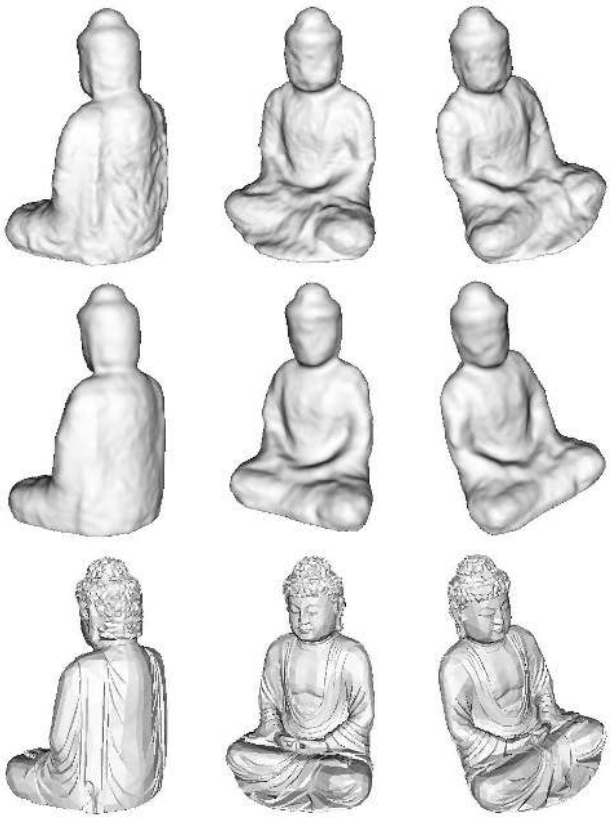

Figure 7: *Estimated shape (top), compared with ground truth (bottom), also compared with the results obtained by the algorithm of [16] (middle).*

underlying *mathematical model* is. Naturally, the closest the scene is to satisfying the assumptions (i.e. the closest it is to smooth shape, diffuse+specular reflection, fixed distant illumination) the smaller the rank $\tilde{R}(P)$ is. However, even though only ideal scenes viewed from noiseless images will satisfy the assumptions exactly, we can still exploit the discrepancy derived from the idealized model to define a constraint that can be used to reconstruct the scene from real images.

Those that object to the restrictiveness of the model laid out in Proposition 1 will be relieved to know that extension to higher ranks is conceptually and computationally trivial. One will need to take more terms from the SVD, but Theorem 1 assures that the gradient flow can be computed essentially in the same way. However, it can be verified experimentally that, for most scenes, an increase in the rank of the model does not yield a significant improvement in the reconstruction, further validating the mathematical model proposed (see Table 1). Indeed, it is possible to explore further reduction in the complexity of

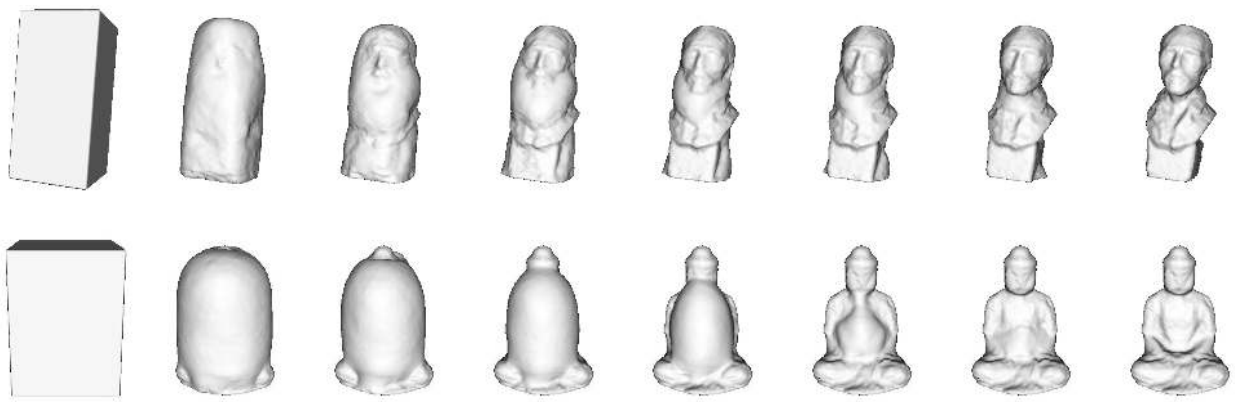

Figure 8: *Shape evolution for Van Gogh (top) and Buddha (bottom).*

the model to obtain more robust and computationally efficient algorithms which are the subject of our current investigation.

