Joint Estimation of Shape and Reflectance using Multiple Images with Known Illumination Conditions

Kuk-Jin Yoon · Emmanuel Prados · Peter Sturm

Abstract We propose a generative model based method for recovering both the shape and the reflectance of the surface(s) of a scene from multiple images, assuming that illumination conditions and cameras calibration 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. The motivations for our approach are threefold. (1) Contrary to previous works which mainly consider specific individual scenarios, our method applies indiscriminately to a number of classical scenarios; in particular it works for classical stereovision, multiview photometric stereo and multiview shape from shading. It works with changing as well as static illumination. (2) Our approach naturally combines stereo, silhouette and shading cues in a single framework. (3) Moreover, unlike most previous methods dealing with only Lambertian surfaces, the proposed method considers general dichromatic surfaces. We verify the method using various synthetic and real data sets.

Keywords 3D reconstruction · Reflectance estimation · Multiview stereo · Photometric stereo · Multiview shape from shading

K.-J. Yoon (✉)
Room C508, Computer Vision Laboratory, Department of Information and Communications, Gwangju Institute of Science and Technology (GIST), 261 Cheomdan-gwagiro (Oryong-dong), Buk-gu, Gwangju 500-712, Republic of Korea
e-mail: kjyoon@gist.ac.kr

E. Prados · P. Sturm
Perception Team, INRIA Grenoble, Montbonnot, Rhône Alpes, France

E. Prados
e-mail: Emmanuel.Prados@inrialpes.fr

P. Sturm
e-mail: Peter.Sturm@inrialpes.fr

1 Introduction and Related Work

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 Seitz et al. (2006) for an evaluation of various recent methods. On the other hand, for a long time, the estimation of surface radiance/reflectance was secondary and was mainly of use to set up the shape reconstruction task (Faugeras and Keriven 1998; Zickler 2006; Zickler et al. 2002). Even some very recent works (Pons et al. 2005, 2007; Goesele et al. 2006; Zach et al. 2006; Tran and Davis 2006; Kolev et al. 2007b, 2007a) 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. For example, Jin, Soatto et al. estimate conjointly the 3D shape and radiance (tensors) (see Jin et al. 2003, 2005; Soatto et al. 2003; Yezzi and Soatto 2003), or the 3D shape and the (piecewise constant) albedo of a Lambertian surface (Jin et al. 2008).

Here, radiance is a combination of lighting, surface reflectance, and the geometry of a scene. In other words, radiance contains shading and shadows and, from raw radiance, it is impossible to correct them when changing the lighting. Therefore, recovering reflectance is required for realistic relighting, which is also fundamental, for example, in virtual reality as well as augmented reality where the lighting con-

ditions when re-synthesizing a scene may be different from the lighting conditions when capturing a scene.

In addition, in real life applications, perfect Lambertian surfaces are rare and, therefore, multiview stereo algorithms have to be *robust to specular reflection*. Many ideas have been exploited to improve the robustness of the algorithms. A widespread idea is to use appropriate *similarity measures* as in Faugeras and Keriven (1998), Jin et al. (2002), Kim et al. (2003), Pons et al. (2005, 2007), Yang et al. (2003), Yoon and Kweon (2006). However, those similarity measures are not generally valid under general lighting conditions and/or not physically motivated. Another common strategy is to modify input images in order to *remove specular highlights* so as to obtain images as if the original surfaces had been purely Lambertian; see Yoon and Kweon (2006), Mallick et al. (2005), Zickler et al. (2008). These methods are based on the well known Neutral Interface Reflection (NIR) assumption (Lee et al. 1990) which supposes that the spectral energy distribution of specular reflection components is similar to the spectral energy distribution of incident light. Nevertheless, these methods are strongly limited by the specific lighting configuration. For example, the method in Yoon and Kweon (2006) is valid only for a single (uniformly) colored illumination. Although Zickler et al. (2008) recently showed that it is always possible to represent an image with $(M - N)$ specularity-independent color channels, where M is the number of color channels of an image and N is the number of different illuminant colors, their image representation may only work with up to two different illuminant colors because images have three color channels in general. Similarly, some authors do not consider the image pixels that potentially have specular reflection components—these are treated as outliers (Hernández Esteban et al. 2008; Birkbeck et al. 2006). The idea in this approach is to work only on data that one is able to model well (and so to ignore what is too complicated to model). The authors have then to increase the amount of data (i.e. the number of input images) in order to compensate for the loss of information. Thus, this strategy cannot be applied to two-frame stereo.

On the other hand, Bhat and Nayar (1998) analyzed the physics of specular reflection and the geometry of stereopsis to reduce errors due to non-Lambertian surfaces, which leads to a relationship between stereo vergence, surface roughness, and the likelihood of a correct match. Zickler et al. (2002) presented the Helmholtz stereopsis to overcome the specular reflection problem. However, these two approaches require *specialized camera/lighting configurations*. Concerning the robustness to non-Lambertian effects, it also worth to cite the work of Jin et al. (2005) which considers the so-called *radiance tensor*. However, although some similarity measures such as normalized cross correlation (Faugeras and Keriven 1998; Pons et al. 2005, 2007) could help to be robust to some illumination changes, the radiance tensor presented in Jin et al. (2005) is *not appropriate*

when images of the scene are taken under several different lighting conditions.

