Toward Global and Model based Multiview Stereo Methods for Shape and Reflectance Estimation

Kuk-Jin Yoon, Amaël Delaunoy, Pau Gargallo, and Peter Sturm
Perception Team, INRIA Rhône-Alpes, Grenoble, France
http://perception.inrialpes.fr

Abstract

In this paper, we present a variational method that recovers both the shape and the reflectance of the Lambertian scene using multiple images. Although we consider only Lambertian surfaces in this paper, the proposed method, which is global and completely model based, is the first and unavoidable stage for reaching a shape and reflectance estimation method for non-Lambertian surfaces.

Basically, our method is a multiview stereo/shape from shading algorithm which allows to recover 3D shapes from Lambertian shading with known illumination conditions. Contrary to previous works that deal with a single material object of the constant albedo, our method works for surfaces with non-constant reflectance parameters, in particular with non-constant albedo. In addition, our algorithm is not based on two or more separate steps – shape and reflectance are jointly recovered in a same process. We verified the proposed method using synthetic images. We will extend our method for non-Lambertian surfaces to improve the robustness to non-Lambertian effects.

1. Introduction

Recovering the three-dimensional surface shape using multiple images is one of the major research topics in computer vision. Many methods have been proposed to solve the problem during these last two decades; refer to [16] for an evaluation of various recent methods. On the other hand, for a long time, the estimation of the surface radiance/reflectance properties was somewhat secondary and was mainly of use to set up the shape reconstruction task [1, 27, 28]. Even some very recent works such as the one of Pons et al. [12, 13] compute the 3D shape without considering radiance estimation. However, the radiance/reflectance estimation has become a concern in multiview reconstruction scenarios in the last decade. For example, Jin, Soatto et al. estimate conjointly the 3D shape and radiance (tensors) [7, 9, 17, 22].

Roughly speaking, the radiance is the combination of the lighting, the reflectance, and the geometry of the scene. For example, radiance contains shading and shadows and, from raw radiance, it is impossible to correct them when changing the lighting. Therefore, recovering reflectance properties is required for realistic relighting, which is also fundamental, for example, in virtual reality as well as augmented reality where the lighting conditions when capturing the object are different from the ones where one resynthesizes it.

In other respects, in real life applications, perfect Lambertian surfaces do not exist. For that reason, multiview stereo algorithms have to be robust to specularities. A number of ideas have been exploited to improve the robustness of the algorithms, also being exhaustive is clearly impossible. A widespread idea consists in using some similarity measures, see for example [1, 8, 10, 12, 13, 21, 23]. The weakness being then the similarity measures proposed are generally not physically motivated. Concerning the robustness to the non-Lambertian effects, it also worth to cite the work of Jin et al. [7] which considers radiance tensor and then does not need similarity measures. However, when some similary measures such as normalized cross correlation [1, 12, 13] could help to be robust to some changes of illumination, the radiance tensor as presented in [7] is not appropriated when the database contains images of the scene lighted by several conditions.

The ultimate and ideal goal of our work is to provide a shape and reflectance estimation method that is global and completely model based as [3]. Also, we hope to finally provide a method that improves the robustness to non-Lambertian effects by directly incorporating a physically based specular model in the mathematical formulation of the problem. By incorporating a complete photometric image formation model, we also hope to exploit prolifically all the photometric phenomena. Also, we thus aim to provide a method that allows to naturally manage with a set of images under with several lighting conditions.

Let us note that actually there already exist recent works that provide solutions in this direction [3, 25]. In [25], Yu and Xu proposed a global and model-based method for re-

covering the 3D shape and the reflectance properties of a non-Lambertian object. Nevertheless, in this last paper, the authors constrain the object to be made by a *single material*; that is to say that the parameters of the reflectance (in particular the albedo) are the same for all the points of the object surface. So, the method in [25] is a "Multiview Shape From Shading" method, similarly as the one proposed by Jin et al. [5] who focuses on the Lambertian case.

Our method provides a multiview stereo/shape from shading algorithm similarly to [5, 6, 11, 20, 24] which allow to recover 3D shapes from Lambertian shading [5, 6, 11] as well as specular shading [20, 24]. Nevertheless, contrary to these previous works [5, 6, 11, 20, 24, 25], in our work, we do not want to restrain ourself to a single material: in other words, the reflectance properties of the object can spatially (strongly) change. In effect, now a day, more and more objects are now printed and so it is fundamental to be able to recover textured and patterned objects. In return, of course, we will not be able to recover lighting conditions as done [5], and we have to use a parallel process which return them. In this work, we assume that lighting conditions are known in advance. Practically, we can use spherical objects with the reference white color to capture the directions and the colors of light sources [14, 26].

Being just at the beginning of our work, we present in this paper the first and unavoidable stage for reaching our goal. This stage consists in adapting Jin et al. work [5] to the objects whose the reflectance properties can spatially change (while keeping the Lambertian model they use), and then in testing its practical feasibility and its numerical convergence and stability. Let us emphasize that we do not claim that the actual method we present here returns better results (for 3D shape and reflectance) than the previous ones since the model we use (the Lambertian model) is really poor and unrealistic. Nevertheless we are convinced that for improving the robustness and the accuracy of the scene reconstruction algorithms, the joint computation of shape and reflectance is crucial and so that this first stage goes in promising direction. Finally let us underline that contrary to [7, 9, 17, 22, 15] which estimate the radiance of the objects, here we propose a global method not for recovering radiance but for recovering the reflectance of the scene along with the shape (shape and reflectance are jointly recovered in a same process): i.e. we separate geometry, reflectance, and illumination from the radiance.

