Multi-view Stereo Beyond Lambert*

Hailin Jin† Stefano Soatto§ Anthony J. Yezzi‡

† Electrical Engineering Department, Washington University, St. Louis - MO 63130, hljin@ee.wustl.edu
§ Computer Science Department, University of California, Los Angeles - CA 90095, soatto@ucla.edu
‡ Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta - GA 30332, ayezzi@ece.gatech.edu

Abstract

We consider the problem of estimating the shape and radiance of an object from a calibrated set of views under the assumption that the reflectance of the object is non-Lambertian. Unlike traditional stereo, we do not solve the correspondence problem by comparing image-to-image. Instead, we exploit a rank constraint on the radiance tensor field of the surface in space, and use it to define a discrepancy measure between each image and the underlying model. Our approach automatically returns an estimate of the radiance of the scene, along with its shape, represented by a dense surface. The former can be used to generate novel views that capture the non-Lambertian appearance of the scene.

Figure 1: *Scenes with strong specularities or made of translucent materials with no distinct point features are a challenge to most stereo algorithms.*

1 Introduction

Multi-frame stereo seeks to reconstruct the three-dimensional shape of a scene from a collection of images taken from different vantage points, and is one of the classical problems of computer vision. The task can be conceptually[1] decomposed into two steps: establishing *correspondence* between points in different views, and *triangulating* their positions in space.

The greatest challenge to most stereo algorithms comes from the first step[2]: successful correspondence depends on the material properties of objects in the scene as well as on the light distribution, which in real environments can be quite complex.

1.1 Relation to prior work

Establishing correspondence requires introducing assumptions on the photometric properties of the scene. This is done, overtly or covertly, in any stereo algorithm[3]. The most common assumption is that 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 the irradiance of individual images. Deviations from Lambertian reflection are often modeled as "noise" or "outliers" and either minimized by choice of cost functionals, or rejected using robust statistical methods. For instance, one can select candidates for correspondence in each image, compute the cross-correlation 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 Fig. 1), this approach fails.

The problem with this classical approach is that it attempts to establish correspondence *from image to image*. This requires that the irradiance profile of corresponding neighborhoods in different images be similar, i.e., that the surface be close to Lambertian, modulo noise and outliers. This requirement can be lifted if one were to use an explicit model of the photometry of the scene, and therefore establish correspondence *from model to images*.

*Work supported by NSF IIS-0208197/CCR-0133736, ONR N00014-02-1-0720, AFOSR F49620-03-1-0095 and Intel 8029. We thank J.-Y. Bouguet and R. Grzeszczuk for data.

[1] Some algorithms merge these two steps into one, and others integrate correspondence information over regions in the image, rather than individual points; we will review these methods in Sect. 1.1.

[2] Multiple view geometry, which addresses the triangulation step, is now well understood, and several textbooks are now available.

[3] 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 [18].

In addressing non-Lambertian reflection, this work relates to several studies on specular reflections in stereo matching and reconstruction. Bhat and Nayar [1] consider the likelihood of correct stereo matching by analyzing the relationship between stereo vergence and surface roughness, and also propose a trinocular system where only two images are used at a time in the computation of depth at a point. Brelstaff and Blake [2, 3] excise specularities as a pre-processing step; similar techniques are used also by Okutomi and Kanade [15], while Nayar et. al. [12] have considered using polarized filters to remove specularities. Ikeuchi formulates the reconstruction problem for specular surfaces in a photometric stereo setting [8], and [13, 19] estimate surface shape with arbitrary reflectance exploiting Helmholtz reciprocity. In space carving techniques [11] brightness constancy is extended to photometric consistency relative to a common model.

There is also a relation between our work and that of Faugeras and Keriven [5], who cast the traditional multi-frame stereo in a variational framework and use level set methods [16] to solve it. They address the correspondence problem by best approximating the brightness constancy assumption at local neighborhoods of the image[4], thus obtaining in effect a dense correspondence wherever the brightness gradient is non-zero. Jin et. al. [10] modify the cost functional to minimize the effects of isolated specularities.

This work also relates to the general problem of estimating reflectance properties as well as shape from sequences of images; for instance, Yu et. al. use known shape to estimate global illumination [18]; see also work on light field rendering, such as [4, 6, 14] and references therein.

None of the algorithms described, however, returns an estimate of both the shape and the (non-Lambertian) reflection of the scene.