In this paper, we propose a method for jointly estimating the shape and the reflectance of scene surfaces from multiple images. This can also be understood as the separation of geometry, reflectance, and illumination from radiance. In fact, our goal is to provide a shape and reflectance estimation method that is global (in the sense that it simultaneously and consistently optimizes shape and reflectance) and completely model based. *The method we propose is robust to non-Lambertian effects by directly incorporating a specular reflectance model* in the mathematical formulation of the problem. By incorporating a complete photometric image formation model, *it also advantageously exploits photometric phenomena*, as is explicitly done in photometric stereo methods. Furthermore, it allows to naturally deal with a set of images taken under several lighting conditions.

Some recent works already provide solutions in this direction. Goldman et al. (2005) present a relevant *photometric stereo* technique that simultaneously recovers shape and spatially-varying reflectance. They model spatially-varying reflectance as a linear combination of a small number of Ward BRDFs, while Hertzmann and Seitz (2005) use a similar representation of reflectance using images of simple objects made of the same material as the modeled scene. In addition, Yu et al. (2004, 2007) propose a model-based method for recovering the 3D shape and the reflectance of a non-Lambertian object. Nevertheless, in this last paper, the authors constrain the object to be made of a *single textureless 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 Yu et al. (2004, 2007) is a "multiview shape from shading" method, similarly as the one proposed by Jin et al. (2004, 2008) which focuses on the Lambertian case. To our knowledge, most 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 Georghiades (2003), Vogiatzis et al. (2005) and for the multiview photometric stereo work of Lu and Little (1995). Only a small number of similar works are able to recover scenes with varying albedo: Birkbeck et al. (2006) and Hernández Esteban et al. (2008). However, in these approaches, specular highlights are filtered out by using a simple thresholding. As a result, only diffuse components are used to estimate shape. Moreover, in Birkbeck et al. (2006), the authors simply compute the light visibility of a point using a surface normal and a light direction and Hernández Esteban et al. (2008) also used a thresholding to detect shadowed pixels that are not visible from light sources, which is however not working in the presence of multiple light sources. Finally, let us emphasize that the method of Hernández Esteban et al. (2008) is specifically a photometric stereo method, i.e. it

requires several different lighting configurations and cannot perform classical stereo-vision.

In our work, we do not want to restrain ourselves to a single textureless material, i.e. the reflectance properties of the object can be spatially variable. Actually, nowadays, more and more objects are 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 in Jin et al. (2004, 2008), and we have to use a separate process which computes them. In this work, we assume that lighting conditions are known in advance. In practice, we can use spherical objects with the reference white color to capture the direction and color of light sources (Powell et al. 2001; Zhou and Kambhamettu 2002).

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 (which is necessary e.g. for realistic re-lighting), this allows to naturally introduce specular models in the mathematical formulation of the multiview reconstruction problem; and thus this allows the method to be robust to highlights. Without any additional effort, this allows also 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 enables to easily incorporate constraints on the reflectance and in particular to exploit shading effects in textureless regions (even if the number of different lighting conditions does not allow to do photometric stereo; e.g. if all images are taken under the same illumination).

Let us emphasize that, contrary to previous works that consider specific scenarios, our method can be applied indiscriminately to a number of classical scenarios—classical stereovision, multiview photometric stereo, and multiview shape from shading. Finally, based on the work of Gargallo et al. (2007), our method allows to naturally and simply combine in a single framework the three main cues available for shape reconstruction: silhouettes/apparent contours, stereo, and shading. To our knowledge, it is the first method which fuses these three cues in such a natural and convenient way. We do not claim that the method presented here gives better results (for 3D shape and reflectance) than previous approaches that are usually specific for certain scenarios: its intended merit is currently rather its generality.

The paper is organized as follows. In Sect. 2, we describe the modeling assumptions and we specify the notations used. In Sect. 3, we formulate the problem in the Bayesian framework; we then describe the associated cost functions in detail in Sect. 4. In Sect. 5, we precisely explain how we are minimizing the global energy. Experimental results on synthetic and real images data sets are shown in Sect. 6. Some discussion and future work are given in Sects. 7 and 8 concludes this paper.

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 $\mathbb{R}^3$. These surfaces are represented by S. More details are given below.

2.1 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 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 (i.e. the area covered by 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 ith camera ($c = 1$ for a gray-scale image and $c = 3$ for a color image). We denote by I the set of input images: $I = \{I_1, I_2, \ldots, I_{n_c}\}$; I_{iF} and I_{iB} are the restrictions of the function I_i to π_{iF} and π_{iB}, respectively.

We consider the visibility function for image i, that is induced by the foreground surfaces S, $v_S^i : \mathbb{R}^3 \to \mathbb{R}$. It is defined as $v_S^i(\mathbf{X}) = 1$ if $\mathbf{X}$ is visible from the ith camera and $v_S^i(\mathbf{X}) = 0$ if it is occluded by the foreground. 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, i.e. for all points $\mathbf{x} \in \pi_{iF}$, $\Pi_{i,S}^{-1}(\mathbf{x})$ is the point on S along the ray joining $\mathbf{X}$ to the optical center of the ith camera, that is closest to the optical center.

2.2 Lighting Conditions

We model the illumination by a finite number of distant point light sources, together with an ambient illumination radiating constant energy isotropically in all directions. Illumination conditions may be different for every image. n_{il} is the number of illuminants corresponding to the ith image and $\mathbf{l}_{ij} \in \mathbb{S}^2$ and $L_{ij} \in \mathbb{R}^c$ are the direction and intensity[1] of the jth illuminant associated with the ith image, respectively. Similarly, $L_{ia} \in \mathbb{R}^c$ is the intensity[1] of the ambient illumination for the ith image.