Our algorithm can handle sharp changes of the radiance profile: In Figure 5, one can actually *read* the text at the base of the bust from the reconstructed radiance. Note that there is no restriction whatsoever imposed on the variation of the diffuse and specular components of the radiance, and nowhere it is assumed that it be constant or smooth. What is assumed to be constant is the surface roughness, *not the albedo*, so we can handle heavily textured objects. On the other hand, our algorithm does not *require* strong texture or point features to be visible, and returns a dense estimate of shape, with no need to interpolate or triangulate a surface from sparse points.

Note also that, although the measured radiance tensor *at a given point* P is assembled using a local approximation of the surface with the tangent plane $T_P S$, this does not mean that our algorithm only works for planar surfaces: In fact, the radiance tensor at a nearby point Q is computed using the tangent plane $T_Q S$ that is not constrained to be similar to $T_P S$. If one thinks of $R(P)$ as a "signature" attached to $P \in S$, the model imposes no constraint that nearby points should have similar signatures.

Acknowledgements

This research is supported in part by NSF IIS-0208197/CCR-0133736, ONR N00014-02-1-0720, AFOSR F49620-03-1-0095 and Intel 8029. We thank Jean-Yves Bouguet, Radek Grzeszczuk and Daniel Wood

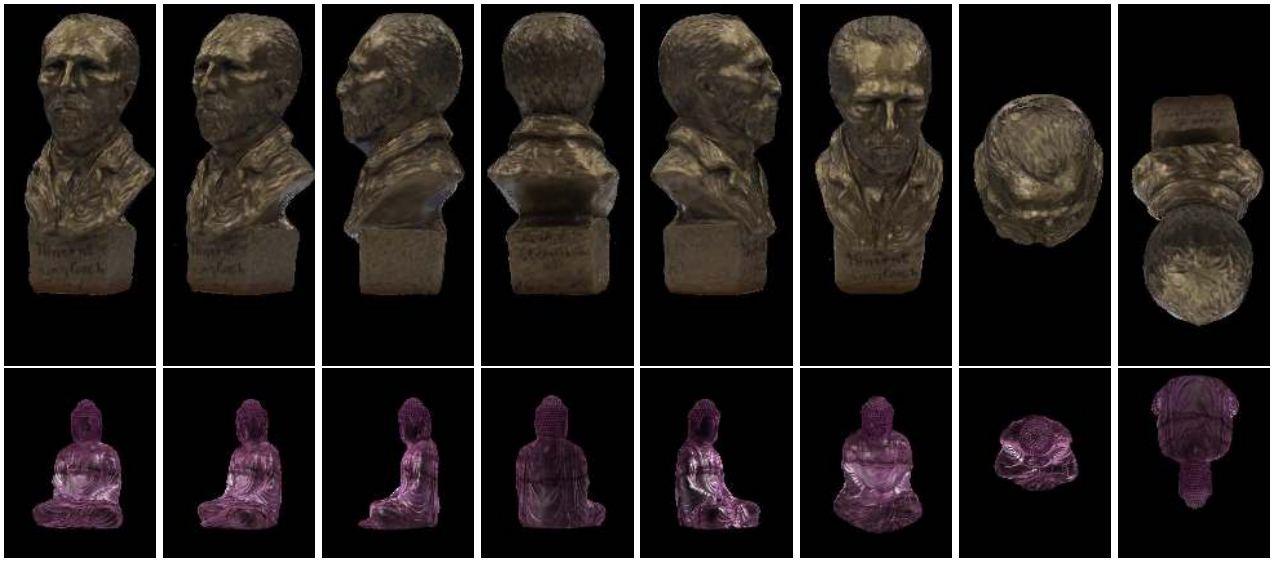

Figure 9: *Synthetic images obtained from the estimated radiance. As it can be seen, the appearance changes significantly with the vantage point.*