2. Modeling Assumptions and Notations

We assume here that the scene can be decomposed into two entities: the foreground, which corresponds to the objects of interest, and the background; these are defined more precisely below. The foreground is composed by a set of (bounded and closed) 2D manifolds of $\mathbb{R}^3$. These surfaces are represented by S.

2.1. Lighting conditions

We assume that the scene is illuminated by a finite number of distant point light sources. We complete them by adding an ambient light term (which partially accounts for interreflections and other complex phenomena), with constant energy radiated isotropically in all directions. Note that, based on Wiener's theorems, [19] shows that such a light distribution can approximate arbitrarily well any positive distribution on the sphere. Let n_l be the number of illuminants and $\mathbf{l}_j \in \mathbb{S}^2$ and $L_j \in \mathbb{R}^c$ be the direction and the intensity[1] of the j^{th} illuminant, respectively. $L_a \in \mathbb{R}^c$ is the intensity[1] of the ambient illumination.

2.2. Cameras, image data, and visibility

Image data are generated by n_c pinhole cameras. The perspective projection, from world to image coordinates, performed by the i^{th} camera, is represented by $\Pi_i : \mathbb{R}^3 \to \mathbb{R}^2$. $\pi_i \subset \mathbb{R}^2$ is the image domain of the i^{th} camera (i.e. the area covered by the pixels). It is split into two parts: the pixels corresponding to the foreground, $\pi_{iF} = \pi_i \cap \Pi_i(S)$, and the other points $\pi_{iB} = \pi_i \setminus \pi_{iF}$ (associated to the background). $I_i : \pi_i \to \mathbb{R}^c$ is the image of the true scene, captured by the i^{th} camera ($c = 1$ for a gray-scale image, and $c = 3$ for a color image). We denote I the set of input images: $I = \{I_1, I_2, \cdots, I_{n_c}\}$; I_{iF} and I_{iB} are the restrictions of the function I_i to π_{iF} and π_{iB}, respectively. In other respects, we consider the visibility function δ_{S_i} defined by: $\delta_{S_i}(\mathbf{X}) = 1$ if $\mathbf{X}$ is visible from the i^{th} camera and $\delta_{S_i}(\mathbf{X}) = 0$ otherwise. S_i denotes the part of S that is visible from the i^{th} camera and $\Pi_{i,S}^{-1}$ is the backprojection from the i^{th} camera onto S: i.e. for all points $\mathbf{x} \in \pi_{iF}$, $\Pi_{i,S}^{-1}(\mathbf{x})$ is the closest point on S along the ray joining $\mathbf{X}$ to the optical center of the i^{th} camera.

2.3. Modeling the background

As suggested by [22], to be sure that the estimated foreground surface does not shrink to an empty set (which is indeed the global optimum for most cost functionals used in other works) it is crucial to define and characterize the background. The choice of model is dictated by the scenario and the applications. For example, in [6, 22], the background is characterized by its radiance which is constrained to be constant or strongly regular. At the opposite extreme, when the background is quite irregular, one can assume that one has at his disposal the background images, i.e. the images of the scene captured by the same cameras without foreground objects. Due to lack of space, we only deal here with the latter scenario. Therefore, in addition to the images I, we assume that we detain the background images $\tilde{I} = \{\tilde{I}_1, \cdots, \tilde{I}_{n_c}\}$. Finally, we define $\tilde{I}_{iF}$ and $\tilde{I}_{iB}$ analogously to I_{iF} and I_{iB}.

[1] Non-normalized color vector, if $c = 3$.

2.4. Modeling the foreground surface

In this work, we model the foreground object(s) by its shape S and its reflectance R. We denote $\Omega = (S, R)$.

When assuming that $I_i(\mathbf{x})$ is equal to the radiance of the surface S at point $\mathbf{X} = \Pi_{i,S}^{-1}(\mathbf{x})$ in the direction of the i^{th} camera, the image I_i is decomposed as

$$I_i = I_{id} + I_{ia}, \tag{1}$$

where I_{id} and I_{ia} are images with the diffuse and ambient reflection component of I_i, respectively.

Diffuse reflection is caused by the subsurface scattering of light and it is independent of viewing direction. By using the cosine law, this image component is described as

$$I_{id}(\mathbf{x}) = \sum_{j=1}^{n_l} \delta_{L_j}(\mathbf{X})\Big(\rho_d(\mathbf{X}) L_j \left(\mathbf{n}(\mathbf{X}) \cdot \mathbf{l}_j\right)\Big), \tag{2}$$

where $\rho_d(\mathbf{X}) \in \mathbb{R}^c$ is the diffuse albedo[1] at point $\mathbf{X}$, $\mathbf{n}(\mathbf{X})$ is the outward normal vector to the surface S at $\mathbf{X}$ and δ_{L_j} represents the light visibility function: S_{L_j} being the part of S visible from the j^{th} illuminant, we define $\delta_{L_j}(\mathbf{X}) = 1$ if $\mathbf{X} \in S_{L_j}$, $\delta_{L_j}(\mathbf{X}) = 0$ otherwise.

The ambient illumination is assumed to be uniform in the scene and the ambient reflection is modeled as

$$I_{ia}(\mathbf{x}) = \rho_d(\mathbf{X}) L_a, \tag{3}$$

where L_a is defined above.