To model occlusions of light sources from 3D points, we use light visibility functions $v_S^{i,j} : \mathbb{R}^3 \to \mathbb{R}$. We define $v_{L_{ij}}(\mathbf{X}) = 1$ if the jth illuminant of the ith image is visible from $\mathbf{X}$, $v_{L_{ij}}(\mathbf{X}) = 0$ otherwise. In addition, $S_{L_{ij}}$ is the part of S that is visible from the jth illuminant of the ith image.

---

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

In other words, $v_{L_{ij}}(\mathbf{X}) = 1$ if $\mathbf{X} \in S_{L_{ij}}$, $v_{L_{ij}}(\mathbf{X}) = 0$ otherwise. Based on the light visibility, we can take into account self-shadowing.

2.3 Modeling the Foreground Surface

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 use a parametric reflectance model. Such a model should be selected in consideration of the applications aimed at. In this work, we consider the popular Blinn-Phong shading model (Blinn 1997). However, the proposed method is not limited to this model. By following the approach we describe, it is straightforward to use any other dichromatic reflection model.

We assume 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. Thus, the images I_i can be decomposed as

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

where I_{id}, I_{is}, and I_{ia} are images containing the diffuse, specular, and ambient reflection components of I_i, respectively.

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

$$I_{id}(\mathbf{x}) = \sum_{j=1}^{n_{il}} v_{L_{ij}}(\mathbf{X})\big(\rho_d(\mathbf{X})L_{ij}(\mathbf{n}(\mathbf{X}) \cdot \mathbf{l}_{ij})\big), \tag{2}$$

where $\rho_d(\mathbf{X}) \in \mathbb{R}^c$ is the diffuse albedo at point $\mathbf{X}$ and $\mathbf{n}(\mathbf{X})$ is the normal vector to the surface S at $\mathbf{X}$.

Specular reflectance is caused by the surface reflection, as with a mirror. This component is expressed as

$$I_{is}(\mathbf{x}) = \sum_{j=1}^{n_{il}} v_{L_{ij}}(\mathbf{X})\big(\rho_s(\mathbf{X})L_{ij}(\mathbf{n}(\mathbf{X}) \cdot \mathbf{h}_{ij}(\mathbf{X}))^{\alpha_s(\mathbf{X})}\big), \tag{3}$$

where $\mathbf{h}_{ij}(\mathbf{X})$ is the bisector of the angle spanned by $\mathbf{X}$, the optical center of the ith camera and the jth illuminant. $\rho_s(\mathbf{X}) \in \mathbb{R}^c$ 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 in the scene and modeled as

$$I_{ia}(\mathbf{x}) = \rho_d(\mathbf{X})L_{ia}, \tag{4}$$

where $\rho_d(\mathbf{X})$ is the diffuse albedo at $\mathbf{X}$, defined above, and L_{ia} is the intensity of the ambient illumination, defined in Sect. 2.2.

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

$$I_i(\mathbf{x}) = \sum_{j=1}^{n_{il}} v_{L_{ij}}(\mathbf{X})\mathbb{L}_{ij}(\mathbf{X}, \mathbf{n}(\mathbf{X})) + \rho_d(\mathbf{X})L_{ia}, \tag{5}$$

where

$$\begin{aligned}
\mathbb{L}_{ij}(\mathbf{X}, \mathbf{n}(\mathbf{X})) &= \mathbb{L}_{ij}^d(\mathbf{X}, \mathbf{n}(\mathbf{X})) + \mathbb{L}_{ij}^s(\mathbf{X}, \mathbf{n}(\mathbf{X})) \\
&= L_{ij}\rho_d(\mathbf{X})\big(\mathbf{n}(\mathbf{X}) \cdot \mathbf{l}_{ij}\big) \\
&\quad + L_{ij}\rho_s(\mathbf{X})\big(\mathbf{n}(\mathbf{X}) \cdot \mathbf{h}_{ij}(\mathbf{X})\big)^{\alpha_s(\mathbf{X})}.
\end{aligned} \tag{6}$$

In the sequel, in order to simplify the notations, we denote $R = (R_d, R_s)$, where $R_d = \rho_d$ and $R_s = (\rho_s, \alpha_s)$.

2.4 Modeling the Background

For most works on multiview stereo, the cost functionals used attain their global optimum when the modeled surface shrinks to an empty set, inducing a *minimal surface bias*. This bias may be avoided as suggested by Yezzi and Soatto (2003) and Gargallo et al. (2007), by modeling the background of the scene, in addition to only the foreground objects. The choice of a background model is dictated by scenarios and applications. For example, in Yezzi and Soatto (2003), Jin et al. (2004), the background is characterized by its radiance which is constrained to be constant or strongly regular. On the other hand, when the background is quite irregular, one can assume to have background images, i.e. images captured by the same cameras, but without the foreground objects in the scene. In this work, we assume that, in addition to the input images I, we have these background images $\tilde{I} = \{\tilde{I}_1, \ldots, \tilde{I}_{n_c}\}$.[2] We also define $\tilde{I}_{iF}$ and $\tilde{I}_{iB}$, analogously to I_{iF} and I_{iB}.

3 Bayesian Formulation of the Problem

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

$$\begin{aligned}
P(\Omega|I) = \frac{P(I|\Omega)P(\Omega)}{P(I)} &\propto P(I|\Omega)P(\Omega) \\
&= P(I|S, R)P(S, R) \\
&= P(I|S, R)P(S)P(R)
\end{aligned} \tag{7}$$

under the assumption that S and R are independent. Here,

---

[2] Another possibility is when the input are silhouette images, i.e. images where the foreground has already been segmented and the rest of the images been "painted" uniformly in some discriminative color. In this case, the background images are defined simply as completely uniform images of that color.

