HAL
open science

Generic Scene Recovery using Multiple Images

Kuk-Jin Yoon, Emmanuel Prados, Peter Sturm

To cite this version:

Kuk-Jin Yoon, Emmanuel Prados, Peter Sturm. Generic Scene Recovery using Multiple Images. SSVM 2009 - 2nd International Conference on Scale Space and Variational Methods in Computer Vision, Jun 2009, Voss, Norway. pp.745-757, 10.1007/978-3-642-02256-2_62 . inria-00357139

HAL Id: inria-00357139

https://inria.hal.science/inria-00357139v1

Submitted on 29 Jan 2009

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Generic Scene Recovery using Multiple Images

Kuk-Jin Yoon[1], Emmanuel Prados[2], and Peter Sturm[2]*

[1] Computer vision Lab., Dept. Information and Communications, GIST, Korea
[2] Perception Lab., INRIA Grenoble - Rhône-Alpes, France

Abstract. In this paper, a generative model based method for recovering both the shape and the reflectance of the surface(s) of a scene from multiple images is presented, assuming that illumination conditions are known in advance. Based on a variational framework and via gradient descents, the algorithm minimizes simultaneously and consistently a global cost functional with respect to both shape and reflectance. Contrary to previous works which consider specific individual scenarios, our method applies to a number of scenarios – mutiview stereovision, multiview photometric stereo, and multiview shape from shading. In addition, our approach naturally combines stereo, silhouette and shading cues in a single framework and, unlike most previous methods dealing with only Lambertian surfaces, the proposed method considers general dichromatic surfaces.

1 Introduction and Related Work

Many methods have been proposed to recover the three-dimensional surface shape using multiple images during these last two decades [1]. On the other hand, for a long time, the estimation of surface radiance/reflectance was secondary. Even some recent works [2–5] compute the 3D shape without considering radiance estimation. However, radiance/reflectance estimation has become a matter of concern in multiview reconstruction scenarios in the last decade [6–8]. Especially, recovering reflectance is required for realistic relighting, which is also fundamental in virtual reality as well as augmented reality. In addition, in real life applications, perfect Lambertian surfaces are rare and, therefore, multiview stereo algorithms have to be robust to specular reflection. Widespread ideas are to use appropriate similarity measures [2, 9, 10] and/or to modify input images in order to remove specular highlights [11, 12]. However, those similarity measures are not generally valid under general lighting conditions and these methods are strongly limited by the specific lighting configuration. Concerning the robustness to non-Lambertian effects, it is also worth to cite [6] which considers the radiance tensor. However, the radiance tensor presented in [6] is not appropriate when the images of the scene are taken under several (different) lighting conditions.

In this paper, our goal is to provide a model based method that simultaneously estimates shape and reflectance by combining stereo, silhouette, and shading cues in a single framework. The method we propose is robust to non-Lambertian effects by directly incorporating a specular reflectance model in the mathematical formulation of

* This work was supported by the Flamenco project (ANR-06-MDCA-007) and by the GIST Dasan project.

the problem. By incorporating a complete photometric image formation model, it also exploits prolifically all the photometric phenomena, as it is explicitly done in photometric stereo methods. Also, it allows to naturally deal with images taken under several lighting conditions. Let us note that actually there already exist recent works that provide solutions in this direction. [13] proposes a model-based method for recovering the 3D shape and the reflectance of a non-Lambertian object. Nevertheless, in this paper, the authors constrain the object to be made of a single textureless material — the parameters of the reflectance (in particular the albedo) are the same for all the points of the object surface. So, the method in [13] is a "multiview shape from shading" method, similarly as the one proposed by [8, 14] which focus on the Lambertian case. To our knowledge, with the exception of [15, 16], all the works going in the same direction as ours are limited to surfaces made of a single (textureless) material. In particular, this is the case for the photometric stereo methods proposed by [17, 18] and for the multiview photometric stereo work of [19]. Only the similar works [15, 16] are able to recover scenes with varying albedo. However, in [15, 16], the authors tried to filter out specular highlights by using a simple thresholding and to use only diffuse components to estimate the shape. [15] also used a thresholding to detect shadowed pixels that are not visible from light sources, which is however not working under multiple light sources.

In our work, we do not want to restrain ourselves to a single textureless material. (In return, we assume that lighting conditions are known in advance.) And, more generally, one of the goals of this paper is to show that the joint computation of shape and reflectance is beneficial from several points of view. In addition to providing the reflectance of the scene, this allows to naturally introduce specular models in the mathematical formulation of the multiview reconstruction problem; and thus the method to be robust to highlights. Without any additional effort, it is also possible to deal with a set of images lighted by several different conditions (which is not possible with radiance only). Moreover in such a case, the method allows to completely exploit the variations of the radiance according to the changes of illumination, as in photometric stereo. Finally, this allows to easily incorporate some constraints on the reflectance and so in particular to naturally exploit shading effects in textureless regions.

Here, let us emphasize that, contrary to previous works that consider specific scenarios, our method can be applied indiscriminately to a number of scenarios — multiview stereovision, multiview photometric stereo, and multiview shape from shading.

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. The foreground is composed by a set of (bounded and closed) 2D manifolds of $\mathbb{R}^3$ and represented by S.