1.2 Contributions of this paper

We address the problem of stereo reconstruction specifically for non-Lambertian objects; we work with a calibrated stereo rig, although extensions to uncalibrated cameras are conceptually straightforward. Unlike traditional stereo, we do not compare image to image, but instead compare each image directly to a model. Our model is not in an explicit functional form for the reflectance distribution function; instead, it is a *constraint on the rank of the radiance tensor field* (Sect. 2.1). We show that this model is implied by (and is therefore more general than) standard diffuse+specular reflection models commonly used in computer graphics (Prop. 1). In addition to robustness to deviations from Lambertian reflection that comes from having an explicit model, our approach enjoys additional benefits that are not found in traditional stereo algorithms. First, it automatically returns an estimate of the *radiance* profile of the scene, along with its *shape*. The former can be used to generate *view-dependent* "radiance maps" (as opposed to "texture maps") that can be used to synthesize novel images that preserve the shiny appearance of objects (Sect. 3.2), yielding results comparable to light field rendering (Fig. 4 and 5). Second, since we work in a variational framework, we estimate a *dense surface* directly, and therefore we do not need to interpolate or triangulate meshes; we integrate our cost functionals on entire regions of the image, and therefore we do not need photometrically distinct "feature points" to be present in the scene (Sect. 3.1). We validate our analysis and algorithms with experiments on real objects with complex reflectance properties (Sect. 4) and ground truth.

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.

2.1 The radiance tensor field

Let S be a (smooth) surface embedded in $\mathbb{R}^3$, $P \in S$ the generic point on it, with coordinates $\mathbf{X} = [X_1, X_2, X_3] \in \mathbb{R}^3$ with respect to an inertial reference frame. We denote with $T_P S$ the tangent plane to the surface at the point P. The generic vector on 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 inertial reference frame and the frame attached to the optical center of the camera, represented by a rotation matrix and a translation vector[5]. Therefore, if $\pi : \mathbb{R}^3 \longrightarrow \mathbb{R}^2$ denotes the canonical central projection[6], the point P projects onto each image in the coordinates $\mathbf{x} = \pi(gP)$.

Our measurements are obtained at a discrete number n of camera poses, $g_1, g_2, \ldots, g_n$, and at a discrete number m of pixels which we represent, for convenience, in a neighborhood of each point P, as the projection of a tessellation of the tangent plane $\Omega_P \subset T_P S$ via the vectors $v_1, v_2, \ldots, v_m$, as in Fig. 2. Therefore, for each point P, we can associate

[4] This is done by looking for corresponding patches that maximize a normalized cross-correlation score, the underlying assumption being that of brightness constancy of corresponding points modulo local contrast and scaling.

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

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

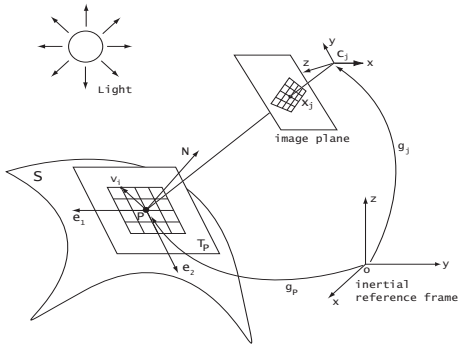

Figure 2: *The local coordinate frame on the tangent plane, its discretization, and the projection onto an image.*

an array of measurements, one column for each view and one row for each pixel in a neighborhood of $\pi(g_j P)$:

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

where $\rho(v_i, g_j)$ can be though of as an approximation of the radiance of the surface at a point. Notice that $R_{ij} \doteq \rho(v_i, g_j)$ depends on P via the irradiance equation ([7] page 208), assuming a pin-hole projection:

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

for all $j = 1, 2, \dots, n$. The map $S \to \mathbb{R}^{m \times n};\ P \mapsto R(P)$ 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

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

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 condition. For instance, for the simplest case of Lambertian reflection, $R(P)$ has rank one at every point since, by the Lambertian assumption, the radiance is independent of the viewpoint, and therefore all the columns of R are equal. For more complex materials, the rank of R will be greater than one in general but, in general, it will be less than full. Prop. 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 inertial reference frame and $\langle \mathbf{e}_1, \mathbf{e}_2, N \rangle$ (see Fig. 2). We can parameterize each unit vector λ in the upper half-sphere at P, H^2_P, with polar coordinates $(\theta_\lambda, \phi_\lambda) \in [0, \pi/2] \times [0, 2\pi]$. In other words, θ_λ is the angle between λ and N and ϕ_λ is the angle between λ and $\mathbf{e}_1$, for all $\lambda \in H^2_P$.