$P(I|\Omega) = P(I|S, R)$ is a likelihood and $P(S)$ and $P(R)$ are priors on shape and reflectance respectively.

3.1 Likelihood

When the camera calibration (i.e. the Π_i) and illumination conditions are given, we can produce synthetic images $\bar{I}_i(\Omega)$ corresponding to the input images I_i, by rendering the current estimate of Ω. The correct estimate of Ω will produce the same images, modulo noise and unmodeled effects of course. This allows us to measure the quality of the current estimate by comparing input images with rendered ones as in Schultz (1994), Yu et al. (2004). When assuming an independent identical distribution (i.i.d.) of measurement errors, the likelihood can be expressed as

$$P(I|\Omega) \propto \prod_{i=1}^{n_c} \exp\left(-\xi_i(\Omega)\right)$$
$$= \prod_{i=1}^{n_c} \exp\left(-\xi(I_i, \bar{I}_i(\Omega))\right), \tag{8}$$

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 on Surface Shape S

A usual and plausible prior for the surface shape S concerns its area.[3] The prior can be expressed as

$$P(S) \propto \exp\left(-\psi(S)\right). \tag{9}$$

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

3.3 Prior on Reflectance R

R is composed of two components, $R = (R_d, R_s)$. Here, unfortunately, reliably estimating specular reflectance for all surface points with only a uniform prior on its parameters, is very difficult unless there are observations of specular reflections for every surface point. For that reason, we need some specific prior on specular reflectance to be able to infer it in spite of the lack of such rich observations.[4]

It is physically valid to assume that specular reflectance varies smoothly within each homogeneous surface patch, i.e. that is made of the same material. This assumption is clearly reasonable in real life applications and in common scenes.

---

[3]In this case, a minimal surface that may be characterized as the surface of minimal area under given boundary conditions will be sought.

[4]We will discuss some special cases that do not need any specific prior on the surface reflectance in Sect. 5.3.3.

It is, however, also very difficult to partition Ω according to the types of materials. In this work, we use the diffuse reflectance of the surface as a soft constraint to partition Ω and define the prior on the surface reflectance as

$$P(R) \propto \exp\left(-\omega(R)\right), \tag{10}$$

where $\omega(R)$ is a function of the intrinsic gradient of the diffuse and specular reflectance of a surface. This function is defined below, in Sect. 4.3.

4 Description of the Cost Functions

Based on the derivations in Sect. 3, the problem is formulated as

$$P(\Omega|I) \propto P(I|\Omega)P(\Omega) = P(I|S, R)P(S, R)$$
$$= P(I|S, R)P(S)P(R)$$
$$\propto \prod_{i=1}^{n_c} \exp\left(-\xi_i(\Omega)\right) \times \left(\exp\left(-\psi(S)\right)\right)$$
$$\times \left(\exp\left(-\omega(R)\right)\right), \tag{11}$$

and it can be expressed in terms of cost functions as

$$E_{total}(\Omega) = E_{data}(\Omega) + E_{shape}(S) + E_{refl}(R)$$
$$= \sum_{i=1}^{n_c} \xi_i(\Omega) + \psi(S) + \omega(R). \tag{12}$$

Maximizing the probability (11) is equivalent to minimizing the total cost (12).

4.1 Data Cost Function

The current estimate of Ω gives a segmentation of each input image I_i into foreground I_{iF} and background I_{iB} and we can synthesize $\bar{I}_{iF}$ according to the image formation model of Sect. 2. As for $\bar{I}_{iB}$, it is generated according to the available background model. In this paper, as mentioned in Sect. 2.4, we use actual background images, i.e. $\bar{I}_{iB} = \tilde{I}_{iB}$. Also, as suggested by Yezzi and Soatto (2003), the similarity measure between observed and rendered images, $\xi_i(\Omega) = \xi(I_i, \bar{I}_i)$, is then rewritten as

$$\xi(I_i, \bar{I}_i) = \xi_F(I_{iF}, \bar{I}_{iF}) + \xi_B(I_{iB}, \bar{I}_{iB})$$
$$= \xi_F(I_{iF}, \bar{I}_{iF}) + \xi_B(I_{iB}, \tilde{I}_{iB})$$
$$= \xi_F(I_{iF}, \bar{I}_{iF}) - \xi_F(I_{iF}, \tilde{I}_{iF})$$
$$\quad + \xi_F(I_{iF}, \tilde{I}_{iF}) + \xi(I_{iB}, \tilde{I}_i)$$
$$= \hat{\xi}_F(I_{iF}, \bar{I}_{iF}) + \xi(I_i, \tilde{I}_i), \tag{13}$$

where $\hat{\xi}_F(I_{iF}, \bar{I}_{iF}) = \xi_F(I_{iF}, \bar{I}_{iF}) - \xi_F(I_{iF}, \tilde{I}_{iF})$.[5] Since $\xi(I_i, \tilde{I}_i)$ is independent of Ω, the data cost function is written as

$$E_{data}(\Omega) = \sum_{i=1}^{n_c} \hat{\xi}_F(I_{iF}, \bar{I}_{iF}) + C, \qquad (14)$$

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

4.1.1 Similarity Measure

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

$$\xi(I_i, \bar{I}_i) = \int_{\pi_i} e_i(\mathbf{x}) d\sigma_i, \qquad (15)$$

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

$$E_{data}(\Omega) = \sum_{i=1}^{n_c} \int_{\pi_{iF}} \hat{e}_i(\mathbf{x}) d\sigma_i + C, \qquad (16)$$

where $\hat{e}_i(\mathbf{x}) = e_i(I_i(\mathbf{x}), \bar{I}_i(\mathbf{x})) - e_i(I_i(\mathbf{x}), \tilde{I}_i(\mathbf{x}))$. We adopt the derivations proposed in Pons et al. (2005) for ξ_i, e_i, and $\partial_2 e_i$.