By combining the diffuse and ambient reflection, we get the image formation equation as

$$I_i(\mathbf{x}) = \sum_{j=1}^{n_l} \delta_{L_j}(\mathbf{X}) \mathbb{L}_{ij}(\mathbf{X}, \mathbf{n}(\mathbf{X})) + \rho_d(\mathbf{X}) L_a, \tag{4}$$

where $\mathbb{L}_{ij}(\mathbf{X}, \mathbf{n}(\mathbf{X})) = L_j \rho_d(\mathbf{X})\big(\mathbf{n}(\mathbf{X}) \cdot \mathbf{l}_j\big)$.

3. Problem Formulation

From a probabilistic point of view, the goal of this work is to estimate the shape S and the reflectance R of a scene surface Ω, that maximize $P(\Omega|I)$ for given I. By Bayes' rule, the problem is then formulated as

$$P(\Omega|I) \propto P(I|\Omega)\, P(\Omega) = P(I|S,R)\, P(S)\, P(R) \tag{5}$$

under the assumption that S and R are independent. Here, $P(I|\Omega) = P(I|S,R)$ is a likelihood while $P(S)$ and $P(R)$ are priors on the shape and reflectance respectively.

3.1. Likelihood

If Π_i and illumination conditions are given, we can produce a synthetic image $\bar{I}_i(\Omega)$ corresponding to an input image I_i by using the current estimation of Ω. Here, the correct estimation of Ω will produce the same images as the input images under the given illumination conditions (modulo noise of course). This allows us to measure the validity of the current estimation by comparing input images with generated ones. When assuming an independent identical distribution (i.i.d) of noise in the observations, the likelihood can be expressed as

$$P(I|\Omega) \propto \prod_{i=1}^{n_c} \exp\big(-\xi_i(\Omega)\big) = \prod_{i=1}^{n_c} \exp\big(-\xi(I_i, \bar{I}_i(\Omega))\big), \tag{6}$$

where $\xi_i(\Omega) = \xi(I_i, \bar{I}_i(\Omega))$ is a function of Ω, measuring the dissimilarity between two images I_i and $\bar{I}_i$.

3.2. Prior

A typical and reasonable prior for the surface shape S is about the area or about the smoothness of a surface. When using the surface area for the prior on S, it is expressed as

$$P(S) \propto \exp\big(-\psi(S)\big). \tag{7}$$

Here, $\psi(S)$ is the monotonic increasing function of the surface area $\int_S d\sigma$ where $d\sigma$ is the Euclidean surface measure.

$P(R)$ can be assumed uniform in general. This means that we do not use any constraints for the surface reflectance and the albedo can be a pattern and texture.

4. Cost Functions

Based on the derivations in Sec. 3, the problem can be expressed in terms of cost functions as

$$E_{total}(\Omega) = E_{data}(\Omega) + E_{shape}(S) = \sum_{i=1}^{n_c} \xi_i(\Omega) + \psi(S), \tag{8}$$

because that maximizing the probability (Eq. (5)) is equal to minimizing the total cost (Eq. (8)).

4.1. Data cost function

The current estimation of Ω gives a segmentation of the input image I_i into foreground I_{iF} and background I_{iB} and we can synthesize $\bar{I}_{iF}$ according to the above image formation model. As for $\bar{I}_{iB}$, it is generated according to the available background model. In this paper, as mentioned in Sec. 2.3, we use actual background images, i.e. $\bar{I}_{iB}$=$\tilde{I}_{iB}$. $\xi_i(\Omega) = \xi(I_i, \bar{I}_i)$ is then rewritten as

$$\begin{aligned} \xi(I_i, \bar{I}_i) &= \xi_F(I_{iF}, \bar{I}_{iF}) + \xi_B(I_{iB}, \bar{I}_{iB}) \\ &= \xi_F(I_{iF}, \bar{I}_{iF}) + \xi_B(I_{iB}, \tilde{I}_{iB}) \\ &= \hat{\xi}_F(I_{iF}, \bar{I}_{iF}) + \xi(I_i, \tilde{I}_i), \end{aligned} \tag{9}$$

where $\hat{\xi}_F(I_{iF}, \bar{I}_{iF}) = \xi_F(I_{iF}, \bar{I}_{iF}) - \xi_F(I_{iF}, \tilde{I}_{iF})$. Since $\xi(I_i, \tilde{I}_i)$ is independent of Ω, the data cost function is written as

$$E_{data}(\Omega) = \sum_{i=1}^{n_c} \hat{\xi}_F(I_{iF}, \bar{I}_{iF}) + C, \tag{10}$$

where $C = \sum_{i=1}^{n_c} C_i = \sum_{i=1}^{n_c} \xi(I_i, \tilde{I}_i)$ is constant.

4.1.1 Similarity Measure

When computing ξ, any statistical correlation among color or intensity patterns such as the sum of squared differences (SSD), cross correlation (CC), and mutual information (MI) can be used. In any case, ξ can be expressed as the integral over the image plane as

$$\xi(I_i, \bar{I}_i) = \int_{\pi_i} e(\mathbf{x}) d\sigma_i, \tag{11}$$

where $d\sigma_i$ is the surface measure and $e(\mathbf{x})$ is the contribution at $\mathbf{x}$ to ξ_i. The data cost function is then given as

$$E_{data}(\Omega) = \sum_{i=1}^{n_c} \int_{\pi_{iF}} \hat{e}(\mathbf{x}) d\sigma_i + C, \tag{12}$$

where $\hat{e}(\mathbf{x}) = e\big(I_i(\mathbf{x}), \bar{I}_i(\mathbf{x})\big) - e\big(I_i(\mathbf{x}), \tilde{I}_i(\mathbf{x})\big)$. We adopt the derivations proposed in [12] for ξ_i, e, and $\partial_2 e$.

