

Tales of Shape and Radiance in Multi-view Stereo

Stefano Soatto[†]

Anthony J. Yezzi[‡]

Hailin Jin^{†*}

[†] Computer Science Department, University of California, Los Angeles, CA 90095, {soatto, hljin}@cs.ucla.edu

[‡] Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, GA 30332, ayezzi@ece.gatech.edu

Abstract

To what extent can three-dimensional shape and radiance be inferred from a collection of images? Can the two be estimated separately while retaining optimality? How should the optimality criterion be computed? When is it necessary to employ an explicit model of the reflectance properties of a scene? In this paper we introduce a separation principle for shape and radiance estimation that applies to Lambertian scenes and holds for any choice of norm. When the scene is not Lambertian, however, shape cannot be decoupled from radiance, and therefore matching image-to-image is not possible directly. We employ a rank constraint on the radiance tensor, which is commonly used in computer graphics, and construct a novel cost functional whose minimization leads to an estimate of both shape and radiance for non-Lambertian objects, which we validate experimentally.

1 Introduction

We address the problem of inferring geometric and photometric properties of a scene from a collection of images. In particular, we wish to understand to what extent it is possible to recover both geometry (shape) and photometry (radiance), and whether the two problems can be meaningfully “separated”. For scenes that are *static*, (piecewise) *smooth* and Lambertian, we present answers to the following questions:

1. To what extent can shape and radiance be inferred from a collection of images? Are there *generic ambiguities*¹, i.e. scenes with different shape *and* different radiance that give rise to the same images under all possible viewing conditions? (Theorem 3)
2. If we pose the problem of inferring shape and radiance as an optimization problem, relative to some cost functional, is it possible to “*separate*” the estimation of shape from the

radiance? (Theorem 1)

3. What cost functional should be minimized in the optimization? How should this *functional* be *computed*? (Theorem 2)

For scenes that are non-Lambertian, we:

4. Remind the reader that irradiance cannot be compared from image to image, but all images must relate to a common underlying model of reflection (equation (25)).
5. Use a rank constraint on the radiance tensor field, recently introduced in computer graphics, as the underlying model (equation (33)).
6. Propose a gradient flow guaranteed to converge to (local) extrema of a novel cost function computed on the image (equation (34)), yielding an algorithm that estimates both shape and radiance from a collection of images (Section 4).

1.1 Relation to prior work

The traditional approach to multi-view stereo dodges question 2 above by *postulating* the separation of radiance from shape: the former is used to establish correspondence among photometric features in different views and then the latter is inferred from the given correspondence. Under this postulate, question 1 has received a lot of attention during the past two decades. Just to mention a few key contributions, Koenderink and van Doorn [11] derived the set of shape ambiguities for point-like features seen with an affine camera, and Faugeras [4] proposed a hierarchy of ambiguities (projective, affine, Euclidean), depending on knowledge on camera calibration. The analysis of ambiguities in the reconstruction of point-wise structure is now well understood (see [6] and references therein), along with a study on how much data is needed for reconstruction, the behavior of the solution in the presence of noise etc. .

In this paper we consider scenes that may or may not have distinct photometric landmarks, and we represent them as “dense” surfaces at the outset. In this case, the measurements are raw images (*not* point correspondences), and the unknowns are (infinite-dimensional) surface shape and radiance, rather than just a collection of point coordinates in \mathbb{R}^3 . As a consequence, the results in this paper relate to work on photometric stereo. Belhumeur et. al. [2] showed that there is a generalized bas relief ambiguity with unknown lighting

^{*}This research was performed while Hailin Jin was with Department of Electrical Engineering, Washington University, Saint Louis, MO 63130.

¹The term “generic” indicates that the ambiguity is present under all viewing conditions. “Pathological” ambiguities, which we do not study here, involve zero-measure sets of viewing conditions.

and fixed viewpoint. This ambiguity affects both geometry and albedo, assuming that objects are Lambertian. Yuille et. al. [16] this work to include changing the viewpoint as well as the lighting. The ambiguity is an affine transformation on the geometry of the object in conjunction with a transformation on the albedo, assuming an affine camera. Baker et. al. [1] addresses pathological conditions, which we do not treat in this paper. 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. [15] use known shape to estimate global illumination, whereas [12, 17], uses Helmholtz reciprocity to infer shape for objects with arbitrary reflection properties. In addition to the analysis, we describe an algorithm that generates an explicit reconstruction; it is related to the work of Faugeras and Keriven [5] and Jin et. al. [9], but different in that we estimate photometry along with geometry, and integrate the cost functional on the image rather than in space.