Images are generated by n_c pinhole cameras. The perspective projection performed by the ith 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 ith camera. It is split into two parts: the pixels corresponding to the foreground, $\pi_{iF} = \pi_i \cap \Pi_i(S)$, and the other points $\pi_{iB} = \pi_i \setminus \pi_{iF}$. $I_i : \pi_i \to \mathbb{R}^c$ is the image of the true scene, captured by the ith camera[3]. I is the set of input images and I_{iF} and I_{iB} are the restrictions of the function I_i to π_{iF} and π_{iB}, respectively. In other respects,

the visibility function $v_S^i : \mathbb{R}^3 \to \mathbb{R}$ is defined by: $v_S^i(\mathbf{X}) = 1$ if $\mathbf{X}$ is visible from the ith camera and $v_S^i(\mathbf{X}) = 0$ if not. S_i denotes the part of S that is visible from the ith camera and $\Pi_{i,S}^{-1}$ is the back-projection from the ith camera onto S_i. We model the scene as being illuminated by a finite number of distant point light sources and an ambient light. n_{il} is the number of illuminants corresponding to the ith camera and $\mathbf{l}_{ij} \in \mathbb{S}^2$ and $L_{ij} \in \mathbb{R}^c$ are the direction and the intensity[3] of the jth illuminant of the ith camera, respectively. Similarly, $L_{ia} \in \mathbb{R}^c$ is the intensity[3] of the ambient illumination of the ith camera. $v_{L_{ij}} : \mathbb{R}^3 \to \mathbb{R}$ is the light visibility function: $v_{L_{ij}}(\mathbf{X}) = 1$ if the jth illuminant of the ith camera is visible from $\mathbf{X}$, $v_{L_{ij}}(\mathbf{X}) = 0$ otherwise.

We model the foreground object(s) by its shape S and its reflectance R. We denote $\Omega = (S, R)$. Contrary to most previous stereovision methods, we want to go beyond the Lambertian model. In order to get a solvable minimization problem without too many unknowns, we represent the reflectance by a parametric model. In this work, we consider the popular Blinn-Phong shading model. In this context, and 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 ith camera, the images I_i are decomposed as $I_i = I_{id} + I_{is} + I_{ia}$, where I_{id}, I_{is}, and I_{ia} are images with the diffuse, specular, and ambient reflection components of I_i, respectively. Here, diffuse reflection can be expressed by using the cosine law as $I_{id}(\mathbf{x}) = \sum_{j=1}^{n_{il}} v_{L_{ij}}(\mathbf{X})\Big(\rho_d(\mathbf{X})L_{ij}\left(\mathbf{n}(\mathbf{X}) \cdot \mathbf{l}_{ij}\right)\Big)$, where $\rho_d(\mathbf{X}) \in \mathbb{R}^c$ is the diffuse albedo[3] at point $\mathbf{X} \in S$, $\mathbf{n}(\mathbf{X})$ is the normal vector to the surface S at $\mathbf{X}$. On the other hand, specular reflection is expressed as $I_{is}(\mathbf{x}) = \sum_{j=1}^{n_{il}} v_{L_{ij}}(\mathbf{X}) \left(\rho_s(\mathbf{X})L_{ij}\big(\mathbf{n}(\mathbf{X}) \cdot \mathbf{h}_{ij}(\mathbf{X})\big)^{\alpha_s(\mathbf{X})}\right)$, where $\mathbf{h}_{ij}(\mathbf{X})$ is the bisector of the angle between the view of the ith camera and the jth illuminant at $\mathbf{X}$, $\rho_s(\mathbf{X}) \in \mathbb{R}^c$ and $\alpha_s(\mathbf{X}) \in \mathbb{R}^+$ are the specular albedo and the shininess parameter at point $\mathbf{X}$. The ambient illumination is assumed to be uniform and modeled as $I_{ia}(\mathbf{x}) = \rho_d(\mathbf{X})L_{ia}$, where L_{ia} is defined above. Finally, the image formation equation is given as

$$I_i(\mathbf{x}) = \sum_{j=1}^{n_{il}} v_{L_{ij}}(\mathbf{X})\mathbb{L}_{ij}(\mathbf{X}, \mathbf{n}(\mathbf{X})) + \rho_d(\mathbf{X})L_{ia}, \tag{1}$$

where $\mathbb{L}_{ij}(\mathbf{X}, \mathbf{n}(\mathbf{X})) = L_{ij}\rho_d(\mathbf{X})\big(\mathbf{n}(\mathbf{X}) \cdot \mathbf{l}_{ij}\big) + L_{ij}\rho_s(\mathbf{X})\big(\mathbf{n}(\mathbf{X}) \cdot \mathbf{h}_{ij}(\mathbf{X})\big)^{\alpha_s(\mathbf{X})}$. We denote $R = (R_d, R_s)$, where $R_d = \rho_d$ and $R_s = (\rho_s, \alpha_s)$.

As suggested by [20, 21], to be sure that the estimated foreground surface does not shrink to an empty set, it is crucial to define and characterize the background. In this work, we assume that we have the background images $\tilde{I} = \{\tilde{I}_1, \cdots, \tilde{I}_{n_c}\}$ and define $(\tilde{I}_{iF}, \tilde{I}_{iB})$ analogously to (I_{iF}, I_{iB}).

3 Bayesian Formulation of the Problem

Clearly, 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,

$$P(\Omega|I) = P(I|\Omega)\ P(\Omega)/P(I) \propto P(I|\Omega)\ P(\Omega) = P(I|S, R)\ P(S)\ P(R) \tag{2}$$

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

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

When Π_i is given, we can produce a synthetic image $\bar{I}_i(\Omega)$ corresponding to I_i by using the current estimation of Ω. This allows us to measure the validity of the current estimation by comparing input images with generated ones. When assuming an independent identical distribution of observations, the likelihood can be expressed as $P(I|\Omega) \propto \prod_{i=1}^{n_c} \exp\big(-\xi_i(\Omega)\big) = \prod_{i=1}^{n_c} \exp\big(-\xi(I_i, \bar{I}_i(\Omega))\big)$, where $\xi_i(\Omega) = \xi(I_i, \bar{I}_i(\Omega))$ is a function of Ω, measuring image dissimilarity.