4.1.2 Decoupling Appearance from Surface Normal

As shown in (5), surface appearance (i.e., the data cost function) is dependent on both the surface normal and the position, and this makes the problem hard to solve and unstable. To resolve this problem, we introduce an auxiliary photometric unit vector field $\mathbf{v}$ satisfying $\|\mathbf{v}\| = 1$ as in Jin et al. (2004), which is used for the computation of surface appearance. The vector field $\mathbf{v}$ is estimated in alternation with all other parameters, i.e. shape and reflectance. Equation (6) is written in terms of $\mathbf{v}$ as

$$\mathbb{L}_{ij}(\mathbf{X}, \mathbf{v}(\mathbf{X})) = L_{ij}\rho_d(\mathbf{X})\big(\mathbf{v}(\mathbf{X}) \cdot \mathbf{l}_{ij}(\mathbf{X})\big)$$
$$+ L_{ij}\rho_s(\mathbf{X})\big(\mathbf{v}(\mathbf{X}) \cdot \mathbf{h}_{ij}(\mathbf{X})\big)^{\alpha_s(\mathbf{X})}, \qquad (17)$$

which is independent of $\mathbf{n}(\mathbf{X})$. To penalize the deviation between the actual normal vector $\mathbf{n}$ and the photometric normal vector $\mathbf{v}$, we add a new term

$$E_{dev}(\Omega) = \tau \int_S \chi(\mathbf{X}) d\sigma = \frac{\tau}{2} \int_S \|\mathbf{n}(\mathbf{X}) - \mathbf{v}(\mathbf{X})\|^2 d\sigma$$

$$= \tau \int_S (1 - (\mathbf{n}(\mathbf{X}) \cdot \mathbf{v}(\mathbf{X}))) d\sigma, \qquad (18)$$

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

$$E_{shape}(S) = \psi(S) = \lambda \int_S d\sigma, \qquad (19)$$

where λ is a control constant.

4.3 Reflectance Discontinuity Cost Function

Based on the assumption on surface reflectance in Sect. 3.3, 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

$$E_{refl}(R) = \omega(R) = \beta \int_S f(\mathbf{X}) d\sigma, \qquad (20)$$

where β is a control constant. $f(\mathbf{X})$ is defined as

$$f(\mathbf{X}) = \zeta\big(R_d(\mathbf{X})\big) \times \eta\big(R_s(\mathbf{X})\big), \qquad (21)$$

where $\zeta(R_d(\mathbf{X}))$ and $\eta(R_s(\mathbf{X}))$ are defined in terms of the magnitude of the intrinsic gradients of diffuse reflectance and specular reflectance respectively as

$$\zeta\big(R_d(\mathbf{X})\big) = \left(1 - \frac{\|\nabla_S R_d(\mathbf{X})\|^2}{M}\right), \qquad (22)$$

$$\eta\big(R_s(\mathbf{X})\big) = \big(\|\nabla_S \rho_s(\mathbf{X})\|^2 + \gamma \|\nabla_S \alpha_s(\mathbf{X})\|^2\big) \qquad (23)$$

where M is a pre-defined constant and ∇_S denotes the intrinsic gradient defined on S. Here, $\zeta(R_d(\mathbf{X}))$ is inversely proportional to the magnitude of the intrinsic gradient of R_d.[6] In addition, $\eta(R_s(\mathbf{X}))$ is proportional to the magnitude of the intrinsic gradient of ρ_s and α_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 with similar diffuse reflectance.

---

[5]Note that (13) is valid only when $\xi(I_i, \bar{I}_i)$ can be expressed as in (15).

[6]One has to use $M \geq 3$ for gray-level images and $M \geq 9$ for color images to make ζ positive.

4.4 Total Cost Function

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

$$E_{total}(\Omega) = E_{data}(\Omega) + E_{dev}(\Omega) + E_{shape}(S) + E_{refl}(R)$$

$$= C + \sum_{i=1}^{n_c} \int_{\pi_i F} \hat{e}_i(\mathbf{x}) d\sigma_i + \tau \int_S \chi(\mathbf{X}) d\sigma$$

$$+ \lambda \int_S d\sigma \; + \; \beta \int_S f(\mathbf{X}) d\sigma. \tag{24}$$