1.2 Assumptions and formalization

So far we have used the term “shape” informally to denote the geometry of the scene, which we represent as a collection of smooth surfaces. Indeed, we consider the case of a single smooth surface S , since extensions of our results to multiple connected components can be derived. So, let $S : \mathbb{R}^2 \rightarrow \mathbb{R}^3$; $\mathbf{u} \mapsto P(\mathbf{u})$ be a smooth surface embedded in \mathbb{R}^3 , and $I_t(\mathbf{x}_t)$ a moving image, i.e. a map $I_t : \Omega \subset \mathbb{R}^2 \rightarrow \mathbb{R}_+$ indexed by $t = 0, 1, \dots, T$. The relative motion between the scene and the camera is described by a rigid body transformation² $g_t \in SE(3)$, $t = 0, \dots, T$. Under the Lambertian assumption, radiance can be represented as a positive-valued function ρ defined on the surface³, $\rho(\cdot) : S \rightarrow \mathbb{R}_+$. The irradiance equation [7, page 208] dictates that, in our notation

$$I_t(\mathbf{x}_t) = \rho(P(\mathbf{u})). \quad (1)$$

Without loss of generality, we will assume that the local coordinates of the image $\mathbf{u} \in \Omega \subset \mathbb{R}^2$ coincide with the image plane on the first camera, with generic coordinate \mathbf{x}_0 . Finally, we assume that cameras are calibrated (intrinsics and extrinsics [6]), so that g_t is known, and the imaging process is governed by a canonical perspective projection⁴, π , although extensions to unknown camera motion and possibly

unknown internal parameters can be derived. We choose to work in a deterministic setting; extension of our results in a probabilistic framework are forthcoming. Therefore, the model of image formation is

$$\begin{cases} I_t(\mathbf{x}_t) = \rho(P(\mathbf{x}_0)) + n_t(\mathbf{x}_t) & t = 0, \dots, T; \mathbf{x}_0 \in \Omega \\ \mathbf{x}_t = \pi(g_t P(\mathbf{x}_0)) \end{cases} \quad (2)$$

where n_t is a “noise” term that quantifies the deviation from the ideal irradiance model. Given a collection of images⁵ (or responses of filters⁶), $\{I_t(\mathbf{x}_t) \mid \forall t, \mathbf{x}_t \in \Omega_t \subset \mathbb{R}^2\}$, one can formulate the problem of recovering both the scene structure $P(\cdot)$ and its radiance $\rho(\cdot)$ by minimizing some discrepancy between the image and the model. For instance, from (2), by choosing some norm for the noise, $\|n_t\|_x$, we have

$$\begin{cases} \hat{\rho}(\cdot), \hat{P}(\cdot) = \arg \min_{\rho, P} \sum_{t=0}^T \|I_t(\mathbf{x}_t) - \rho(P(\mathbf{x}_0))\|_x \\ \text{subject to } \mathbf{x}_t = \pi(g_t P(\mathbf{x}_0)). \end{cases} \quad (3)$$

Whether the problem (3) has a unique, well-behaved solution depends on whether there are different choices of ρ , P and g_t in (2) that produce the same images. Since the optimization above is over infinite-dimensional spaces (both ρ and P are functions), regularization may be necessary, as we discuss in later sections.

2 Separability of shape from radiance in Lambertian stereo

Let us consider the problem of jointly estimating shape and radiance from a collection of images. For simplicity, we carry out our discussion for the case of two images, but all considerations apply to an arbitrary number of images [8]. The model that generates the data is therefore of the form

$$\begin{aligned} I_1(\mathbf{x}) &= \rho(P(\mathbf{x})) + n_1(\mathbf{x}) \\ I_2(\pi(gP(\mathbf{x}))) &= \rho(P(\mathbf{x})) + n_2(\mathbf{x}). \end{aligned} \quad (4)$$

In a calibrated setting, we choose a norm $\|\cdot\|_x$ for n_i , and minimize it with respect to ρ and P subject to equation (4). The norm should be defined in the space where the measurements (image irradiance) I_i live, i.e. on the image plane. For instance, the squared 2-norm $\|n\|_2^2 \doteq \int_{\Omega} (n(\mathbf{x}))^2 d\mathbf{x}$, or the 1-norm $\|n\|_1 \doteq \int_{\Omega} |n(\mathbf{x})| d\mathbf{x}$ where $\Omega \subset \mathbb{R}^2$ is a domain on the image plane. Given the additive model, we can substitute the expression of the noise as $n_i = I_i - \rho$. We

² A rigid motion g acts on points P via gP . In coordinates, if P is represented by $\mathbf{X} \in \mathbb{R}^3$, then g is represented by a rotation matrix $R \in SO(3)$ and a translation vector $T \in \mathbb{R}^3$ such that gP is represented by $R\mathbf{X} + T$. The action of g on a vector is represented by g_* , which in coordinates is given by $g_*V = RV$ for a vector $V \in \mathbb{R}^3$.

³ For non-Lambertian surfaces, the radiance function is substituted with a tensor field, $\rho : S \rightarrow \{R : \mathbb{S}^2 \rightarrow \mathbb{R}_+\}$, where at each $P \in S$ and for each direction $v \in \mathbb{S}^2$, $R_P(v) \in \mathbb{R}_+$ measures the amount of energy irradiated from the point P in the direction v . In Section 3 we introduce a discretized version of the tensor R .

⁴ $\pi : \mathbb{R}^3 \rightarrow \mathbb{P}^2$; $\mathbf{X} = [X_1, X_2, X_3]^T \mapsto \mathbf{X}/X_3$.

⁵ Here $\Omega_t = \{\pi(g_t P(\mathbf{x}_0)) \mid \forall \mathbf{x}_0 \in \Omega\}$.