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[7]). This is a function of two directions in H^2_P, the incident direction, parameterized by (θ_i, ϕ_i) and the reflected direction, parameterized by (θ_o, ϕ_o), as well as the wavelength and polarization of the incident radiation, which we will ignore. Ward's (anisotropic) elliptical Gaussian model [17] approximates the BRDF β with a combination of a diffuse term and a specular term:

$$\begin{aligned} &\beta(\theta_i, \phi_i, \theta_o, \phi_o) = \\ &= \frac{\rho_d}{\pi} + \frac{\rho_s \exp[-\tan^2 \delta(\cos^2\gamma/\alpha_x^2 + \cos^2\gamma/\alpha_y^2)]}{4\pi\alpha_x\alpha_y\sqrt{\cos\theta_i \cos\theta_o}} \end{aligned} \quad (4)$$

where ρ_d is the diffuse reflectance coefficient 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. They are related to the properties of the material and we will consider them to be constant in a neighborhood of P. Let $\mathbf{h}$ be the half vector between the direction (θ_i, ϕ_i) and (θ_o, ϕ_o); δ is the angle between $\mathbf{h}$ and N, γ is the angle between $\mathbf{h}$ and $\mathbf{e}_1$; (δ, γ) are the polar coordinates for $\mathbf{h}$, and are therefore a function of $(\theta_i, \phi_i, \theta_o, \phi_o)$. The radiance in the direction determined by the point $\mathbf{x}_j$ is given by integrating the BRDF against the light distribution L in all directions (θ_i, ϕ_i):

$$\begin{aligned} \rho(0, g_j) \quad = \quad & \int_0^{2\pi}\int_0^{\pi/2} \beta(\theta_i, \phi_i, \theta_o, \phi_o) \cdot \\ & L(\theta_i, \phi_i) \cos\theta_i \sin\theta_i d\theta_i d\phi_i \end{aligned} \quad (5)$$

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 5), is represented in polar coordinates (θ_o, ϕ_o).

Proposition 1 (radiance tensor rank). *Let S be made of a material that obeys a reflection model (4). Furthermore, consider a surface patch $\Omega_P \subset T_P S$ that is small compared*

[7]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.

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*

$$\textit{rank}(R(P)) \leq 2 \quad \forall\, P \in S.$$

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 $\tilde{g}_j(0)$. 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 $\rho(v_i|\lambda)$ the radiance of point v_i along the direction λ, by equation (5), the radiance in the direction toward c_j is given by

$$\begin{aligned}
& \rho(v_i, g_j) = \rho\left(v_i|\tilde{g}_j(v_i)\right) \cong \rho\left(v_i|\tilde{g}_j(0)\right) \\
= & \int \beta(v_i, \theta_i, \phi_i, \theta_o, \phi_o) L(\theta_i, \phi_i) \cos\theta_i \sin\theta_i d\theta_i d\phi_i \\
= & \int_0^{2\pi}\!\!\int_0^{\pi/2} \frac{\rho_d(v_i)}{\pi} L(\theta_i, \phi_i) \cos\theta_i \sin\theta_i d\theta_i d\phi_i \\
+ & \int \frac{\rho_s(v_i) \exp[-\tan^2\delta(\cos^2\gamma/\alpha_x^2 + \cos^2\gamma/\alpha_y^2)]}{4\pi\alpha_x\alpha_y\sqrt{\cos\theta_i \cos\theta_o}} \cdot \\
& L(\theta_i, \phi_i) \cos\theta_i \sin\theta_i d\theta_i d\phi_i \\
= & \rho_d(v_i) s_1 + \rho_s(v_i) s_2(g_j)
\end{aligned}$$

where

$$\begin{aligned}
& s_1 \doteq \int_0^{2\pi}\!\!\int_0^{\pi/2} \frac{1}{\pi} L(\theta_i, \phi_i) \cos\theta_i \sin\theta_i d\theta_i d\phi_i \\
& s_2(g_j) \doteq s_2(\theta_o, \phi_o) = \\
& = \int_0^{2\pi}\!\!\int_0^{\pi/2} \frac{\exp[-\tan^2\delta(\cos^2\gamma/\alpha_x^2 + \cos^2\gamma/\alpha_y^2)]}{4\pi\alpha_x\alpha_y\sqrt{\cos\theta_i \cos\theta_o}} \cdot \\
& L(\theta_i, \phi_i) \cos\theta_i \sin\theta_i d\theta_i d\phi_i.
\end{aligned}$$