A typical and reasonable prior for the surface shape S is about the area given as $P(S) \propto \exp\big(-\psi(S)\big)$. Here, $\psi(S)$ is the monotonic increasing function of the surface area $\int_S d\sigma$ where $d\sigma$ is the Euclidean surface measure.

In other respects, a prior on the reflectance is also required because there are not enough observations exhibiting specular reflection at every surface point. To overcome the lack of observations, we assume that specular reflectance varies smoothly within each homogeneous material surface patch. This prior is clearly reasonable in real life applications and in common scenes. Thus, in this work, we use the diffuse reflectance of a surface as a soft constraint to partition Ω and define the prior on the surface reflectance as $P(R) \propto \exp\big(-\omega(R)\big)$, where $\omega(R)$ will be defined later.

4 Description of the Cost Functions

Based on the section 3, the problem can be expressed in terms of cost functions as $E_{total}(\Omega) = E_{data}(\Omega) + E_{shape}(S) + E_{refl}(R) = \sum_{i=1}^{n_c} \xi_i(\Omega) + \psi(S) + \omega(R)$. Maximizing the probability $P(\Omega|I)$ is equivalent to minimizing the total cost.

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, we use actual background images, i.e. $\bar{I}_{iB}$=$\tilde{I}_{iB}$. Also, as suggested by [20], $\xi_i(\Omega) = \xi(I_i, \bar{I}_i)$ is then rewritten as

$$\xi(I_i, \bar{I}_i) = \xi_F(I_{iF}, \bar{I}_{iF}) + \xi_B(I_{iB}, \bar{I}_{iB}) = \hat{\xi}_F(I_{iF}, \bar{I}_{iF}) + \xi(I_i, \tilde{I}_i), \tag{3}$$

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 $E_{data}(\Omega) = \sum_{i=1}^{n_c} \hat{\xi}_F(I_{iF}, \bar{I}_{iF}) + C$, where $C = \sum_{i=1}^{n_c} C_i = \sum_{i=1}^{n_c} \xi(I_i, \tilde{I}_i)$ is constant.

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 $\xi(I_i, \bar{I}_i) = \int_{\pi_i} e_i(\mathbf{x}) d\sigma_i$, where $d\sigma_i$ is the surface measure and $e_i(\mathbf{x})$ is the contribution at $\mathbf{x}$ to ξ_i. The data cost function is then given as

$$E_{data}(\Omega) = \sum_{i=1}^{n_c} \int_{\pi_{iF}} \hat{e}_i(\mathbf{x}) d\sigma_i + C, \tag{4}$$

where $\hat{e}_i(\mathbf{x}) = e_i\big(I_i(\mathbf{x}), \bar{I}_i(\mathbf{x})\big) - e_i\big(I_i(\mathbf{x}), \tilde{I}_i(\mathbf{x})\big)$.

Decoupling Appearance from Surface Normal As shown in Eq. (1), surface appearance 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 a photometric unit vector field $\mathbf{v}$ satisfying $\|\mathbf{v}\| = 1$ as in [14], which is used for the computation of surface appearance. To penalize the deviation between the actual normal vector $\mathbf{n}$ and the photometric normal vector $\mathbf{v}$, we add a new term

$$E_{dev}(\Omega) = \tau \int_S \chi(\mathbf{X}) d\sigma = \tau \int_S (1 - (\mathbf{n}(\mathbf{X}) \cdot \mathbf{v}(\mathbf{X}))) \, d\sigma, \tag{5}$$

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

Shape Area and Reflectance Discontinuity Cost Functions By using the area of a surface for the prior, $E_{shape}(S)$ is simply defined as $E_{shape}(S) = \psi(S) = \lambda \int_S d\sigma$, where λ is a control constant. Based on the assumption in section 3, we define a discontinuity cost function of surface reflectance, which makes the discontinuities of specular reflectance generally coincide with the discontinuities of diffuse reflectance, as

$$E_{refl}(R) = \omega(R) = \beta \int_S f(\mathbf{X}) d\sigma = \beta \int_S \zeta\big(R_d(\mathbf{X})\big) \times \eta\big(R_s(\mathbf{X})\big) d\sigma, \tag{6}$$

where β is a control constant, and $\zeta\big(R_d(\mathbf{X})\big)$ and $\eta\big(R_s(\mathbf{X})\big)$ are defined as

$$\zeta\big(R_d(\mathbf{X})\big) = \left(1 - \frac{\|\nabla_S R_d(\mathbf{X})\|^2}{M}\right), \eta\big(R_s(\mathbf{X})\big) = \big(\|\nabla_S \rho_s(\mathbf{X})\|^2 + \gamma \|\nabla_S \alpha_s(\mathbf{X})\|^2\big) \tag{7}$$

with a pre-defined constant M.[4] ∇_S denotes the intrinsic gradient defined on S. By using the proposed discontinuity cost function of surface reflectance, surface points that do not have enough specular observations get assigned specular reflectance inferred from the specular reflectance of neighboring surface points.