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

$$d\sigma_i = -\frac{\mathbf{d}_i(\mathbf{X}) \cdot \mathbf{n}(\mathbf{X})}{z_i(\mathbf{X})^3} d\sigma, \tag{25}$$

where $\mathbf{d}_i(\mathbf{X})$ is the vector connecting the center of the ith camera and $\mathbf{X}$ and $z_i(\mathbf{X})$ is the depth of $\mathbf{X}$ relative to the ith camera, we can replace the integral over the image plane by an integral over the surface as in Pons et al. (2007):

$$E_{data}(\Omega) = C - \sum_{i=1}^{n_c} \int_{S_i} \left(\hat{e}_i(\Pi_i(\mathbf{X})) \frac{\mathbf{d}_i(\mathbf{X}) \cdot \mathbf{n}(\mathbf{X})}{z_i(\mathbf{X})^3} \right) d\sigma$$

$$= C - \int_S \left(\sum_{i=1}^{n_c} v_S^i(\mathbf{X}) \hat{e}_i(\Pi_i(\mathbf{X})) \right.$$

$$\left. \times \frac{\mathbf{d}_i(\mathbf{X}) \cdot \mathbf{n}(\mathbf{X})}{z_i(\mathbf{X})^3} \right) d\sigma. \tag{26}$$

As a result, the total cost function (24) is expressed as

$$E_{total}(\Omega) = C + \int_S \left(-\sum_{i=1}^{n_c} \left(v_S^i \hat{e}_i \frac{\mathbf{d}_i \cdot \mathbf{n}}{z_i{}^3} \right) \right.$$

$$\left. + \tau \chi + \lambda + \beta f \right) d\sigma. \tag{27}$$

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

$$g(\mathbf{X}, \mathbf{n}(\mathbf{X})) = \left(-\sum_{i=1}^{n_c} \left(v_S^i \hat{e}_i \frac{\mathbf{d}_i \cdot \mathbf{n}}{z_i{}^3} \right) \right.$$

$$\left. + \tau \chi + \lambda + \beta f \right), \tag{28}$$

Equation (24) is simply rewritten as

$$E_{total}(\Omega) = C + \int_S g(\mathbf{X}, \mathbf{n}(\mathbf{X})) d\sigma. \tag{29}$$
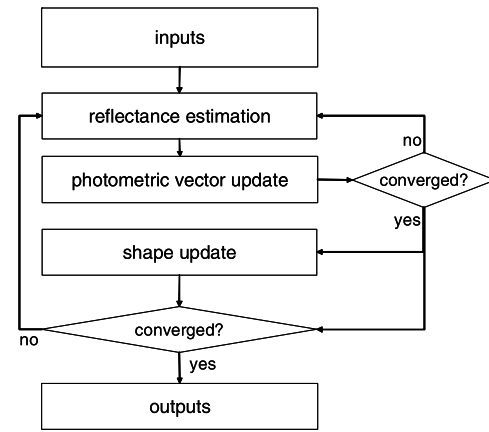

Fig. 1 Overall procedure of the proposed method. It is composed of three parts: shape update, reflectance estimation, and update of the auxiliary vector field $\mathbf{v}$

Here, although the total cost function is an integral over the surface, it does not suffer from the usual minimal surface bias mentioned in Sect. 2.4: most functionals used in multiview stereo have an empty set as globally optimal surface, since they do not "explain" all pixels in the input images. Our approach, like Yezzi and Soatto (2003), takes into account all pixels in the cost function, not only those covered by the current estimate of the foreground object's shape, using both the estimated foreground and the available background information.

5 Scene Recovery

Recently, based on graph cuts or convexity, several global optimization methods have been proposed for the classical multiview stereovision problem, see Snow et al. (2000), Paris et al. (2006), Vogiatzis et al. (2007), Kolev et al. (2007b, 2007a). 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 in the previous section. In this work, scene recovery is achieved by minimizing E_{total} via gradient descents.

In other respects, S and R are highly coupled and 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. The overall procedure is shown in Fig. 1.

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 correspond-
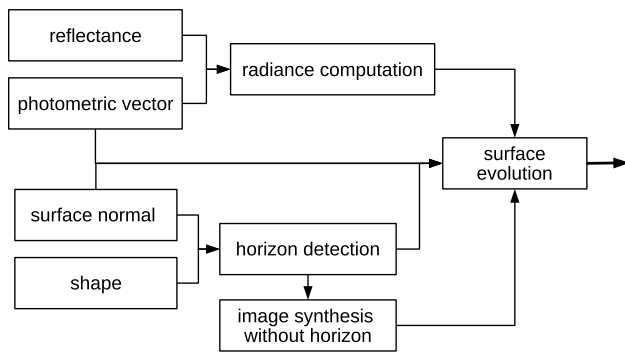
Fig. 2 Shape update

ing to the cost functions respectively. The final gradient descent flow is then given by

$$S_t = \left(S_t|_{data} + S_t|_{dev} + S_t|_{shape} + S_t g|_{refl}\right); \tag{30}$$

where $S_t|_{data}$, $S_t|_{dev}$, $S_t|_{shape}$ and $S_t|_{refl}$ are described below. The shape update scheme is shown in Fig. 2. For more details about gradient descent flows, we refer the inexperienced reader to Solem and Overgaard (2005) who nicely detail the geometric formulation of gradient descent in such a context.