4.1.2 Decoupling appearance from surface normal

As shown in Eq. (4), surface appearance (i.e., the data cost function) is dependent on both the surface normal and position, and this makes the problem hard to solve and unstable. To resolve this problem, we introduce an auxiliary unit vector field $\mathbf{v}$ satisfying $\|\mathbf{v}\| = 1$ as in [6], which is used for the computation of surface appearance. To penalize the deviation between the actual normal vector $\mathbf{n}$ and the auxiliary normal vector $\mathbf{v}$, we add a new term

$$\begin{aligned} E_{dev}(\Omega) &= \tau \int_S \chi(\mathbf{X}) d\sigma = \frac{\tau}{2} \int_S \|\mathbf{n}(\mathbf{X}) - \mathbf{v}(\mathbf{X})\|^2 d\sigma \\ &= \tau \int_S \left(1 - (\mathbf{n}(\mathbf{X}) \cdot \mathbf{v}(\mathbf{X}))\right) d\sigma, \end{aligned} \tag{13}$$

to the cost function, where τ is a control constant.

4.2. Shape area cost function

By using the area of a surface for the prior, the shape area cost function is simply defined as

$$E_{shape}(S) = \psi(S) = \lambda \int_S d\sigma, \tag{14}$$

where λ is a control constant.

4.3. Total cost function

By combining the cost functions defined in the previous sections, the total cost function is given by

$$E_{total}(\Omega) = E_{data}(\Omega) + E_{dev}(\Omega) + E_{shape}(S). \tag{15}$$

Here, it is worthy of notice that $E_{dev}(\Omega)$ and $E_{shape}(S)$ are defined over the scene surface while $E_{data}(\Omega)$ is defined as an integral over the image plane. By the change of variable

$$d\sigma_i = -\frac{\mathbf{d}_i(\mathbf{X}) \cdot \mathbf{n}(\mathbf{X})}{z_i(\mathbf{X})^3} d\sigma, \tag{16}$$

where $\mathbf{d}_i(\mathbf{X})$ is the vector connecting the center of the i_{th} camera and $\mathbf{X}$ and $z_i(\mathbf{X})$ is the depth of $\mathbf{X}$ relative to the i_{th} camera, we can replace the integral over the image plane by an integral over the surface:

$$\begin{aligned} E_{data}(\Omega) &= C - \sum_{i=1}^{n_c} \int_{S_i} \Big(\hat{e}(\mathbf{x}) \frac{\mathbf{d}_i(\mathbf{X}) \cdot \mathbf{n}(\mathbf{X})}{z_i(\mathbf{X})^3} \Big) d\sigma \\ &= C - \int_S \left(\sum_{i=1}^{n_c} \delta_{S_i}(\mathbf{X}) \hat{e}(\mathbf{x}) \frac{\mathbf{d}_i(\mathbf{X}) \cdot \mathbf{n}(\mathbf{X})}{z_i(\mathbf{X})^3} \right) d\sigma \end{aligned} \tag{17}$$

When denoting $g(\mathbf{X}, \mathbf{n}(\mathbf{X})) : \mathbb{R}^3 \times \Omega \to \mathbb{R}$ as

$$g(\mathbf{X}, \mathbf{n}(\mathbf{X})) = \left(-\sum_{i=1}^{n_c} \Big(\delta_{S_i} \hat{e} \frac{\mathbf{d}_i \cdot \mathbf{n}}{{z_i}^3} \Big) + \tau\chi + \lambda \right), \tag{18}$$

Eq. (15) is simply rewritten as

$$E_{total}(\Omega) = C + \int_S g(\mathbf{X}, \mathbf{n}(\mathbf{X})) d\sigma. \tag{19}$$

Here, although the total cost function is an integral over the surface, it does not suffer from the usual minimal surface bias: most functionals used in multiple stereo have an empty set as globally optimal surface, since they do not "explain" all pixels in the input images. Our approach, like [22], takes into account all pixels in the cost function, using both the estimated foreground and the available background information.

5. Scene Recovery

Scene recovery is achieved by minimizing E_{total} while updating S and R. Unfortunately, S and R are highly coupled and, therefore, it is very complicated to estimate all unknowns simultaneously. To efficiently solve the problem, we adopt an alternating scheme, updating S for a fixed R and then R for a fixed S. This procedure is repeated until E_{total} no longer decreases and S and R no longer change.