⁶ If filter responses are available instead of image irradiance, it is sufficient to substitute them to $I_t(\mathbf{x}_t)$. In this case, $\rho : S \rightarrow \mathbb{R}^K$ is a vector-valued function, with as many components as filter responses at any given pixel.

therefore write the optimization above using the cost functional

$$\phi(P, \rho) \doteq \|I_1 - \rho\|_x + \|I_2 - \rho\|_x \quad (5)$$

so that

$$\hat{P}, \hat{\rho} = \arg \min_{P, \rho} \phi(P, \rho). \quad (6)$$

For simplicity, we often omit the argument of the functions I_i and ρ . The reader can easily reinstate the notation from equation (4).

At the outset, we notice that, if the shape $P_{true}(\cdot)$ was given, then radiance $\rho(\cdot)$ could be easily estimated. For instance, if we choose the 2-norm, the radiance at the point $\mathbf{X} = P_{true}(\mathbf{x})$ is given by the mean of the reprojected images, or “texture maps”:

$$\hat{\rho}_2(\mathbf{X}) = \frac{1}{2} \sum_{i=1}^2 I_i(\pi(g_i \mathbf{X})) = \text{mean}\{I_i(\pi(g_i \mathbf{X}))\}_{i=1 \dots N} \quad (7)$$

where $g_1 = e$, the identity. If instead we choose the 1-norm, then it is well-known that

$$\hat{\rho}_1(\mathbf{X}) = \text{median}\{I_i(\pi(g_i \mathbf{X}))\}_{i=1 \dots N}. \quad (8)$$

Now, it would be desirable if one could “separate” the problem of estimating shape $P(\cdot)$ from estimating radiance $\rho(\cdot)$, since this would simplify the optimization task. One way to do so is to obtain the ρ that minimizes ϕ for any value of P , and substitute it back into ρ . We therefore have a “reduced” cost functional $\hat{\phi}(P) \doteq \min_{\rho} \phi(P, \rho)$ and the optimization (6) becomes

$$\hat{P} = \arg \min_P \left[\hat{\phi}(P) \doteq \phi(P, \hat{\rho}(P)) \right] \quad (9)$$

where $\hat{\rho}(P) = \arg \min_{\rho} \phi(P, \rho)$. In general, however, an analytic expression of $\hat{\rho}$ as a function of P is difficult to obtain. Therefore, one may be tempted to “eliminate” ρ directly from the model (4); for instance, $\rho = I_1 - n_1$ from the first equation can be substituted into the second equation to yield

$$\tilde{P} = \arg \min_P \left[\|I_1(\mathbf{x}) - I_2(\pi(gP(\mathbf{x})))\|_x \doteq \psi(P) \right] \quad (10)$$

for some choice of norm x . In general, however, this may be incorrect, since $\hat{\phi}(P) \neq \psi(P)$ and therefore the minimizers may be different $\hat{P} \neq \tilde{P}$. Notice that, from equation (5), the functional ϕ denotes the sum of the norms of the error in each image – a very natural criterion – whereas ψ denotes the norm of the *difference* between errors in the image: $\psi = \|n_2 - n_1\|_x$, not quite as principled. Notice also that ψ is what is most commonly used for stereo matching, where images are compared to images as in (10), as opposed to all images being compared to a common radiance, as in (5).

The following theorem shows that, somewhat unintuitively, the two cost functionals $\hat{\phi}$ and ψ , although different, are equivalent in the sense of having the same minimizer. This has consequences for the stereo problem, because it implies that one can first solve for shape, P , by minimizing $\psi(P)$, and then estimate ρ , for instance via (7) or (8) under the Lambertian assumption.

Theorem 1 (separability of shape from radiance). *Let \hat{P} be such that $\hat{P}, \hat{\rho} = \arg \min_{P, \rho} \phi(P, \rho)$, where ϕ is defined in equation (5), and let \tilde{P} be such that $\tilde{P} = \arg \min_P \psi(P)$, where $\psi(P)$ is defined in equation (10). Then $\tilde{P} = \hat{P}$. Consequently, $\hat{\rho}$ can be estimated as $\hat{\rho} = \arg \min_{\rho} \phi(\tilde{P}, \rho)$.*

Proof. We first notice that, by the triangular inequality, $\|I_1 - I_2\|_x = \|(I_1 - \rho) - (I_2 - \rho)\|_x \leq \|I_1 - \rho\|_x + \|I_2 - \rho\|_x$ for any choice of norm x and any function ρ . In particular, the inequality is true for the particular ρ that minimizes the right hand-side, that is $\psi(P) \leq \min_{\rho} \phi(P, \rho)$. On the other hand, $\min_{\rho} (\|I_1 - \rho\| + \|I_2 - \rho\|) \leq \|I_1 - I_2\|$, because one can always find at least one function ρ for which the two expressions are equal, namely $\rho(\cdot)$ such that $\rho(P(\mathbf{x})) = I_1(\mathbf{x})$, and therefore the minimum of the left hand-side must make the expression smaller than the right hand-side. Therefore, we have $\psi(P) \leq \min_{\rho} \phi(P, \rho) \leq \psi(P)$ which proves that $\psi(P) = \min_{\rho} \phi(P, \rho)$. \square

An extension of this proof to an arbitrary number of views can be found in [8].