□

The intuition behind this proposition is that, in the limit where the light sources are far, and the patch Ω_P is small, the specularity is either absent, or it "washes out" the entire patch. Of course, these conditions are only a mathematical idealization, and are not verified in practice. Indeed, this very fact is exploited in the next section to set up a cost function for stereo reconstruction. In view of the claim above, one can then write the radiance tensor as the sum of two rank-one matrices. The relevance of Prop. 1 will be discussed in Sect. 5.

Corollary 1 (local radiance model). *At each point P of an ideal surface S that obeys the conditions of Prop. 1, the radiance tensor field is given by*

$$R(P) = d_1(v) s_1^T(g) + d_2(v) s_2^T(g) \tag{6}$$

where $d_i(v)$ stands for $[d_i(v_1), d_i(v_2) \ldots, d_i(v_m)]^T$, and $s_i(g)$ stands for $[s_i(g_1), s_i(g_2), \ldots, s_i(g_n)]^T$, $i = 1, 2$.

The reader should notice that $d_i(v)$, $s_i(g)$, $i = 1, 2$ are functions of the point P on the surface. The notation d is suggestive of the fact that d_i is mainly due to the *diffuse* component of the radiance (that does not depend on the viewpoint), whereas s is suggestive of *specular,* since s_i depends only on the viewing direction.

2.2 A discrepancy measure for non-Lambertian scenes

Naturally, real scenes do not satisfy the conditions of Prop. 1, so the *measured* tensor $\tilde{R}(P)$ has rank greater than 2. 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)):

$$\Phi(P) \doteq \|\tilde{R}(P) - d_1(v) s_1^T(g) - d_2(v) s_2^T(g)\|_F^2 \tag{7}$$

where we have chosen the squared Frobenius norm to compare radiance tensors, although any other matrix norm would do. 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

$$\phi_{ij} = \tilde{R}_{ij} - d_1(v_i) s_1(g_j) - d_2(v_i) s_2(g_j), \tag{8}$$

where $\tilde{R}_{ij}$ is the (i, j)-th element of $\tilde{R}(P)$, 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 $E \doteq \int_S \Phi(P) dA$:

$$\hat{S} \doteq \arg\min_S \int_S \Phi(P) dA \tag{9}$$

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

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

where $r = \min\{m, n\}$. Since, from the rank constraint of Prop. 1, we can choose the basis of R arbitrarily, we can have

$$d_i(v) = \tilde{d}_i(v) \quad \text{and} \quad s_i(g) = \tilde{s}_i(g) \quad i = 1, 2 \tag{10}$$

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

$$\Phi(P) = \|\tilde{d}_3(v)\tilde{s}_3^T(g) + \tilde{d}_4(v)\tilde{s}_4^T(g) + \cdots + \tilde{d}_r(v)\tilde{s}_r^T(g)\|_F^2. \tag{11}$$

By the properties of the SVD, we have that

$$\langle \tilde{d}_i(v), \tilde{d}_j(v) \rangle = a_i^2 \delta_{ij} \quad \text{and} \quad \langle \tilde{s}_i(g), \tilde{s}_j(g) \rangle = b_i^2 \delta_{ij} \tag{12}$$

where $a_i, b_i \in \mathbb{R}$ and $\delta_{ij} = 1$, if $i = j$; $\delta_{ij} = 0$, otherwise.

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.

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 E. The first-order optimality condition is given in the following theorem, which is not proven here for reasons of space (see [9]):

Theorem 1 (optimality condition). *Let $\Phi_{\mathbf{X}}, \Phi_N$ be the first-order derivative 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 is always possible in that Φ_{NN} is real and symmetric). Let H be the mean curvature and $\mathbf{II}(t)$ be the second fundamental form of a vector $t \in T_P(S)$. Then we have that, at the optimum S,*

$$\begin{aligned} &2H\Phi - \langle \Phi_{\mathbf{X}}, N \rangle - 2H \langle \Phi_N, N \rangle - \text{trace}(\Phi_{\mathbf{X}N}) \\ &+ N^T \Phi_{\mathbf{X}N} N + \sum_{i=1}^k \lambda_i \mathbf{II}((I - NN^T)p_i) = 0 \end{aligned} \tag{13}$$

A flow based on the first-order derivatives is given by the following partial differential equation:

$$S_t = 2H\Phi - \langle \Phi_{\mathbf{X}}, N \rangle - 2H \langle \Phi_N, N \rangle . \tag{14}$$

The calculation of the flow above reveals some interesting structure, as major simplification occur after equation (12).