5.1. Shape estimation – Surface evolution

When assuming that R is given, E_{total} is a function of S. In this case, the gradient of $E_{total}(S)$ is given according to the derivation in [4] and [18] as

$$\nabla_S E_{total}(S) = (\nabla_S g) \cdot \mathbf{n} + 2gH + \nabla_S \cdot g_{\mathbf{n}}, \tag{20}$$

where H is the mean curvature and $g_{\mathbf{n}}$ represents the gradient on the unit sphere. Accordingly, the gradient descent

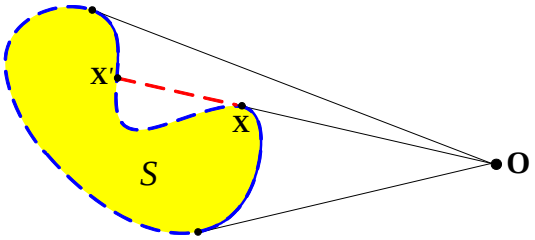


Figure 1. Horizon point $\mathbf{X}$ and its terminator point $\mathbf{X}'$

surface evolution that minimizes the total cost function is characterized by the normal velocity of the evolution given as

$$S_t = -\left((\nabla_S g)\cdot\mathbf{n} + 2gH + \nabla_S\cdot g_{\mathbf{n}}\right). \quad (21)$$

In this work, we derive the gradient descent flows corresponding to the cost functions respectively. The final gradient descent flow is then given by

$$S_t = \left(S_t\big|_{data} + S_t\big|_{dev} + S_t\big|_{shape}\right)\mathbf{n}. \quad (22)$$

5.1.1 Gradient descent flow for the data cost

As shown in Eq. (17), the data cost is a function of the visibility of a surface point, which is dependent on the whole surface shape. According to our previous derivation in [2], $S_t\big|_{data}$ is given as

$$S_t\big|_{data} = \sum_{i=1}^{n_c}\Big(\frac{\delta_{S_i}(\hat{e}_i-\hat{e}_i')}{z_i^3}(\nabla_S h(-\mathbf{d}_i\cdot\mathbf{n})\cdot\mathbf{d}_i) + \frac{\delta_{S_i}}{z_i^3}\left((\partial_2\hat{e}_i\nabla\bar{I}_i)\cdot\mathbf{d}_i\right)\Big), \quad (23)$$

where $h(\cdot)$ is a Heaviside step function and $\hat{e}_i'$ is an error computed by using the radiance at point $\mathbf{X}'$ in the direction of the i^{th} camera, which is the terminator of a horizon point $\mathbf{X}$ as shown in Fig. 1. When a horizon point has no terminator point on the surface, $\hat{e}_i' = 0$ because the terminator point is from the background. $\nabla\bar{I}_i$ is expressed by using Eq. (4) as

$$\nabla\bar{I}_i = \sum_{j=1}^{n_l}\{(\nabla\delta_{L_j})\mathbb{L}_{ij} + \delta_{L_j}(\nabla\mathbb{L}_{ij})\} + (\nabla\rho_a)L_a, \quad (24)$$

where

$$\nabla\mathbb{L}_{ij} = L_j(\nabla\rho_d)(\mathbf{v}\cdot\mathbf{l}_j) + L_j\rho_d\left(\nabla(\mathbf{v}\cdot\mathbf{l}_j)\right). \quad (25)$$

This gradient descent flow includes both the variation related to the camera visibility changes (the first term in Eq. (23)) and the variation related to the image changes (the second term in Eq. (23)), which also includes the variation due to the light visibility changes. Here, it is worthy of notice that the gradient descent flow for the data cost is not dependent on the image gradient, which is sensitive to image noise, but on the shape/reflectance estimation.

5.1.2 Gradient descent flows for the normal deviation cost and the shape area cost

$S_t\big|_{dev}$ (originating from $E_{dev}(\Omega)$) is computed as

$$S_t\big|_{dev} = (-2\tau H + \tau(\nabla_S\cdot\mathbf{v})) \quad (26)$$

and $S_t\big|_{shape}$ (from $E_{shape}(S)$) is the mean curvature flow

$$S_t\big|_{shape} = -2\lambda H. \quad (27)$$

5.2. Updating the auxiliary vector field $\mathbf{v}$

The computed gradient descent flows minimize the total cost with respect to given reflectance and $\mathbf{v}$. We then update the auxiliary vector field $\mathbf{v}$ to minimize the total cost with respect to given shape and reflectance. The $\mathbf{v}$ that minimizes the total cost satisfies the following equation.

$$\frac{\partial g}{\partial\mathbf{v}} = \left(-\sum_{i=1}^{n_c}\delta_{S_i}\partial_2\hat{e}\frac{\partial\bar{I}_i}{\partial\mathbf{v}}\frac{\mathbf{d}_i\cdot\mathbf{n}}{{z_i}^3}\right) + (-\tau\mathbf{n}) = 0 \quad (28)$$

Here, $\frac{\partial\bar{I}_i}{\partial\mathbf{v}}$ is given as

$$\frac{\partial\bar{I}_i}{\partial\mathbf{v}} = \sum_{j=1}^{n_l}\delta_{L_j}L_j\rho_d\mathbf{l}_j. \quad (29)$$

We update $\mathbf{v}$ by performing gradient descent using the following PDE, with the constraint $\|\mathbf{v}\| = 1$.

$$\frac{\partial\mathbf{v}}{\partial t} = \left(-\sum_{i=1}^{n_c}\delta_{S_i}\partial_2\hat{e}\frac{\partial\bar{I}_i}{\partial\mathbf{v}}\frac{\mathbf{d}_i\cdot\mathbf{n}}{{z_i}^3}\right) + (-\tau\mathbf{n}) \quad (30)$$

5.3. Reflectance estimation

We estimate R for a fixed S, still minimizing the total cost function. Since E_{dev} and E_{shape} do not depend on R at all, we seek an optimal R by minimizing $E_{data}(\Omega)$.

For given S, we estimate ρ_d that minimizes the following.