2.1 Where should stereo discrepancy be computed?

The norm $\|I_2 - I_1\|_x$ involves integrating the functions $I_1(\cdot)$ and $I_2(\cdot)$ on a domain. There are essentially two ways of performing this integral. One is to integrate on the image domain

$$\psi_{im}(P) \doteq \|I_2 - I_1\|_{im} = \int_{\Omega} h(I_2(\pi(gP(\mathbf{x}))) - I_1(\mathbf{x})) d\mathbf{x} \quad (11)$$

where h is a norm on \mathbb{R} , for instance $h(n) = n^2$ or $h(n) = |n|$. The other is to integrate on the surface, by back-projecting the image onto it:

$$\psi_s(P) \doteq \|I_2 - I_1\|_s = \int_S h(I_2(\pi(gP)) - I_1(\pi(P))) dA(P) \quad (12)$$

where $dA(P)$ is the area form on the surface. Of course, one could also write the first integral $\psi_{im}(P)$ on the surface by a change of variable

$$d\mathbf{x} = P \cdot N / Z^3 dA \quad (13)$$

where Z is the third component of $P = [X, Y, Z]^T$ and $N(P)$ is the inward unit normal at P :

$$\psi_{im}(P) = \int_S h(I_2(\pi(gP)) - I_1(\pi(P))) \frac{gP \cdot g_*N(P)}{Z^3} dA(P). \quad (14)$$

This highlights the difference between the two cost functionals (12) and (14): one can see that the two approaches are *not* equivalent, because of the presence of the term $\frac{P \cdot N}{Z^3}$. Does one of the approaches have an advantage over the other?

Discussion 1. *Before answering this question, which we do analytically in the next paragraph, we note that the minimization of ψ is performed over the surface $P(\cdot)$, which in general lives in an infinite-dimensional space. Therefore, in order to render the optimization well-posed, one has to impose regularity by restricting the class of allowable surfaces, $P \in \mathcal{S}$. Integrating on the surface has the advantage of implicitly imposing regularization, since the surface P appears in the area form dA . This could be seen as an advantage, because minimizing ψ_s automatically enforces some sort of regularity. On the other hand, having regularity imposed via the area form causes the discrepancy between images to change depending on the surface being estimated. A more principled approach to imposing regularity in the surface is to separate the data fidelity term (which should only depend on the data, and not on the surface) from the regularizer (which should only depend on the surface, and not on the data). This is best achieved by performing the integral on the image, and adding a regularization term, if necessary (regularization can also be imposed by restricting the class of allowable surfaces, for instance by choosing a parametric class), rather than “twisting” the metric to act as a regularizer.*

In Section 2 we have shown that, when the radiance is unconstrained (ρ can be an arbitrary function), then comparing images to images (through ψ) or images to the model (through ϕ) is equivalent. This is no longer the case when an explicit model of radiance is present. We now show that, when a model of radiance is present, by computing the discrepancy measure on the images (i.e. ψ_{im}) the variation of the cost functional does not depend on derivatives of the images. This means that we can write gradient flow algorithms for estimating shape that do not entail computing derivatives of the image but compute derivatives of the *model* instead. This is a very desirable property in view of the fact that images are only measured up to noise, which is amplified by differentiation, whereas the model is not subject to noise. We will note that this property does not hold if one were to compute the discrepancy measure on the surface instead (i.e. ψ_s).

Let us start with the case of two images; extension to an arbitrary number is straightforward. We rewrite the cost

functionals on the image and on the surface for convenience as follows:

$$\phi_s(\rho, P) = \int_S h(I_1 - \rho) + h(I_2 - \rho) dA \quad (15)$$

$$\phi_{im}(\rho, P) = \int_S \frac{h(I_1 - \rho)}{Z^3} P \cdot N + \frac{h(I_2 - \rho)}{Z^3} P_2 \cdot N_2 dA \quad (16)$$

where $P_2 = gP$, $N_2 = g_*N$ and, again, h is a norm in \mathbb{R} and we have omitted the arguments of $\rho(P)$, $I_1(\pi(P))$ and $I_2(\pi(gP))$. We now show that the variation of ϕ_s depends directly on the gradient of the images I_1, I_2 , whereas the variation of ϕ_{im} does not. The gradient of the image can be thought of as a 2-D gradient $\nabla_{\mathbf{x}} I \in \mathbb{R}^{1 \times 2}$ or, by writing the image point \mathbf{x} in homogeneous coordinates as $\bar{\mathbf{x}}$, as a 3-D gradient $\nabla_{\bar{\mathbf{x}}} I \in \mathbb{R}^{1 \times 3}$, where the image is extended in space as being constant along the projection ray P . In this case, clearly

$$\nabla_{\bar{\mathbf{x}}} I(\pi(P)) \cdot P = 0 \quad (17)$$

since I is constant along P . For simplicity, we carry out the derivation only for the first of the two terms in the integrand in equation (15), which we write in a more concise form respectively as

$$\int_S h(I - \rho) dA \quad \text{and} \quad \int_S \frac{h(I - \rho)}{Z^3} P \cdot N dA. \quad (18)$$