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*

$$\dot{\Phi} = \sum_{i,j=1}^{m,n} 2\phi_{ij} \dot{\tilde{R}}_{ij} \tag{15}$$

where the dot indicates differentiation with respect to ξ.

Proof. We define $\phi_i \doteq \tilde{R}_i - d_1(v_i)s_1(g) - d_2(v_i)s_2(g)$ and $\phi_j \doteq \tilde{R}_j - d_1(v)s_1(g_j) - d_2(v)s_2(g_j)$, where $\tilde{R}_i$ and $\tilde{R}_j$ are shorthands to $[\tilde{R}_{i1}, \tilde{R}_{i2}, \ldots, \tilde{R}_{im}]^T$ and $[\tilde{R}_{1j}, \tilde{R}_{2j}, \ldots, \tilde{R}_{nj}]^T$ respectively. Expanding the derivative we get

$$\begin{aligned} \dot{\Phi} &= \sum_{i,j=1}^{n,m} \dot{\phi}_{ij}^2 = \sum_{i,j=1}^{n,m} 2\phi_{ij} \Big(\dot{\tilde{R}}_{ij} - \dot{d}_1(v_i)s_1(g_j) - \\ &\quad d_1(v_i)\dot{s}_1(g_j) - \dot{d}_2(v_i)s_2(g_j) - d_2(v_i)\dot{s}_2(g_j) \Big) \\ &= \sum_{i,j=1}^{n,m} 2\phi_{ij} \dot{\tilde{R}}_{ij} \\ &\quad + \sum_{i=1}^{n} \dot{d}_1(v_i) \langle \phi_i, s_1(g) \rangle + \sum_{j=1}^{m} \dot{s}_1(g_j) \langle \phi_j, d_1(v) \rangle \\ &\quad + \sum_{i=1}^{n} \dot{d}_2(v_i) \langle \phi_i, s_2(g) \rangle + \sum_{j=1}^{m} \dot{s}_2(g_j) \langle \phi_j, d_2(v) \rangle . \end{aligned}$$

However, from equation (11) we see that ϕ_i is in the span of $\tilde{s}_3(g), \tilde{s}_4(g), \ldots, \tilde{s}_r(g)$ and ϕ_j is in the span of $\tilde{d}_3(v), \tilde{d}_4(v), \ldots, \tilde{d}_r(v)$. Therefore, from equations (10) and (12), we can see that the only term to contribute to the derivative is $\sum_{i,j=1}^{n,m} 2\phi_{ij} \dot{\tilde{R}}_{ij}$. □

As a consequence of the previous result, the flow can be written explicitly as:

$$S_t = 2H\Phi - \sum_{i,j=1}^{m,n} 2\phi_{ij} \left\langle \frac{\partial \tilde{R}_{ij}}{\partial \mathbf{X}} + 2H \frac{\partial \tilde{R}_{ij}}{\partial N}, N \right\rangle \tag{16}$$

We implement the flow (16) using level set methods [16].

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 $d_i(v)$ and $s_i(g)$ for $i = 1, 2$, 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 (10). Given a novel vantage point g', the corresponding function $s_i(g')$, $i = 1, 2$ is interpolated from the existing $s_i(g_j)$. Notice that $d_i(v)$, $i = 1, 2$ does not depend on the viewpoint, and therefore does not need to be interpolated. This desirable byproduct of our framework results from comparing each image to a model. 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 $s_i(g)$ depend directly on the