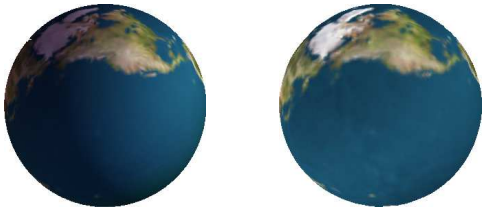
$$E_{data} = \int_S\Big(-\sum_{i=1}^{n_c}\delta_{S_i}\hat{e}\frac{\mathbf{d}_i\cdot\mathbf{n}}{{z_i}^3}\Big)d\sigma \quad (31)$$

Here, ρ_d that minimizes the total cost function will satisfy the Euler-Lagrange equation given as

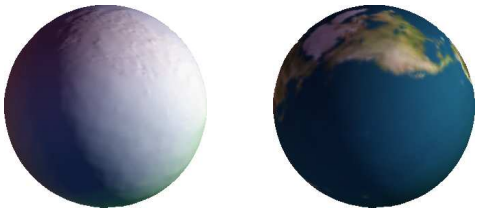
$$-\sum_{i=1}^{n_c}\delta_{S_i}\partial_2\hat{e}\frac{\partial\bar{I}_i}{\partial\rho_d}\frac{\mathbf{d}_i\cdot\mathbf{n}}{{z_i}^3} = 0, \quad (32)$$

where Δ_S denotes the Laplace-Beltrami operator defined on the surface S and $\frac{\partial\bar{I}_i}{\partial\rho_d}$ is given as

$$\frac{\partial\bar{I}_i}{\partial\rho_d} = \sum_{j=1}^{n_l}\delta_{L_j}L_j\left(\mathbf{v}\cdot\mathbf{l}_j\right) + L_a. \quad (33)$$



(a) input image (b) estimated reflectance



(c) estimated shading (d) synthesized image

Figure 2. Estimation results for a "sphere" image set

(a) original input image (b) synthesized image

Figure 3. Scene synthesis under different lighting conditions

We solve the PDE by performing gradient descent using the following PDE.

$$\frac{\partial \rho_d}{\partial t} = \left(-\sum_{i=1}^{n_c} \delta_{S_i} \partial_2 \hat{e} \frac{\partial \bar{I}_i}{\partial \rho_d} \frac{\mathbf{d}_i \cdot \mathbf{n}}{{z_i}^3} \right) \tag{34}$$

6. Experiments

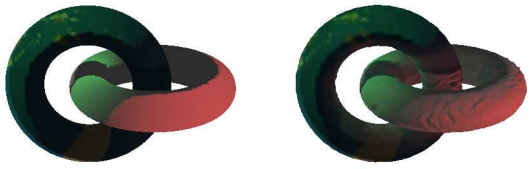
To verify the proposed method, we generated synthetic data sets by specifying illumination conditions and surface reflectance with known 3D models.

Instead of implementing the surface evolution directly, we have implemented the gradient descent surface evolution in the level set framework, in which the topological changes of surfaces are handled automatically. In all experiments, $(128 \times 128 \times 128)$ grids were used except the "dragon" image set[2]. The algorithm starts from the visual hull obtained by rough silhouette images to reduce the computational time and to avoid local minima. The camera and light visibility are computed by using the OpenGL and the simple L^2-norm is used to compute the image similarity, e.

Figure 2 shows one of 32 input images and the synthesized image generated by using the estimated shape (and shading) and reflectance. We can generate the images of the scene with different lighting conditions as shown in Fig. 3 by using the estimated shape and reflectance. The results for the "torus" image set are also shown in Fig. 4 and the results for more complex object is shown in Fig. 5. We can see that the images synthesized by using the estimation closely resemble the input images while the shading and the reflectance are successfully separated. The image synthesized by using the estimated shape and reflectance with different lighting conditions and a different viewpoint is shown in Fig. 6.

The estimated shape is quantitatively evaluated in terms of accuracy and completeness as in [16]. We used 95% for accuracy and the 10.0mm error for completeness. Here, beside the shape evaluation, we also evaluated the estimated reflectance in the same manner. For each point on a estimated surface, we found the nearest point on the surface

[2] $(160 \times 80 \times 128)$ grids was used for the "dragon" image set.

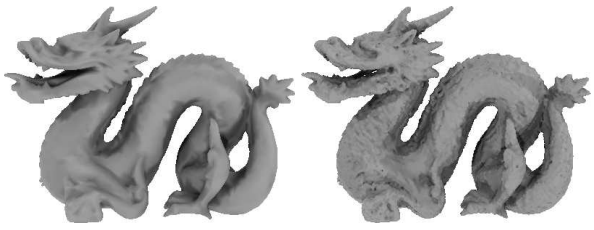


(a) input image (b) synthesized image

Figure 4. Estimation results for a "torus" image set

(a) input image (b) synthesized image

(c) true reflectance of (a) (d) estimated reflectance of (a)