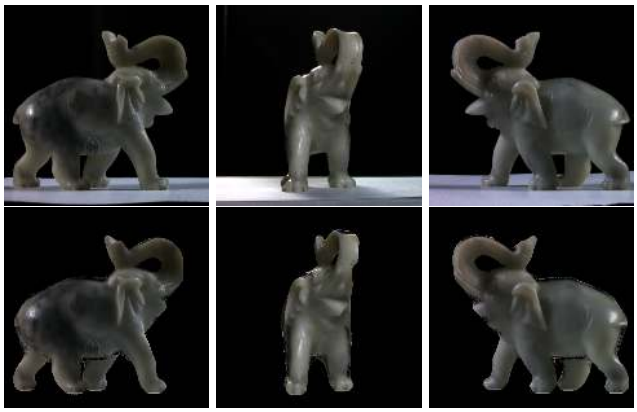

Figure 10: *Top row: three images from the elephant dataset (courtesy of Daniel Wood, University of Washington). Bottom row: synthetic views generated using the estimated radiance. The structure and position of specular highlights is correctly captured; there are some visualization artifacts at the boundaries, but note that even the text on the small label is visible on the left image. Note that this is an estimate of the radiance,* not *a texture map.*

for providing us testing data.

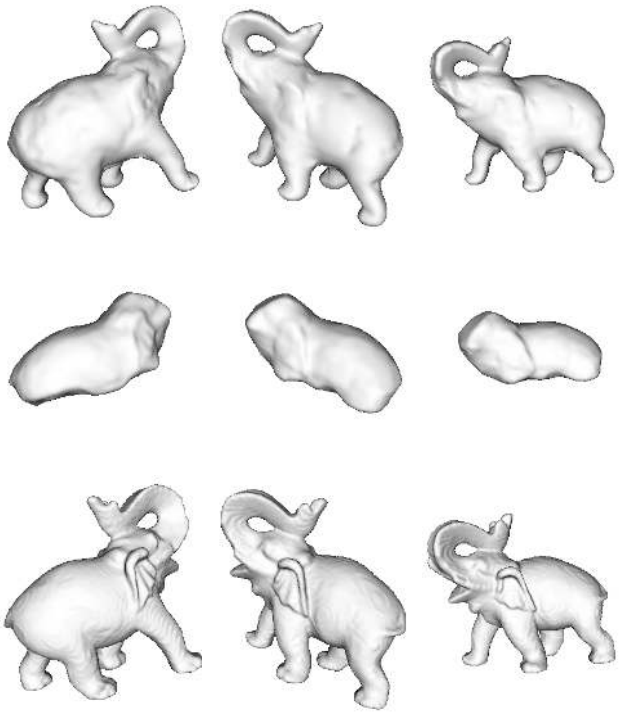

Figure 11: *Estimated shape (top row) of the scene in Figure 10, compared with pseudo-ground truth (bottom row), obtained with a 3D laser scanner and mesh polishing. Our results improve those obtained with [16] (middle row). The ear is not clear in the reconstruction, although it is well captured as radiance (Figure 10).*

References

[1] D. N. Bhat and S. K. Nayar. Stereo in the presence of specular reflection. In *Proc. Int. Conf. on Computer Vision*, pages 1086–1092, 1995.

[2] A. Blake. Specular stereo. In *Proc. Int. J. Conf. on Artificial Intell.*, pages 973–976, 1985.

[3] G. Brelstaff and A. Blake. Detecting specular reflections using Lambertian constraints. In *Proc. Int. Conf. on Computer Vision*, pages 297–302, 1988.

[4] T. F. Chan and L. A. Vese. Active contours without edges. *IEEE Trans. on Image Processing*, 10(2):266–277, February 2001.

[5] W. Chen, J.-Y. Bouguet, M. Chu, and R. Grzeszczuk. Light Field Mapping: Efficient Representation and Hardware Rendering of Surface Light Fields. In *Proc. ACM SIGGRAPH*, 2002.

[6] C. L. Epstein and M. Gage. The curve shortening flow. Wave motion: theory, modelling, and computation (Berkeley, Calif., 1986), *Math. Sci. Res. Inst. Publ.*, 7: 15–59, 1987.

[7] O. Faugeras. *Three-Dimensional Computer Vision.* MIT Press, 1993.

[8] O. Faugeras and R. Keriven. Variational principles, surface evolution, pde's, level set methods and the stereo problem. *IEEE Trans. on Image Processing*, 7(3):336–344, 1998.

[9] S. Gortler, R. Grzeszczuk, R. Szeliski, and M. Cohen. The lumigraph. In *Proc. ACM SIGGRAPH*, pages 43–54, 1996.

[10] R. Hartley and A. Zisserman. *Multiple view geometry in computer vision*. Cambridge University Press, 2000.