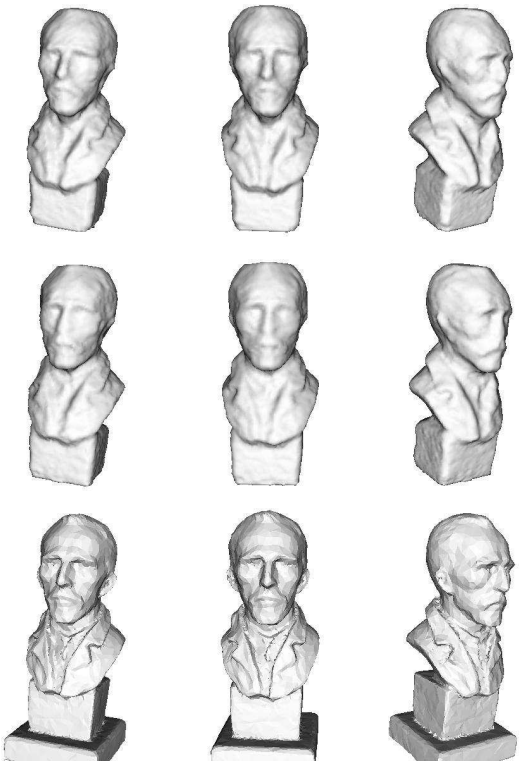

Figure 3: *Estimated shape (top), compared with pseudo-ground truth (bottom), obtained with a 3D laser scanner and manual mesh cleaning. Our results improve those obtained with the algorithm of [10] (middle).*

Figure 4: *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 a true radiance estimate,* not *a texture map.*

viewpoint, and therefore one can generate radiometrically accurate synthetic images from an arbitrary vantage point. When the viewpoint moves, the highlights move on the estimated surface, giving an overall result that is comparable with purely image-based rendering techniques.

4 Experiments

In this section we test the algorithm on the two objects shown in Fig. 1, both courtesy of J.-Y. Bouguet and R. Grzeszczuk (Intel). Van Gogh is made of polished metal, and is highly specular. Pseudo-ground truth has been generated by laser scanning followed by manual mesh polishing (Fig. 3). Buddha is actually a synthetic scene, meant to simulate translucent material. Ground truth is available (Fig. 6). In Fig. 3 we show the estimates of shape produced by the algorithm described in Sect. 3.1, together with the estimates obtained with the algorithm of [10], both compared with pseudo ground truth, obtained with a laser scanner. Our estimate is obviously not as crisp as the ground truth, but it does capture important details on the face. Fig. 7 shows the evolution of the estimate of shape. In Fig. 8 we show synthetic images generated using the radiance map, as described in Sect. 3.2. Note that the specularities move with the viewpoint (best viewed in the movies downloadable from `http://vision.ucla.edu`). In Fig. 4 we show a few synthetic images compared with the real images from the same vantage point. In Fig. 6 we show the estimated shape for the Buddha in Fig. 1. In this case, ground truth is available since the images are synthetic. We also show the results obtained with the algorithm of [10]. In Fig. 5 we show images synthesized from the model, compared with corresponding true images. In Fig. 7 we show the evolution of shape, and in Fig. 8 we show several novel views.

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 equivalent to diffuse+specular reflection models, commonly used in computer graphics, in the sense elucidated in Prop. 1. While one could dismiss the analysis and just introduce the cost func-