Total Cost Function Finally, the total cost function is given by

$$E_{total}(\Omega) = C + \sum_{i=1}^{n_c} \int_{\pi_{iF}} \hat{e}_i(\mathbf{x}) d\sigma_i + \tau \int_S \chi(\mathbf{X}) d\sigma + \lambda \int_S d\sigma + \beta \int_S f(\mathbf{X}) d\sigma. \tag{8}$$

Here, it is worthy of notice that $E_{dev}(\Omega)$, $E_{shape}(S)$, and $E_{refl}(R)$ are defined over the scene surface while $E_{data}(\Omega)$ is defined as an integral over the image plane. By the change of variable, $d\sigma_i = -\frac{\mathbf{d}_i(\mathbf{X}) \cdot \mathbf{n}(\mathbf{X})}{z_i(\mathbf{X})^3} d\sigma$, 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 [7]. When denoting $g(\mathbf{X}, \mathbf{n}(\mathbf{X})) : \mathbb{R}^3 \times \Omega \to \mathbb{R}$ as

$$g(\mathbf{X}, \mathbf{n}(\mathbf{X})) = \left(- \sum_{i=1}^{n_c} \left(v_S^i \hat{e}_i \frac{\mathbf{d}_i \cdot \mathbf{n}}{{z_i}^3} \right) + \tau\chi + \lambda + \beta f \right), \tag{9}$$

Eq. (8) is simply rewritten as $E_{total}(\Omega) = C + \int_S g(\mathbf{X}, \mathbf{n}(\mathbf{X})) d\sigma$.

[4] Be sure that $M \geq 3$ for gray-level images and $M \geq 9$ for color images.

5 Scene Recovery

Recently, via graph cuts or convexity, some authors have proposed some global optimization methods for the classical multiview stereovision problem [5,22,23]. Nevertheless, because of the presence of the normal but also of the visibility in the cost function, the state of the art in optimization does not allow to compute the global minimum of the energy we have designed. Also, here, scene recovery is achieved by minimizing E_{total} via gradient descents. In other respects, S and R being highly coupled, it is very complicated to estimate all unknowns simultaneously. To solve the problem, we adopt an alternating scheme, updating S for a fixed R and then R for a fixed S.

5.1 Shape Estimation – Surface Evolution

When assuming that R is given, E_{total} is a function of S. In this work, we derive the gradient descent flows corresponding to the cost functions respectively. The final gradient descent flow is then given by $S_t = \Big(S_t\big|_{data} + S_t\big|_{dev} + S_t\big|_{shape} + S_t\big|_{refl} \Big)$, where $S_t\big|_{data}$, $S_t\big|_{dev}$, $S_t\big|_{shape}$ and $S_t\big|_{refl}$ are described below.

The data cost is a function of the visibility of a surface point, which is dependent on the whole surface shape. According to [20, 21] for correctly dealing with the visibility of non-convex objects, $S_t\big|_{data}$ is given by

$$S_t\big|_{data} = \sum_{i=1}^{n_c} \Big(- \frac{v_S^i \, (\hat{e}_i - \hat{e}_i')}{z_i^3} \left(\mathbf{d}_i^t \nabla \mathbf{n} \mathbf{d}_i^t \delta(\mathbf{d}_i \cdot \mathbf{n}) \right) \; + \; \frac{v_S^i}{z_i^3} \left(\left(\partial_2 \hat{e}_i \nabla \bar{I}_i \right) \cdot \mathbf{d}_i \right) \Big), \quad (10)$$

where $\delta(\cdot)$ is the delta function and $\hat{e}_i'$ is an error computed by using the radiance at point $\mathbf{X}'$ in the direction of the ith camera, which is the terminator of a horizon point $\mathbf{X}$ [21]. 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. (1) as $\nabla \bar{I}_i = \sum_{j=1}^{n_{il}} \{ (\nabla v_{L_{ij}}) \mathbb{L}_{ij} + v_{L_{ij}} (\nabla \mathbb{L}_{ij}) \} + (\nabla \rho_a) L_{ia}$. This gradient descent flow includes both the variation related to the camera visibility changes (the first term) and the variation related to the image changes (the second term), which also includes the variation due to the light visibility changes.

In addition, similarly as [8, 14], the gradient descent flows for the normal deviation cost $S_t\big|_{dev}$ (originating from $E_{dev}(\Omega)$) is $S_t\big|_{dev} = (-2\tau H + \tau(\nabla \cdot \mathbf{v}))$. Also $S_t\big|_{shape}$ (from $E_{shape}(S)$) is the mean curvature flow as $S_t\big|_{shape} = -2\lambda H$.

Due to the complexity of the discontinuity cost function of surface reflectance, it needs more attention to derive the gradient descent flow. By using the derivation in [24], we get the following equation for surface evolution.

$$S_t\big|_{refl} = -2\beta \left(\frac{1}{M} m(\rho_d) \eta(R_s) \; - (m(\rho_s) + \gamma m(\alpha_s)) \, \zeta(R_d) \right). \quad (11)$$

Here, $m(y) = \Big(\mathrm{II}\big(\nabla_S \, y \times \mathbf{n}\big) + \|\nabla_S \, y\|^2 H \Big)$, where $\mathrm{II}(\mathbf{t})$ is the second fundamental form for a tangent vector $\mathbf{t}$ with respect to $\mathbf{n}$.