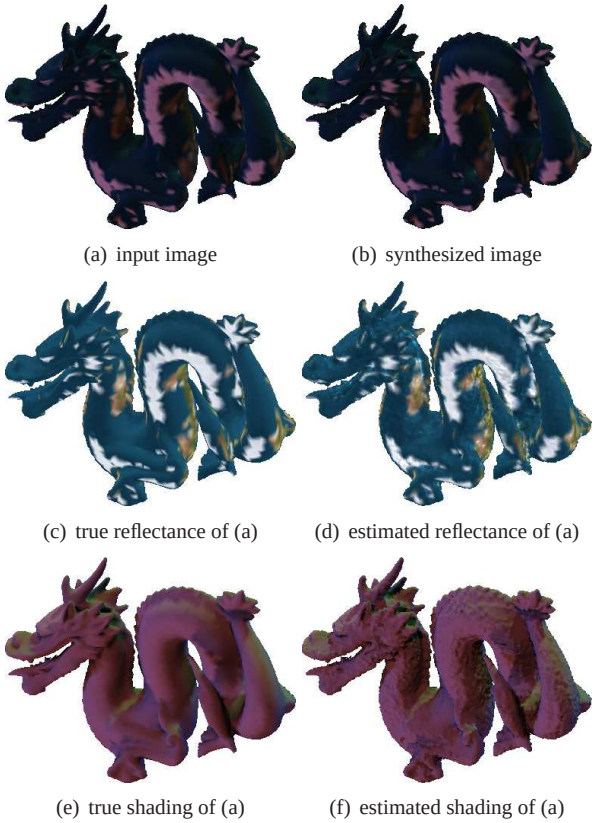


(e) true shading of (a) (f) estimated shading of (a)

Figure 5. Estimation results for a "dragon" image set

Table 1. Performance of the proposed method

	accuracy (95%) (shape, ρ_{dr}, ρ_{dg}, ρ_{db})	completeness (10.0mm) (shape, ρ_{dr}, ρ_{dg}, ρ_{db})	e_{imgae}
sphere	14.04mm, 0.0254, 0.0189, 0.0167	97.17%, 0.0228, 0.0175, 0.0161	0.6026
dragon	2.63mm, 0.0897, 0.0734, 0.0655	99.88%, 0.0658, 0.0575, 0.0543	5.4812



(a) input image (b) synthesized image

(c) true reflectance of (a) (d) estimated reflectance of (a)

(e) true shading of (a) (f) estimated shading of (a)

Figure 6. Synthesized image with different lighting conditions and a different viewpoint

and compute the distance and reflectance differences and vice versa. In addition, for the more quantitative evaluation of surface shape and reflectance, we computed the average of the differences between input images and synthesized images using the L^2-norm as

$$e_{image} = \frac{1}{n_c}\sum_{i=1}^{n_c}\frac{1}{A}\int_{\pi_i} \|\big(I_i(\mathbf{x}) - \bar{I}_i(\mathbf{x})\big)\| d\sigma_i, \qquad (35)$$

where $A = \int_{\pi_i} d\sigma_i$. The performance of the proposed method is summarized in Table 1.

7. Discussion: single-material non-Lambertian surface

When dealing with a single-material non-Lambertian surface[3], it is possible to apply our method to estimate both diffuse/specular reflectance. In order to get a solvable minimization problem without too many unknown variables, we can choose to represent the reflectance by a parametric model. Of course the chosen model directly depends on the applications aimed at; as an example, we consider the popular Blinn-Phong shading model. In this case, the image I_i is decomposed as

$$I_i = I_{id} + I_{is} + I_{ia}, \qquad (36)$$

where I_{is} is an image with the specular reflection component of I_i. Specular reflection is caused by the surface reflection, as with a mirror. This component is expressed as

$$I_{is}(\mathbf{x}) = \sum_{j=1}^{n_l} \delta_{L_j}(\mathbf{X})\left(\rho_s L_j\big(\mathbf{n}(\mathbf{X})\cdot\mathbf{h}_{ij}(\mathbf{X})\big)^{\alpha_s}\right), \qquad (37)$$

where $\mathbf{h}_{ij}(\mathbf{X})$ is the bisector of the angle between the view of the i^{th} camera and the j^{th} illuminant at $\mathbf{X}$, $\rho_s \in \mathbb{R}^c$ and $\alpha_s \in \mathbb{R}^+$ are the specular albedo and the shininess parameter of the surface. These can be computed by performing gradient descent using the following PDEs.

$$\frac{\partial \rho_s}{\partial t} = \int_S \left(-\sum_{i=1}^{n_c}\delta_{S_i}\partial_2\hat{e}\frac{\partial \bar{I}_i}{\partial \rho_s}\frac{\mathbf{d}_i\cdot\mathbf{n}}{{z_i}^3}\right) d\sigma \qquad (38)$$

$$\frac{\partial \alpha_s}{\partial t} = \int_S \left(-\sum_{i=1}^{n_c}\delta_{S_i}\partial_2\hat{e}\frac{\partial \bar{I}_i}{\partial \alpha_s}\frac{\mathbf{d}_i\cdot\mathbf{n}}{{z_i}^3}\right) d\sigma \qquad (39)$$

8. Conclusion and Further Work

In this paper, we have presented a variational method that recovers both the shape and the reflectance of the scene using multiple images. We modeled the scene and the image formation using given information about cameras and illuminants. We then formulated the problem via Bayes' rule and defined the global cost functional. Scene recovery is achieved by minimizing the global cost functional alternatively. As a result, the proposed method produces the complete description of a scene surface enough for resynthesizing the images with different lighting conditions.

As we mentioned, the method presented in this paper is the first and unavoidable stage for reaching a global and completely model-based method for non-Lambertian surfaces. Therefore, we will first apply the method to real images with the estimated lighting and will also extend our method to improve the robustness to non-Lambertian effects by directly incorporating a physically based specular model

[3]This does not mean that the surface has the constant diffuse reflectance. Only specular reflectance is constant over the surface.

in the mathematical formulation of the problem. By incorporating a complete photometric image formation model, we will also exploit prolifically all the photometric phenomena to provide a method that allows to naturally manage with a set of images under with several lighting conditions.