5.1.1 Gradient Descent Flow for the Data Cost

As shown in (26), the data cost is a function of the visibility of a surface point, which is dependent on the whole surface shape, not only on the normal at the considered point. To correctly handle visibility for non-convex objects, self-occlusions of the foreground object must be taken into account, i.e. parts of the surface that occlude other parts of it from cameras. This is done according to Yezzi and Soatto (2003) and Gargallo et al. (2007) (which extends the work of Solem and Overgaard 2005), by writing $S_t|_{data}$ as

$$S_t|_{data} = \sum_{i=1}^{n_c} \left(- \frac{v_S^i(\hat{e}_i - \hat{e}'_i)}{z_i^3} \left(\mathbf{d}_i^t \nabla \mathbf{n} \mathbf{d}_i^t \delta(\mathbf{d}_i \cdot \mathbf{n})\right) \right.$$
$$\left. + \frac{v_S^i}{z_i^3} \left((\partial_2 \hat{e}_i \nabla \bar{I}_i) \cdot \mathbf{d}_i\right) \right). \tag{31}$$

Here $\mathbf{d}_i$ is, as defined in Sect. 4.4, the vector connecting the center of the ith camera and a surface point $\mathbf{X}$, and $\delta(\cdot)$ is the delta function. Hence, the term $\delta(\mathbf{d}_i \cdot \mathbf{n})$ in the above expression is non-zero exactly at horizon points, i.e. points on (self-)occluding contours of the foreground. Let $\mathbf{X}'$ be the terminator of a horizon point $\mathbf{X}$ as shown in Fig. 3 (for more details see Gargallo et al. 2007; Gargallo 2008). $\hat{e}'_i$ is a similarity measure computed using the radiance of $\mathbf{X}'$ in the direction of the ith camera and the intensity at the image position corresponding to $\mathbf{X}$ and $\mathbf{X}'$ (cf. Sect. 4.1.1). Hence, the
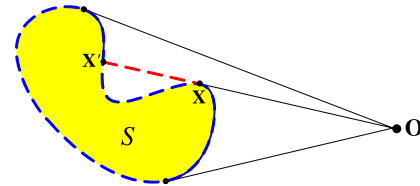

Fig. 3 Horizon point $\mathbf{X}$ and its terminator point $\mathbf{X}'$. $\mathbf{X}'$ is used to consider visibility changes. See Gargallo et al. (2007) for details

first term of the sum in (31) accounts for what happens to the likelihood when the visibility of non-convex objects changes due to the surface evolution. When a horizon point has no terminator point on the foreground surface itself, $\hat{e}'_i = 0$ because the terminator point is from the background. $\nabla \bar{I}_i$ is expressed by using (5) as

$$\nabla \bar{I}_i = \sum_{j=1}^{n_{il}} \{(\nabla v_{\mathbb{L}_{ij}})\mathbb{L}_{ij} + v_{L_{ij}}(\nabla \mathbb{L}_{ij})\} + (\nabla \rho_a)L_{ia}, \tag{32}$$

where

$$\nabla \mathbb{L}_{ij} = \nabla \mathbb{L}_{ij}^d + \nabla \mathbb{L}_{ij}^s, \tag{33}$$

and

$$\nabla \mathbb{L}_{ij}^d = L_{ij}(\nabla \rho_d)(\mathbf{v} \cdot \mathbf{l}_{ij}) + L_{ij}\rho_d \left(\nabla(\mathbf{v} \cdot \mathbf{l}_{ij})\right), \tag{34}$$

$$\nabla \mathbb{L}_{ij}^s = L_{ij}(\nabla \rho_s)(\mathbf{v} \cdot \mathbf{h}_{ij})^{\alpha_s} + L_{ij}\rho_s \left(\nabla(\mathbf{v} \cdot \mathbf{h}_{ij})^{\alpha_s}\right). \tag{35}$$

This gradient descent flow includes both the variation related to the camera visibility changes (the first term in (31)) and the variation related to the image changes (the second term in (31)), 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

Similarly as in Jin et al. (2004, 2008), the gradient descent flow for the normal deviation cost $S_t|_{dev}$ (originating from $E_{dev}(\Omega)$) is

$$S_t|_{dev} = (-2\tau H + \tau(\nabla \cdot \mathbf{v})), \tag{36}$$

where H is the mean curvature. Also $S_t|_{shape}$ (from $E_{shape}(S)$) is the mean curvature flow as

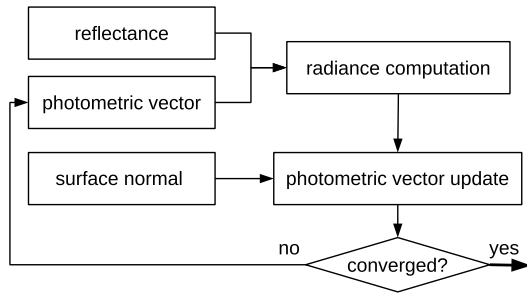$$S_t|_{shape} = -2\lambda H. \tag{37}$$

Fig. 4 Photometric unit vector update

5.1.3 Gradient Descent Flow for the Reflectance Discontinuity Cost

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 Jin et al. (2003), we get the following equation for surface evolution.

$$S_t|_{refl} = -2\beta \left(\frac{1}{M} m(\rho_d)\eta(R_s) - (m(\rho_s) \right.$$
$$\left. + \gamma m(\alpha_s))\zeta(R_d) \right). \tag{38}$$

Here,

$$m(\rho_s) = \left(\mathrm{II}(\nabla_S \rho_s \times \mathbf{n}) + \|\nabla_S \rho_s\|^2 H \right), \tag{39}$$

$$m(\alpha_s) = \left(\mathrm{II}(\nabla_S \alpha_s \times \mathbf{n}) + \|\nabla_S \alpha_s\|^2 H \right), \tag{40}$$

$$m(\rho_d) = \left(\mathrm{II}(\nabla_S \rho_d \times \mathbf{n}) + \|\nabla_S \rho_d\|^2 H \right), \tag{41}$$

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,

$$\frac{\partial g}{\partial \mathbf{v}} = \left(-\sum_{i=1}^{n_c} v_S^i \partial_2 \hat{e}_i \frac{\partial \bar{I}_i}{\partial \mathbf{v}} \frac{\mathbf{d}_i \cdot \mathbf{n}}{z_i^3} \right) + (-\tau \mathbf{n}) = 0. \tag{42}$$