The derivation for the second term in the integrand follows the same steps, with a slightly more involved notation because of the pose g . Now, the gradient flow corresponding to the integral on the image (left in equation (18)) is given by

$$\frac{dP_s}{dt} = 2H(h(I - \rho) - \nabla h(I - \rho)) N \quad (19)$$

where the term $\nabla h(I - \rho)$ involves derivatives of the image I . On the other hand, the integral on the right in equation (18), which calculates the discrepancy on the image, can be thought of as the flux of the vector field $\frac{h}{Z^3} P$ through the surface S . This can be written as the volume integral of the divergence of the vector field in the volume contained in S , $\int_V \nabla \cdot \left(\frac{h(I - \rho)}{Z^3} P \right) dV$. The gradient flow in this case is given simply by

$$\frac{dP_{im}}{dt} = \nabla \cdot \left(\frac{h(I - \rho)}{Z^3} P \right) N. \quad (20)$$

Now, the divergence of the product of a scalar, h/Z^3 , with a vector, P , is the inner product of the gradient of the scalar part by the vector, plus other terms that do not contain derivatives of the scalar part. Therefore, the only term in the equation above that involves derivatives of the image is $\nabla \cdot \left(\frac{h(I - \rho)}{Z^3} \right) \cdot P$, where the gradient of the image appears, through the chain rule, as an inner product with the vector part P . However, in virtue of equation (17), this term cancels out! Carrying out the derivation for the second term in

the integrand in equation (15) leads to a proof of the following theorem.

Theorem 2 (independence of image derivatives). *The first variation of the functional $\phi_{im}(P, \rho)$ in equation (15) does not depend the gradient of the images ∇I_i .*

Note that this result is made possible exactly by the presence of the term $\frac{P \cdot N}{Z^3}$ in the image integral ϕ_{im} , which is what distinguishes it from the surface integral ϕ_s .

Naturally, the flow of equation (20) has to be complemented with a regularization term, as we have discussed in Section 2.1.

2.2 Ambiguities in shape and radiance

As we have noted in Section 2, if the “true” $P(\cdot)$ and g_t were given, the best estimate of ρ would be given, for instance, by equation (7) or (8), depending on the choice of the norm. When the true shape is not known, Theorem 1 guarantees that one can substitute it with the best estimate of shape \hat{P} in the expressions (7) or (8) and obtain the best estimate of radiance. Therefore, if there is an ambiguity in the recovery of P or g_t , the resulting ρ would also be subject to an ambiguity. Vice versa, Theorem 1 suggests that lack of knowledge of the radiance does not introduce additional ambiguities. This is clarified in the following claim, whose proof we carried out only in the absence of visibility limitations (it is omitted due to space constraints, and appears in a technical report [14]). It is worth noting that this result reinforces well-known facts for stereo and structure for motion based on point features, and extends it to arbitrary surfaces supporting a Lambertian reflection.

Theorem 3 (radiance–shape ambiguity). *The set of surface shapes $\tilde{P}(\cdot)$, radiances $\tilde{\rho}(\cdot)$ and motions \tilde{g}_t , $t = 0, \dots, T$ that cannot be distinguished from $P(\cdot), \rho(\cdot), g_t$, $t = 0, \dots, T$ for any number of images T and for all points $\mathbf{x}_0 \in \Omega$ is given by*

$$\tilde{P}(\mathbf{X}) = \alpha g_w P(\mathbf{X}) \forall g_w \in SE(3) \quad (21)$$

$$\tilde{\rho}(\mathbf{X}) = \rho\left(\frac{\mathbf{X}}{\alpha}\right) \forall \alpha \in \mathbb{R}_+ \quad (22)$$

$$\tilde{g}_t = g_t g_w^{-1} \forall t = 0, \dots, T. \quad (23)$$

There are no additional generic ambiguities.

We would like to emphasize that Theorem 3 addresses generic ambiguities in estimating shape and radiance from multiple views. It does not apply to pathological ambiguities [1].

3 Non-Lambertian scenes

The results obtained so far apply to scenes that exhibit Lambertian reflection. If the scene is non-Lambertian, reflection

can no longer be represented by a function $\rho : S \rightarrow \mathbb{R}_+$; instead, a tensor is necessary to represent the amount of energy that the surface radiates at each point P in any direction v (see footnote 3). In the next subsection we describe a discretized representation of this tensor that was derived from computer graphics [3] and first used in the context of reconstruction by Jin et. al. [9].

3.1 The radiance tensor field

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 $v \in \mathbb{R}^3$. Measurements are obtained at a discrete number n of camera poses, g_1, g_2, \dots, 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, \dots, v_m . Therefore, for each point P , we can associate an array of measurements, one column for each view and one row for each vector in the grid on Ω_P :

$$R_P = \begin{bmatrix} \rho(v_1, g_1) & \dots & \rho(v_1, g_n) \\ \vdots & & \vdots \\ \rho(v_m, g_1) & \dots & \rho(v_m, g_n) \end{bmatrix} \quad (24)$$

where $\rho(v_i, g_j)$ represents the local radiance of the surface. Notice that $\tilde{R}_{ij} \doteq \rho(v_i, g_j)$ relates to the image irradiance via of the coordinates of P (the underscore “ P ” indicates this dependency); assuming a pin-hole projection we have, for all $j = 1, 2, \dots, n$,