5.2 Photometric Unit Vector Field Update

The computed gradient descent flows minimize the total cost with respect to given reflectance and $\mathbf{v}$. We then update the photometric unit 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 equation $\frac{\partial g}{\partial \mathbf{v}} = \left(-\sum_{i=1}^{n_c} v_S^i \partial_2 \hat{e}_i \frac{\partial \bar{I}_i}{\partial \mathbf{v}} \frac{\mathbf{d}_i \cdot \mathbf{n}}{{z_i}^3}\right) + (-\tau \mathbf{n}) = 0$. Here, we have to keep $\|\mathbf{v}\| = 1$. Since $\mathbf{v} \in \mathbb{S}^2$, $\mathbf{v}$ can be expressed in spherical coordinates as $[\cos\theta_v \sin\phi_v, \sin\theta_v \sin\phi_v, \cos\phi_v]^{\mathrm{T}}$ where θ_v and ϕ_v are the coordinates of $\mathbf{v}$. Therefore, we update θ_v and ϕ_v to update $\mathbf{v}$. As before, the θ_v and ϕ_v that minimize the total cost satisfy the following two equations by the chain rule.

$$\frac{\partial g}{\partial \theta_v} = \frac{\partial g}{\partial \mathbf{v}} \cdot \frac{\partial \mathbf{v}}{\partial \theta_v} = 0, \quad \frac{\partial g}{\partial \phi_v} = \frac{\partial g}{\partial \mathbf{v}} \cdot \frac{\partial \mathbf{v}}{\partial \phi_v} = 0 \tag{12}$$

So, we update $\mathbf{v}$ by performing gradient descent using above two PDEs.

5.3 Reflectance Estimation

Here, we estimate R for fixed S and $\mathbf{v}$, 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 $(E_{data}(\Omega) + E_{refl}(R))$. Here, because it is also complex to estimate diffuse and specular reflectance at the same time due to the high coupling between them, we alternatively estimate surface reflectance one by one while assuming that the rest are given and fixed. We repeat the procedure until they no longer change.

Diffuse Reflectance Estimation For given S and R_s, we estimate ρ_d that minimizes the cost $(E_{data} + E_{refl})$. Here, ρ_d that minimizes the total cost function will satisfy the Euler-Lagrange equation given as $-\sum_{i=1}^{n_c} v_S^i \partial_2 \hat{e}_i \frac{\partial \bar{I}_i}{\partial \rho_d} \frac{\mathbf{d}_i \cdot \mathbf{n}}{{z_i}^3} + \frac{2\beta}{M} \eta(R_s) \Delta_S \rho_d = 0$, where Δ_S denotes the Laplace-Beltrami operator defined on the surface S. We solve the PDE by performing gradient descent using the following PDE:

$$\frac{\partial \rho_d}{\partial t} = \left(-\sum_{i=1}^{n_c} v_S^i \partial_2 \hat{e}_i \frac{\partial \bar{I}_i}{\partial \rho_d} \frac{\mathbf{d}_i \cdot \mathbf{n}}{{z_i}^3}\right) + \left(\frac{2\beta}{M} \eta(R_s)\right) \Delta_S \rho_d. \tag{13}$$

Specular Reflectance Estimation We then estimate $R_s = (\rho_s, \alpha_s)$ for given S and R_d in the same manner. ρ_s that minimizes the total cost function will satisfy the Euler-Lagrange equation given as $\left(-\sum_{i=1}^{n_c} v_S^i \partial_2 \hat{e}_i \frac{\partial \bar{I}_i}{\partial \rho_s} \frac{\mathbf{d}_i \cdot \mathbf{n}}{{z_i}^3}\right) - 2\beta(\Delta_S \rho_s)\zeta(\rho_d) = 0$. We again solve the PDE by performing gradient descent using the following PDE.

$$\frac{\partial \rho_s}{\partial t} = -\sum_{i=1}^{n_c} \left(v_S^i \partial_2 \hat{e}_i \frac{\partial \bar{I}_i}{\partial \rho_s} \frac{\mathbf{d}_i \cdot \mathbf{n}}{{z_i}^3}\right) - 2\beta(\Delta_S \rho_s)\zeta(\rho_d). \tag{14}$$

α_s is also estimated in the same manner by solving the PDE as

$$\frac{\partial \alpha_s}{\partial t} = -\sum_{i=1}^{n_c} \left(v_S^i \partial_2 \hat{e}_i \frac{\partial \bar{I}_i}{\partial \alpha_s} \frac{\mathbf{d}_i \cdot \mathbf{n}}{{z_i}^3}\right) - 2\beta\gamma(\Delta_S \alpha_s)\zeta(\rho_d). \tag{15}$$

Single-Material Surface Case When dealing with a single-material surface that has a single specular reflectance R_s, the discontinuity cost function of surface reflectance, $E_{refl}(R)$, can be excluded because $f(\mathbf{X})$ is zero everywhere on the surface. The PDE used for the ρ_d estimation, Eq. (13), is then simplified as $\frac{\partial \rho_d}{\partial t} = -\sum_{i=1}^{n_c} v_S^i \partial_2 \hat{e}_i \frac{\partial \bar{I}_i}{\partial \rho_d} \frac{\mathbf{d}_i \cdot \mathbf{n}}{{z_i}^3}$. ρ_s and α_s are also computed by performing gradient descent using PDEs given as

$$\frac{\partial \rho_s}{\partial t} = \int_S -\sum_{i=1}^{n_c} v_S^i \partial_2 \hat{e}_i \frac{\partial \bar{I}_i}{\partial \rho_s} \frac{\mathbf{d}_i \cdot \mathbf{n}}{{z_i}^3} d\sigma, \frac{\partial \alpha_s}{\partial t} = \int_S -\sum_{i=1}^{n_c} v_S^i \partial_2 \hat{e}_i \frac{\partial \bar{I}_i}{\partial \alpha_s} \frac{\mathbf{d}_i \cdot \mathbf{n}}{{z_i}^3} d\sigma. \quad (16)$$