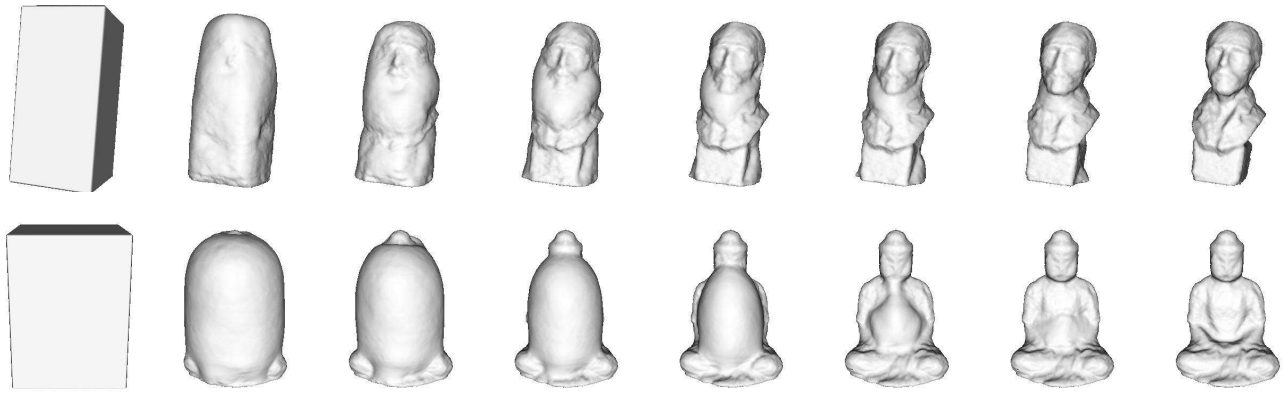

Figure 7: *Shape evolution for Van Gogh (top) and Buddha (bottom).*

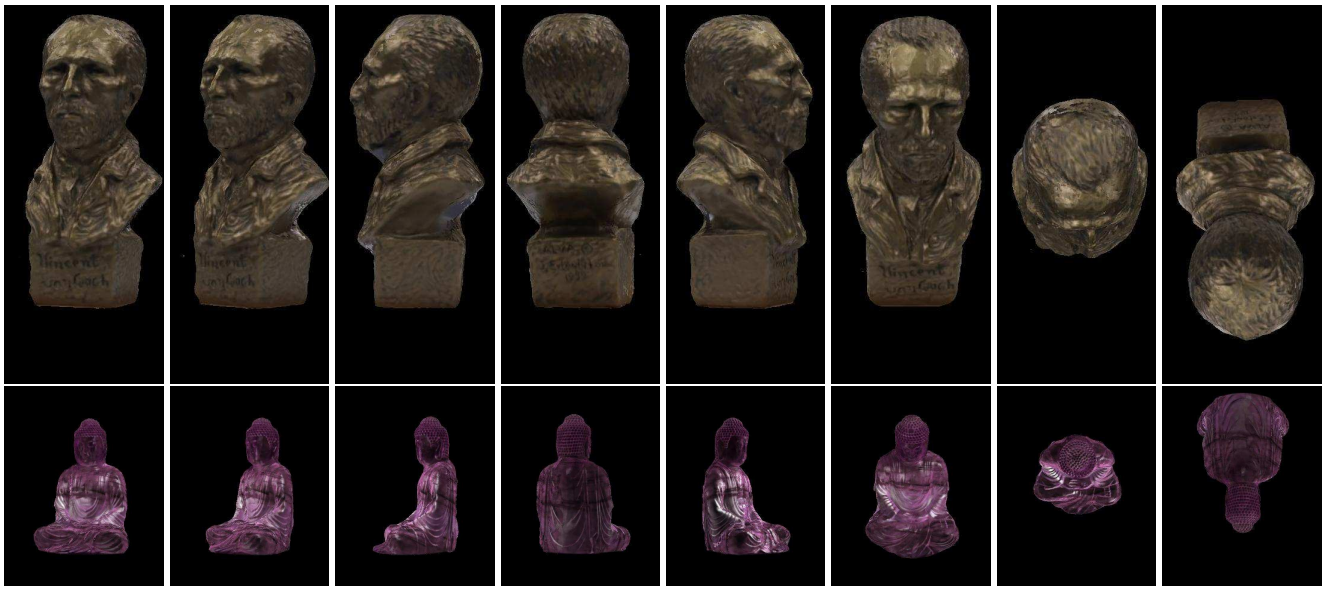

Figure 8: *Synthetic images obtained from the estimated radiance. As it can be seen, the appearance changes with the vantage point.*

tion (7) 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 underlying *mathematical model* is. Of course, as we have pointed out, real surfaces *do not* satisfy the conditions of Prop. 1. Instead, it is this very discrepancy from the idealized model that we exploit to define a constraint that can be used for reconstruction. Those that object to the restrictiveness of the model laid out in Prop. 1 will be relieved to know that extension to higher rank is conceptually and computationally trivial. 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.

Our algorithm can handle sharp changes of the radiance profile: In Fig. 4, one can actually *read* the text at the base of the bust from the reconstructed radiance. 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.

More analysis and experimental results are collected in

a technical report [9] and displayed at the website `http://vision.ucla.edu`.

Figure 5: *Synthetic images obtained with the estimated radiance tensor field (top) compared with the true images taken from the same vantage point.*

Figure 6: *Estimated shape (top), compared with ground truth (bottom), also compared with the results obtained by the algorithm of [10] (middle).*

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] 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.

[5] 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.

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

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

[8] 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.

[9] H. Jin, S. Soatto and A. J. Yezzi. Multi-view Stereo Beyond Lambert. Technical report UCLA-CSD-TR02-0040, 2002.

[10] 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.

[11] 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.

[12] 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.

[13] 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.

[14] 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.

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

[16] 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.

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

[18] 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.

[19] 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.