[11] A. Hertzmann and S. M. Seitz. Shape and Materials by Example: A Photometric Stereo Approach. *Proc. IEEE Conf on Comp Vision and Pattern Recogn.*, June 2003.

[12] B. K. P. Horn. *Robot vision*. MIT press, 1986.

[13] K. Ikeuchi. Determining surface orientations of specular surfaces by using the photometric stereo method. *IEEE Trans. on Pattern Analysis and Machine Intell.*, 3(6):661–669, 1981.

[14] H. Jin. Variational Methods for Shape Reconstruction in Computer Vision. Ph.D. thesis, Washington University, Saint Louis, Missouri, August 2003.

[15] H. Jin, S. Soatto and A. J. Yezzi. Multi-view Stereo Reconstruction of Dense Shape and Complex Appearance. *UCLA CAM Report 04-53*, Department of Mathematics, University of California at Los Angeles, September 2004.

[16] H. Jin, A. J. Yezzi and S. Soatto. Variational multiframe stereo in the presence of specular reflections. In *Proc. of the First Intl. Symp. on 3D Data Processing Visual. and Trans.*, June 2002.

[17] H. Jin, A. J. Yezzi and S. Soatto. Region-based Segmentation on Evolving Surfaces with Application to 3D Reconstruction of Shape and Piecewise Constant Radiance. In *Proc. of Euro. Conf. on Computer Vision.*, May 2004.

[18] H. Jin, A. J. Yezzi, Y.-H. Tsai, L.-T. Cheng and S. Soatto. Estimation of 3D Surface Shape and Smooth Radiance from 2D Images: A Level Set Approach. *J. of Scientific Computing*, 19(1-3): 267–292, December 2003.

[19] R. Keriven. Equations aux Dérivées Partielles, Evolutions de Courbes et de Surfaces et Espaces d'Echelle: Applications à la Vision par Ordinateur. Ph.D. thesis. ENPC, France, 1997.

[20] J. Kim, V. Kolmogorov and R. Zabih. Visual Correspondence Using Energy Minimization and Mutual Information. In *Proc. of Intl. Conf. on Computer Vision*, October 2003.

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

[22] Y. Ma, S. Soatto, J. Kosecka and S. Sastry, *An Invitation to 3-D Vision*. Springer-Verlag, 2003.

[23] S. Magda, T. Zickler, D. J. Kriegman, and P. N. Belhumeur. Beyond Lambert: Reconstructing Surfaces with Arbitrary BRDFs. In *Proc. Intl. Conf. on Computer Vision*, 2001.

[24] S. K. Nayar, X. S. Fang, and T. Boult. Removal of specularities using color and polarization. In *Proc. IEEE Conf. on Comp. Vision and Pattern Recogn.*, pages 585–590, 1993.

[25] K. Nishino, Y. Sato, and K. Ikeuchi. Eigen-texture method: appearance compression based on 3D model In *Proc. IEEE Conf on Comp Vision and Pattern Recogn.*, 1999.

[26] M. Okutomi and T. Kanade. A multiple baseline stereo. *IEEE Trans. on Pattern Analysis and Machine Intell.*, 15(4):353–363, 1993.

[27] S. J. Osher and J. A. Sethian. Fronts propagating with curvature-dependent speed: algorithms based on hamilton-jacobi equations. *J. of Comp. Physics*, 79:12–49, 1988.

[28] S. Soatto, A. J. Yezzi and H. Jin. Tales of Shape and Radiance in Multi-view Stereo. In *Proc. Intl. Conf. on Computer Vision*, pages 171–178, October 2003.

[29] Y. Tsin, S. B. Kang and R. Szeliski. Stereo Matching with Reflections and Translucency. *Proc. IEEE Conf. on Comp. Vis. and Patt. Recog*, June 2003.

[30] G. Ward. Measuring and modeling anisotropic reflection. In *Proc. ACM SIGGRAPH*, pages 265–272, 1992.

[31] A. J. Yezzi and S. Soatto. Stereoscopic Segmentation. *Intl. J. of Computer Vision*, 53(1):31-43, June 2003

[32] Y. Yu, P. Debevec, J. Malik, and T. Hawkins. Inverse global illumination: Recovering reflectance models of real scenes from photographs. In *Proc. ACM SIGGRAPH*, 1999.

[33] T. Zickler, P. N. Belhumeur, and D. J. Kriegman. Helmholtz Stereopsis: Exploiting Reciprocity for Surface Reconstruction. *Intl. J. of Computer Vision*, 49(2-3):215-227, 2002.