$$R_{ij} = I_j(\pi(g_j(P + v_i))) \quad \forall v_i \in \Omega_P. \quad (25)$$

Unlike its Lambertian counterpart equation (1), here I_j depends on v_i , and therefore R cannot be eliminated to yield a discrepancy of the type (10). This addresses point 4 in the introduction (Section 1). The map $S \rightarrow \mathbb{R}^{m \times n}$, $P \mapsto R_P$ defines a tensor field on S , R 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

$$\tilde{I}_j(\mathbf{x}) = I_j(\mathbf{x}) + w_j(\mathbf{x}); \quad \tilde{R}_{ij} = R_{ij} + w_{ij} \quad (26)$$

where $w_j(\mathbf{x})$ is a noise term that can be modeled 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 radiance tensor field obtained by substituting the noisy images \tilde{I} in equation (25).

⁷Not to be confused with the rotation matrix $R \in SO(3)$ in footnote 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, for most materials, it will be less than full. As shown by Jin et. al. in [9], constraining the rank to be 2 or less represents a good compromise between richness of the model (it allows capturing specular highlights) and simplicity; allowing the rank to be higher, say 3 or 4, is straightforward, but offset by the complexity of the resulting model. We refer the reader to [9] for a more thorough exposition. For the purpose of illustration, we will restrict our attention to the case of rank 2, which is often used in computer graphics for light field rendering. Therefore, at each point P , we consider the radiance tensor field to be given by

$$R_P = d_1(v)s_1^T(g) + d_2(v)s_2^T(g) \quad (27)$$

for some unknown functions d_i, s_i , where $d_i(v)$ stands for $[d_i(v_1), d_i(v_2), \dots, d_i(v_m)]^T$, and $s_i(g)$ stands for $[s_i(g_1), s_i(g_2), \dots, s_i(g_n)]^T$, $i = 1, 2$. The reader should notice that $d_i(v), s_i(g)$, $i = 1, 2$ are functions of the point P on the surface.

3.2 A discrepancy measure for non-Lambertian scenes

Jin et. al. [9] proposed using the rank constraint in (27) as a discrepancy function for stereo matching. 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 (26)):

$$\Phi(P) \doteq \|\tilde{R}_P - d_1(v)s_1^T(g) - d_2(v)s_2^T(g)\|_F^2 \quad (28)$$

where the subscript F indicates the Frobenius norm. 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

$$\phi_{ij} = \tilde{R}_{ij} - d_1(v_i)s_1(g_j) - d_2(v_i)s_2(g_j), \quad (29)$$

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 $E \doteq \int_S \Phi(P) dA$.

In this paper, motivated by the analysis of Section 2.1, we propose using a different cost functional, obtained by integrating the square of the Frobenius norm on the *image* rather than on the surface:

$$\hat{S} \doteq \arg \min_S \int_S \sum_{i,j} \phi_{ij}^2 \sigma_j dA, \quad (30)$$

where $\sigma_j = \frac{\langle g_j P, g_{j*} N(P) \rangle}{\langle g_j P, e_3 \rangle^3}$ and $e_3 = [0, 0, 1]^T$. This is equivalent to replacing \tilde{R}_P with $\Sigma \tilde{R}_P$ in equation (28) where $\Sigma = \text{diag}(\sqrt{\sigma_1}, \sqrt{\sigma_2}, \dots, \sqrt{\sigma_n})$. Note that the rank of $\Sigma \tilde{R}_P$ is the same as that of \tilde{R}_P , as Σ is a full-rank diagonal matrix. Naturally, the actual measured tensor $\Sigma \tilde{R}$ will not satisfy the rank constraint; therefore, it will in general have full rank. We can write it, for each P , using the singular value decomposition (SVD) as

$$\Sigma \tilde{R}_P = \sum_{i=1}^r \tilde{d}_i(v) \tilde{s}_i^T(g) \quad (31)$$

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

$$d_i(v) = \tilde{d}_i(v) \quad \text{and} \quad s_i(g) = \tilde{s}_i(g) \quad i = 1, 2 \quad (32)$$

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

$$\Phi(P) = \|\tilde{d}_3(v) \tilde{s}_3^T(g) + \tilde{d}_4(v) \tilde{s}_4^T(g) + \dots + \tilde{d}_r(v) \tilde{s}_r^T(g)\|_F^2. \quad (33)$$

3.3 Estimation of shape and radiance

Let the subscript denote partial differentiation, and H the mean curvature. It can be shown (the proof is omitted due to space limitations) that the flow based on the first-order derivatives of the gradient descent flow for the functional (30) is given by the following partial differential equation:

$$S_t = \left(\alpha H - \sum_{i,j=1}^{m,n} 2\phi_{ij} \sigma_j \left\langle \tilde{R}_{ij} \mathbf{X}, N \right\rangle \right) N. \quad (34)$$

Note that αH comes from regularization (H is mean curvature and α a tuning parameter). We implement the flow (34) using the level set method of Osher and Sethian [13].

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 SVD of the measured radiance tensor $\Sigma \tilde{R}$, according to equation (32). Given a novel vantage point g_{new} , the corresponding function $s_i(g_{new})$, $i = 1, 2$ 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. We refer the reader again to [9] for more details on estimating the radiance.