Acknowledgement

This work has been financially supported by the Flamenco project ANR-06-MDCA-007.

References

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

[2] P. Gargallo, E. Prados, and P. Sturm. Minimizing the reprojection error in surface reconstruction from images. In *IEEE International Conference on Computer Vision*, 2007.

[3] A. S. Georghiades. Incorporating the torrance and sparrow model of reflectance in uncalibrated photometric stereo. In *IEEE International Conference on Computer Vision*, volume 02, pages 816–823, 2003.

[4] B. Goldluecke and M. Magnor. Weighted minimal hypersurfaces and their applications in computer vision. In *European Conference on Computer Vision*, volume 2, pages 366–378, 2004.

[5] H. Jin, D. Cremers, D. Wang, E. Prados, A. Yezzi, and S. Soatto. 3-d reconstruction of shaded objects from multiple images under unknown illumination. *To appear in the International Journal of Computer Vision*, 2007.

[6] H. Jin, D. Cremers, A. J. Yezzi, and S. Soatto. Shedding light on stereoscopic segmentation. *IEEE Conference on Computer Vision and Pattern Recognition*, 01:36–42, 2004.

[7] H. Jin, S. Soatto, and A. J. Yezzi. Multi-view stereo reconstruction of dense shape and complex appearance. *International Journal of Computer Vision*, 63(3):175–189, 2005.

[8] H. Jin, A. Yezzi, and S. Soatto. Variational multiframe stereo in the presence of specular reflections. In *International Symposium on 3D Data Processing Visualization and Transmission*, pages 626–630, 2002.

[9] 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. *Journal of Scientific Computing*, 19(1-3):267–292, 2003.

[10] J. Kim, V. Kolmogorov, and R. R. Zabih. Visual correspondence using energy minimization and mutual information. In *IEEE International Conference on Computer Vision*, pages 1033–1040, 2003.

[11] J. Lu and J. Little. Reflectance function estimation and shape recovery from image sequence of a rotating object. In *IEEE International Conference on Computer Vision*, pages 80–86, 1995.

[12] J.-P. Pons, R. Keriven, and O. Faugeras. Modelling dynamic scenes by registering multi-view image sequences. In *IEEE Conference on Computer Vision and Pattern Recognition*, volume 2, pages 822–827, 2005.

[13] J.-P. Pons, R. Keriven, and O. Faugeras. Multi-view stereo reconstruction and scene flow estimation with a global image-based matching score. *International Journal of Computer Vision*, 72(2):179–193, 2007.

[14] M. W. Powell, S. Sarkar, and D. Goldgof. A simple strategy for calibrating the geometry of light sources. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(9):1022–1027, 2001.

[15] D. Samaras, D. Metaxas, PascalFua, and Y. G. Leclerc. Variable albedo surface reconstruction from stereo and shape from shading. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 480–487, 2000.

[16] S. M. Seitz, B. Curless, J. Diebel, D. Scharstein, and R. Szeliski. A comparison and evaluation of multi-view stereo reconstruction algorithms. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 519–528, 2006.

[17] S. Soatto, A. J. Yezzi, and H. Jin. Tales of shape and radiance in multi-view stereo. In *IEEE International Conference on Computer Vision*, pages 974–981, 2003.

[18] J. E. Solem and N. C. Overgaard. A geometric formulation of gradient descent for variational problems with moving surfaces. In *Scale-Space*, pages 419–430, 2005.

[19] A. Vedaldi and S. Soatto. Viewpoint invariance for non-planar scenes. Technical Report TR050012, UCLA CSD, 2006.

[20] G. Vogiatzis, P. Favaro, and R. Cipolla. Using frontier points to recover shape, reflectance and illumunation. In *IEEE International Conference on Computer Vision*, volume 1, pages 228–235, 2005.

[21] R. Yang, M. Pollefeys, and G. Welch. Dealing with textureless regions and specular highlights-a progressive space carving scheme using a novel photo-consistency measure. In *IEEE International Conference on Computer Vision*, pages 576–583, 2003.

[22] A. Yezzi and S. Soatto. Stereoscopic segmentation. *International Journal of Computer Vision*, 53(1):31–43, 2003.

[23] K.-J. Yoon and I.-S. Kweon. Correspondence search in the presence of specular highlights using specular-free two-band images. In *Asian Conference on Computer Vision*, pages 761–770, 2006.

[24] T. Yu, N. Xu, and N. Ahuja. Recovering shape and reflectance model of non-lambertian objects from multiple views. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 226–233, 2004.

[25] T. Yu, N. Xu, and N. Ahuja. Shape and view independent reflectance map from multiple views. *International Journal of Computer Vision*, 73(2):123–138, 2007.

[26] W. Zhou and C. Kambhamettu. Estimation of illuminant direction and intensity of multiple light sources. In *European Conference on Computer Vision*, pages 206–220, 2002.

[27] T. Zickler. Reciprocal image features for uncalibrated helmholtz stereopsis. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1801–1808, 2006.

[28] T. Zickler, P. N. Belhumeur, and D. J. Kriegman. Helmholtz stereopsis: Exploiting reciprocity for surface reconstruction. *International Journal of Computer Vision*, 49(2-3):215–227, 2002.