6 Experiments

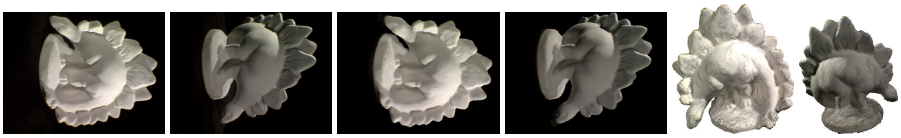
We have implemented the gradient descent surface evolution in the level set framework. The proposed method starts with the visual hull obtained by rough silhouette images to reduce computational time and to avoid local minima. We also adopt a multi-scale strategy. 640×480 or 800×600 images were used as inputs and the simple L^2-norm was used to compute the image similarity, e.

For synthetic data sets, the estimated shape is quantitatively evaluated in terms of accuracy and completeness as in [1]. We used 95% for accuracy and the 1.0mm error for completeness. For easy comprehension, the size of a target object is normalized so that it is smaller than [100mm 100mm 100mm]. Here, beside the shape evaluation, we also evaluated the estimated reflectance in the same manner. In addition, we computed the average difference between input images and synthesized images as $e_{image} = \frac{1}{n_c} \sum_{i=1}^{n_c} \frac{1}{A} \int_{\pi_i} \| \big(I_i(\mathbf{x}) - \bar{I}_i(\mathbf{x}) \big) \| d\sigma_i$, where $A = \int_{\pi_i} d\sigma_i$.

Due to the generality of the proposed method, it can be applied to various types of image sets with different camera/light configurations. Here, knowledge of illumination allows to factorize radiance into reflectance and geometry. In practice, depending on the scenario, that knowledge may not be required, e.g. for recovering shape and radiance of Lambertian surfaces with static illumination. In this case, the proposed method can be applied even without lighting information, assuming only an ambient illumination, and the proposed method works much like the conventional multiview stereo methods. Figure 1 shows the result for the "dino" image set [1], for which no lighting information is required. The proposed method successfully recovers the shape as well as the radiance.

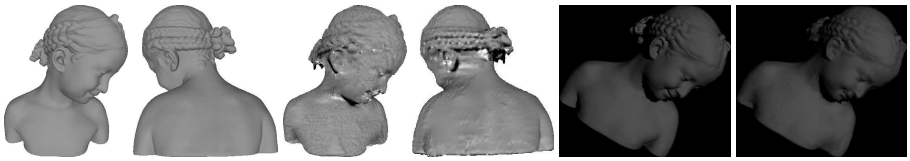
The proposed method can also be applied to images taken under varying illumination. Results using images of textureless/textured Lambertian surfaces are shown in Fig. 2 and Fig 3. In the case of Fig. 2, the proposed method works as a multiview photometric stereo method and recovers the shape and the diffuse reflectance of each surface point. Based on these, we can synthesize images of the scene for different lighting conditions

We then applied our method to the images of textureless/textured non-Lambertian surfaces showing specular reflection. Note that, unlike [15, 16], we do not use any thresholding to filter out specular highlight pixels. The result for the smoothed "bimba" data set is shown in Fig. 4. In this case, the surface has uniform diffuse/specular reflectance and each image was taken under a different illumination. Here, we used the method with Eq. (16) to estimate the specular reflectance. Although there is high-frequency noise in the estimated shape, the proposed method estimates the specular reflectance well — the ground-truth specular reflectance is (ρ_s=0.7, α_s=50) while the



(a) input images (b) synthesized images (c) results

Fig. 1. "dino" image set (16 images) — Lambertian surface case (static illumination).



(a) ground-truth model (b) estimated model (c) input vs. synthesized image

Fig. 2. "bimba" image set (18 images) — textureless Lambertian surface case (varying illumination and viewpoint). 95% accuracy (shape, ρ_{dr}, ρ_{dg}, ρ_{db})=(2.16mm, 0.093, 0.093, 0.093), 1.0mm completeness (shape, ρ_{dr}, ρ_{dg}, ρ_{db}) = (82.63%, 0.104, 0.104, 0.104), e_{image}=1.44.

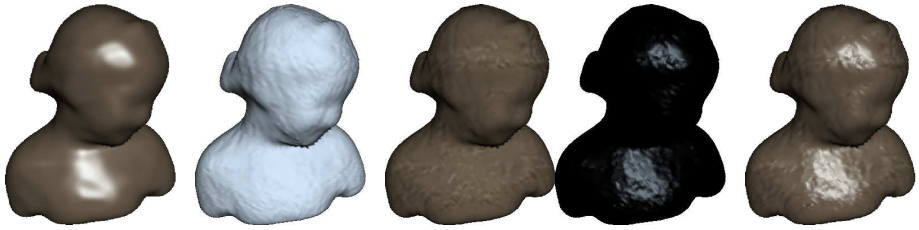
(a) input image (b) true refl. (c) true shading (d) est. refl. (e) est. shading

Fig. 3. "dragon" image set (32 images) — textured Lambertian surface case (static illumination and varying viewpoint). 95% accuracy (shape, ρ_{dr}, ρ_{dg}, ρ_{db})=(1.28mm, 0.090, 0.073, 0.066), 1.0mm completeness (shape, ρ_{dr}, ρ_{dg}, ρ_{db}) = (97.11%, 0.064, 0.056, 0.052), e_{image}=1.25.

estimated one is (ρ_s=0.61, α_s=41.8). Here, note that small errors in estimated surface normals can cause large errors in specular reflectance due to its sensitivity to the surface normal. For instance, $0.7 \times (0.98)^{50} (= 0.255) \approx 0.61 \times (0.979)^{41.8} (= 0.251)$.