4 Experiments

In Figure 1 (top left) we show an image of a textured object together with the “pseudo-ground truth” (top right). In Figure 1 (bottom-right) we show the reconstruction obtained by a gradient flow algorithm based on equation (19),

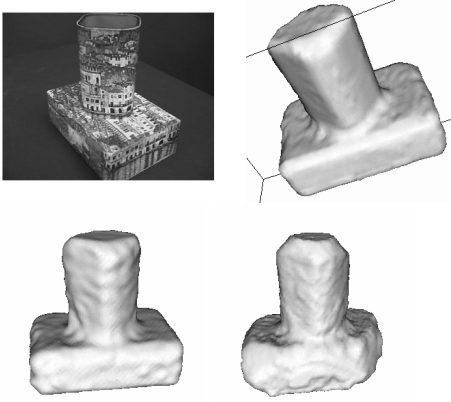


Figure 1: (Top:) Sample view of a textured object used in the experiment (left) and the “pseudo-ground truth” reconstruction (right). (Bottom:) Reconstruction via the gradient flow (35) obtained by integrating the cost functional on the image (left). Compare with the results provided by a gradient flow that integrates the cost function on the surface (right), according to equation (19). Edges are smoothed out due to the uneven weighting of the image data at occluding boundaries. Regularization is fixed by the area form and cannot be controlled. Integrating on the image results in a 19% decrease of the re-projection error.

computed by integrating the cost functional on the surface. One can see that the edges of the object are “smoothed out” by the algorithm, which includes implicitly a regularizing term dA , as discussed in Section 2.1. This is due to the fact that, when the cost functional is integrated on the surface, at the occluding boundary little image information is allowed to affect large portions of the surface via the Jacobian (13). We have then implemented a regularized version of the gradient flow (20), in particular

$$\frac{dP}{dt} = \frac{dP_{im}}{dt} + \alpha H \quad (35)$$

where dP_{im}/dt is given in equation (20) and α was chosen to match the level of smoothness implicit in (19). In Figure 1 (bottom-left) we show the results, that exhibit much sharper boundaries. The re-projection error from the estimated surface to ground truth is reduced by 19% when integrating on the image.

None of these algorithms, however, would work on the scene in Figure 2 (top row, data courtesy of Daniel Wood, University of Washington), due to significant deviations from the Lambertian model. The object is made of polished ceramic, and is highly specular. We therefore test the model-based algorithm described in Section 3 on this object. Note that there is no restriction whatsoever imposed on the variation of the diffuse component of the radiance,

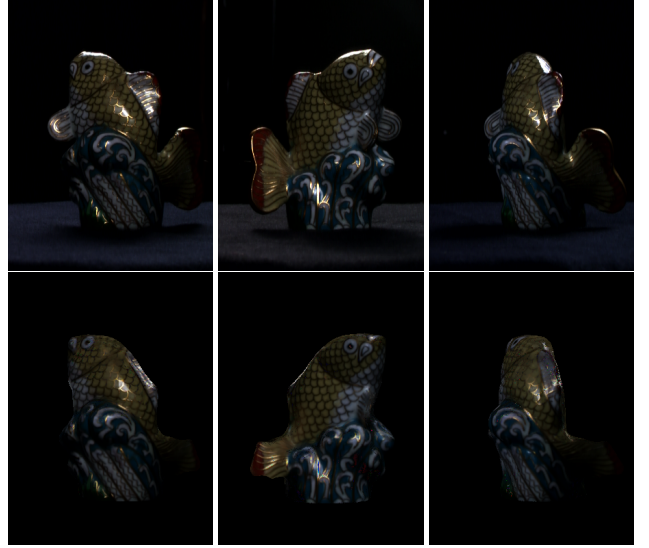


Figure 2: Original images of the fish dataset (top row), courtesy of Daniel Wood, University of Washington and estimated radiance (bottom row).

and nowhere it is assumed that it be constant or smooth. Feature points or curves are effortlessly handled (indeed, welcome) by the algorithm. In Figure 3 we show the estimates of shape produced by the algorithm described in Section 3.3, compared with pseudo-ground truth. We found the rank-2 condition described in the previous section to be a good compromise between robustness and generality; increasing the rank is conceptually straightforward although the resulting algorithms would be less robust with respect to image noise due to the increased number of functions d_i , s_i to be estimated. Note that our algorithm and that in [9] return the estimates of both shape and radiance, while other existing variational algorithms (such as [5, 10]) can only estimate shape. In the bottom row of Figure 2 we show a few synthetic images compared with the real images from same vantage points. Notice that the specularities move with the viewpoint, an indication of the non-Lambertian nature of the model. This is a true estimate of the radiance, *not* a texture map.

5 Conclusions

We have shown that, for Lambertian smooth scenes, comparing images to models (and therefore jointly estimating shape and radiance) is equivalent to comparing images to images (and therefore estimate shape and radiance separately). This result is remarkably simple but compelling since the two cost functionals appear to be very different, one integrating the sum of the norm of the the noise in ev-

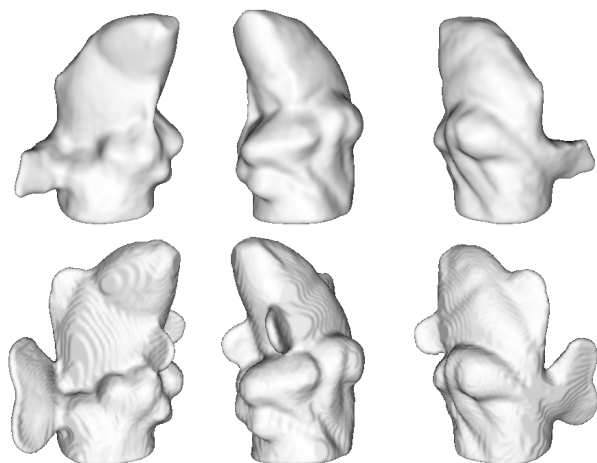


Figure 3: *Estimated shape (top row) and ground truth shape (bottom row). This scene is particularly challenging due to a highly specular surface and dim lighting. With the exception of the fins, the shape and radiance are correctly captured.*

ery image, the other integrating the pairwise difference of noise terms in individual images. When an explicit model of radiance is present, however, comparing images to the model allows enforcing such constraints explicitly. The cost functional that describes the discrepancy of corresponding images should be integrated on the images themselves, and not on the (model) surface, because this avoids the computation of derivatives of the image in the first-order optimality conditions, which are used in gradient flow algorithms. Jointly recovering radiance and shape is subject to the same ambiguities that one would expect when performing point feature-based stereo reconstruction. The results we have presented hold for any choice of norm, and can be generalized in a variety of ways. For non-Lambertian reflection, we have presented a novel cost functional based on a rank constraint of the radiance tensor, as used in computer graphics, which yields a (locally) optimal algorithm to reconstruct both shape and radiance for scenes that exhibit a “specular+diffuse” reflection.

Acknowledgments

We thank Alan Yuille for stimulating discussions. This work is supported by NSF IIS-0208197, AFOSR F49620-03-1-0095, and ONR N00014-02-1-0720.

References

- [1] S. Baker, T. Sim, and T. Kanade. When is the shape of a scene unique given its light-field: a fundamental theorem of

- 3d vision? *IEEE Trans. Pattern Anal. Mach. Intell.*, 25(1), 2002.
- [2] P. Belhumeur, D. Kriegman, and A. L. Yuille. The generalized bas relief ambiguity. *Int. J. of Computer Vision*, 35:33–44, 1999.
- [3] W. C. Chen, J.-Y. Bouguet, M. H. Chu, and R. Grzeszczuk. Light field mapping: efficient representation and hardware rendering of surface light fields. In *Proc. of the ACM SIGGRAPH*, 2002.
- [4] O. D. Faugeras. Stratification of three-dimensional vision: projective, affine and metric representations. *J. Opt. Soc. Am. A*, 12:465–484, 1995.
- [5] O. D. Faugeras and R. Keriven. Variational principles, surface evolution pdes, level set methods and the stereo problem. *INRIA Technical report*, 3021:1–37, 1996.
- [6] R. Hartley and A. Zisserman. *Multiple view geometry in computer vision*. Cambridge University Press, 2000.
- [7] B. K. P. Horn. *Robot vision*. MIT press, 1986.
- [8] H. Jin. Variational Methods for Shape Reconstruction in Computer Vision. *Ph.D. Thesis*, Washington University, Saint Louis, August 2003.
- [9] H. Jin, S. Soatto, and A. J. Yezzi. Multi-view stereo beyond lambert. In *Proc. of IEEE Conf. Computer Vision and Pattern Recog.*, pages 171–178, June 2003.
- [10] H. Jin, A. Yezzi, and S. Soatto. Variational multiframe stereo in the presence of specular reflections. In *Proc. of the Intl. Conf. on 3DPVT*, pages 626–630, June 2002.
- [11] J. J. Koenderink and A. J. Van Doorn. Affine structure from motion. *J. Optic. Soc. Am.*, 8(2):377–385, 1991.
- [12] S. Magda, T. Zickler, D. J. Kriegman, and P. B. Belhumeur. Beyond lambert: reconstructing surfaces with arbitrary brdfs. In *Proc. of the ICCV*, pages 391–398, 2001.
- [13] S. Osher and J. Sethian. Fronts propagating with curvature-dependent speed: algorithms based on hamilton-jacobi equations. *J. of Comp. Physics*, 79:12–49, 1988.
- [14] S. Soatto and A. J. Yezzi. Tales of Shape and Radiance in Multi-view Stereo. *Technical Report CSD-020039*, University of California, Los Angeles, November 2002.
- [15] Y. Yu, P. Debevec, J. Malik, and T. Hawkins. Inverse global illumination: Recovering reflectance models of real scenes from photographs. In *Proc. of the AMS SIGGRAPH*, 1999.
- [16] A. L. Yuille, D. Snow, R. Epstein, and P. Belhumeur. Determining generative models of objects under varying illumination: shape and albedo from multiple images using svd and integrability. *Int. J. of Computer Vision*, 35:203–222, 1999.
- [17] T. Zickler, P. N. Belhumeur, and D. J. Kriegman. Helmholtz stereopsis: exploiting reciprocity for surface reconstruction. In *Proc. of the ECCV*, pages 869–884, 2002.