Note that most previous methods do not work for image sets taken under varying illumination and, moreover, they have difficulties to deal with specular reflection even if the images are taken under static illumination. For example, Fig. 5 shows results obtained by the method of [2] and our result for comparison. We ran the original code provided by the authors many times while changing parameters and used mutual information (MI) and cross correlation (CCL) as similarity measures to get the best results under specular reflection. As shown, the method of [2] fails to get a good shape even when the shape is very simple, while our method estimates it accurately. Also, with such images, given the large proportion of overbright surface parts, it seems intuitive that the strategy chosen by [16] and [15] (who consider bright pixels as outliers) might return less accurate results, because it removes too much information.

We also used real image sets of textured glossy objects, which were taken by using fixed cameras/light sources, while rotating the objects as in [15, 16]. Here, we simply



(a) true model (b) est. shape (c) diffuse & specular images (d) synthesized

Fig. 4. Smoothed "bimba" image set (36 images) — textureless non-Lambertian surface case (uniform specular reflectance, varying illumination and viewpoint). 95% accuracy (shape, ρ_{dr}, ρ_{dg}, ρ_{db}, ρ_s, α_s)=(0.33mm, 0.047, 0.040, 0.032, 0.095, 8.248), 1.0mm completeness (shape, ρ_{dr}, ρ_{dg}, ρ_{db}, ρ_s, α_s) = (100%, 0.048, 0.041, 0.032, 0.095, 8.248), e_{image}=1.63.

(a) two input images (b) results using [2] (MI, CCL) (c) our result

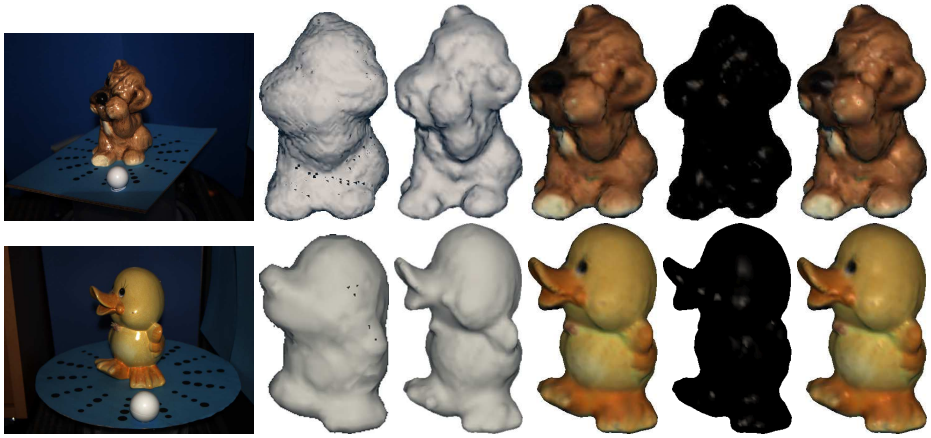
Fig. 5. Comparison using the "ellipse" image set (16 images) — textureless non-Lambertian surface case (uniform specular reflectance, static illumination and varying viewpoint).

assumed a single-material surface. $(72 \times 72 \times 72)$ grids were used for the "saddog" (59 images) and "duck" (26 images) image sets. Figure 6 shows that, although sparse grid volumes were used, the proposed method successfully estimated the shape of the glossy object even under specular reflection while estimating specular reflectance. In addition, although the estimated specular reflectance may not be so accurate because of the inaccuracy of lighting calibration, saturation, and some unexpected photometric phenomenon such as interreflection, it really helps to recover the shape well.

Finally, we applied our method to the most general case — textured non-Lambertian surfaces with spatially varying diffuse and specular reflectance and shininess, cf. Fig. 7. $(64 \times 125 \times 64)$ grids were used in this case. We can see that the proposed method yields plausible specular/diffuse images and shape. However, there is high-frequency noise in the estimated shape. Moreover, the error in reflectance estimation is rather larger compared to the previous cases. This result shows that, although the proposed discontinuity cost function of surface reflectance helps to infer the specular reflectance of all points with sparse specular reflection observation, reliably estimating specular reflectance for all surface points is still difficult unless there are enough observations of specular reflection for every surface point.

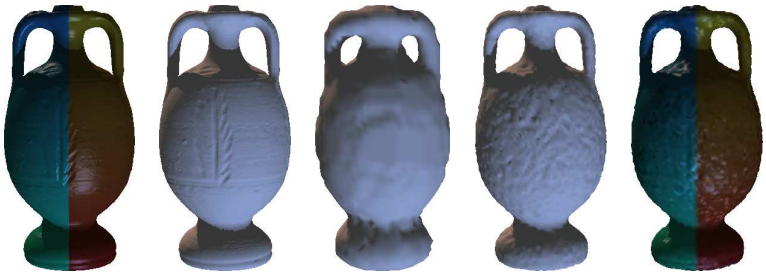
7 Conclusion

In this paper, we have presented a variational method that recovers both the shape and the reflectance of surfaces using multiple images. Scene recovery is achieved by min-



(a) (b) (c) (d) (e) (f)

Fig. 6. Result for real image sets. (a) input image (b) initial shape (c) estimated shape (d) diffuse image (e) specular image (f) synthesized image



(a) input image (b) true shading (c) init. shape (d) est. shading (e) syn. image

Fig. 7. Result for the "amphora" image set (36 images) — textured non-Lambertian surface case (spatially varying specular reflectance, static illumination, and varying viewpoint). 95% accuracy (shape, ρ_{dr}, ρ_{dg}, ρ_{db}, ρ_s, α_s)=(0.59mm, 0.041, 0.047, 0.042, 0.226, 13.59), 1.0mm completeness (shape, ρ_{dr}, ρ_{dg}, ρ_{db}, ρ_s, α_s) = (89.73%, 0.042, 0.047, 0.042, 0.226, 13.55), e_{image}=1.99.

imizing a global cost functional by alternation. As a result, the proposed method produces a complete description of scene surfaces. Contrary to previous works that consider specific scenarios, our method can be applied indiscriminately to a number of classical scenarios — it naturally fuses and exploits several important cues (silhouettes, stereo, and shading) and allows to deal with most of the classical 3D reconstruction scenarios such as stereo vision, (multi-view) photometric stereo, and multiview shape from shading. In addition, our method can deal with strong specular reflection, which is difficult even in some other state of the art methods using complex similarity measures.

References

1. Seitz, S.M., Curless, B., Diebel, J., Scharstein, D., Szeliski, R.: A comparison and evaluation of multi-view stereo reconstruction algorithms. In: IEEE CVPR. (2006) 519–528
2. Pons, J.P., Keriven, R., Faugeras, O.: Multi-view stereo reconstruction and scene flow estimation with a global image-based matching score. IJCV **72**(2) (2007) 179–193
3. Goesele, M., Curless, B., Seitz, S.M.: Multi-view stereo revisited. In: IEEE CVPR. Volume 02. (2006) 2402–2409
4. Tran, S., Davis, L.: 3d surface reconstruction using graph cuts with surface constraints. In: ECCV. Volume 2. (2006) 219–231
5. Kolev, K., Klodt, M., Brox, T., Esedoglu, S., Cremers, D.: Continuous global optimization in multiview 3d reconstruction. In: EMMCVPR. Volume 4679. (August 2007) 441–452
6. Jin, H., Soatto, S., Yezzi, A.J.: Multi-view stereo reconstruction of dense shape and complex appearance. IJCV **63**(3) (2005) 175–189
7. Soatto, S., Yezzi, A.J., Jin, H.: Tales of shape and radiance in multi-view stereo. In: IEEE ICCV. (2003) 974–981
8. Jin, H., Cremers, D., Wang, D., Prados, E., Yezzi, A., Soatto, S.: 3-d reconstruction of shaded objects from multiple images under unknown illumination. IJCV **76**(3) (2008) 245–256
9. Jin, H., Yezzi, A., Soatto, S.: Variational multiframe stereo in the presence of specular reflections. In: 3DPVT. (2002) 626–630
10. Yang, R., Pollefeys, M., Welch, G.: Dealing with textureless regions and specular highlights-a progressive space carving scheme using a novel photo-consistency measure. In: IEEE ICCV. (2003) 576–583
11. Yoon, K.J., Kweon, I.S.: Correspondence search in the presence of specular highlights using specular-free two-band images. In: ACCV. (2006) 761–770
12. Zickler, T., Mallick, S.P., Kriegman, D.J., Belhumeur, P.: Color subspaces as photometric invariants. To appear in IJCV (2008)
13. Yu, T., Xu, N., Ahuja, N.: Shape and view independent reflectance map from multiple views. IJCV **73**(2) (2007) 123–138
14. Jin, H., Cremers, D., Yezzi, A.J., Soatto, S.: Shedding light on stereoscopic segmentation. In: IEEE CVPR. Volume 01. (2004) 36–42
15. Esteban, C.H., Vogiatzis, G., Cipolla, R.: Multiview photometric stereo. IEEE TPAMI **30**(3) (2008) 548–554
16. Birkbeck, N., Cobzas, D., Sturm, P., Jägersand, M.: Variational shape and reflectance estimation under changing light and viewpoints. In: ECCV. Volume 1. (2006) 536–549
17. Georghiades, A.S.: Incorporating the torrance and sparrow model of reflectance in uncalibrated photometric stereo. In: IEEE ICCV. Volume 02. (2003) 816–823
18. Vogiatzis, G., Favaro, P., Cipolla, R.: Using frontier points to recover shape, reflectance and illumunation. In: IEEE ICCV. Volume 1. (2005) 228–235
19. Lu, J., Little, J.: Reflectance function estimation and shape recovery from image sequence of a rotating object. In: IEEE ICCV. (1995) 80–86
20. Yezzi, A., Soatto, S.: Stereoscopic segmentation. IJCV **53**(1) (2003) 31–43
21. Gargallo, P., Prados, E., Sturm, P.: Minimizing the reprojection error in surface reconstruction from images. In: IEEE ICCV. (2007)
22. Paris, S., Sillion, F.X., Quan, L.: A surface reconstruction method using global graph cut optimization. IJCV **66**(2) (2006) 141–161
23. Vogiatzis, G., Esteban, C.H., Torr, P.H.S., Cipolla, R.: Multiview stereo via volumetric graph-cuts and occlusion robust photo-consistency. IEEE TPAMI **29**(12) (2007) 2241–2246
24. Jin, H., Yezzi, A.J., Tsai, Y.H., Cheng, L.T., Soatto, S.: Estimation of 3D surface shape and smooth radiance from 2D images: A level set approach. J. Sci. Comput. **19**(1-3) (2003) 267–292