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

$$\frac{\partial \bar{I}_i}{\partial \mathbf{v}} = \sum_{j=1}^{n_{il}} v_{L_{ij}} L_{ij} \left(\rho_d \mathbf{l}_{ij} + \rho_s \alpha_s \left(\mathbf{v} \cdot \mathbf{h}_{ij} \right)^{\alpha_s - 1} \mathbf{h}_{ij} \right). \tag{43}$$

We can update $\mathbf{v}$ by performing gradient descent using the following PDE:

$$\frac{\partial \mathbf{v}}{\partial t} = \left(-\sum_{i=1}^{n_c} v_S^i \partial_2 \hat{e}_i \frac{\partial \bar{I}_i}{\partial \mathbf{v}} \frac{\mathbf{d}_i \cdot \mathbf{n}}{z_i^3} \right) + (-\tau \mathbf{n}). \tag{44}$$

However, because we have to keep $\|\mathbf{v}\| = 1$, we can not use (44) directly. 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.

$$\frac{\partial g}{\partial \theta_v} = \frac{\partial g}{\partial \mathbf{v}} \cdot \frac{\partial \mathbf{v}}{\partial \theta_v} = 0, \tag{45}$$

$$\frac{\partial g}{\partial \phi_v} = \frac{\partial g}{\partial \mathbf{v}} \cdot \frac{\partial \mathbf{v}}{\partial \phi_v} = 0. \tag{46}$$

Here, $\frac{\partial \mathbf{v}}{\partial \theta_v}$ and $\frac{\partial \mathbf{v}}{\partial \phi_v}$ are given as

$$\frac{\partial \mathbf{v}}{\partial \theta_v} = \begin{bmatrix} -\sin\theta_v \sin\phi_v \\ \cos\theta_v \sin\phi_v \\ 0 \end{bmatrix}, \qquad \frac{\partial \mathbf{v}}{\partial \phi_v} = \begin{bmatrix} \cos\theta_v \cos\phi_v \\ \sin\theta_v \cos\phi_v \\ -\sin\phi_v \end{bmatrix}. \tag{47}$$

So, we update $\mathbf{v}$ by updating θ_v and ϕ_v by performing gradient descent using the following two PDEs:

$$\frac{\partial \theta_v}{\partial t} = \left(\left(-\sum_{i=1}^{n_c} v_S^i \partial_2 \hat{e}_i \frac{\partial \bar{I}_i}{\partial \mathbf{v}} \frac{\mathbf{d}_i \cdot \mathbf{n}}{z_i^3} \right) + (-\tau \mathbf{n}) \right)$$
$$\cdot \begin{bmatrix} -\sin\theta_v \sin\phi_v \\ \cos\theta_v \sin\phi_v \\ 0 \end{bmatrix} \tag{48}$$

and

$$\frac{\partial \phi_v}{\partial t} = \left(\left(-\sum_{i=1}^{n_c} v_S^i \partial_2 \hat{e}_i \frac{\partial \bar{I}_i}{\partial \mathbf{v}} \frac{\mathbf{d}_i \cdot \mathbf{n}}{z_i^3} \right) + (-\tau \mathbf{n}) \right)$$
$$\cdot \begin{bmatrix} \cos\theta_v \cos\phi_v \\ \sin\theta_v \cos\phi_v \\ -\sin\phi_v \end{bmatrix}. \tag{49}$$

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 $(E_{data}(\Omega) + E_{refl}(R))$. Since it is complex to estimate diffuse and specular reflectance at the same time due to the high coupling between them, we alternatively estimate surface reflectance components one by one while assuming that

Fig. 5 Reflectance estimation
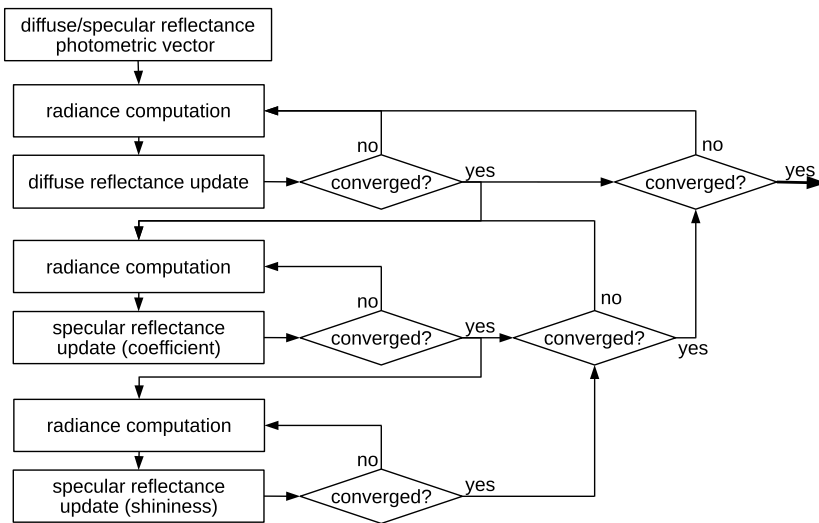
the rest are given and fixed. We repeat the procedure until they no longer change. Figure 5 shows the whole scheme we have used for the reflectance estimation. Below, we are detailing the intermediate steps.

5.3.1 Diffuse Reflectance Estimation

For given S and R_s, we estimate ρ_d that minimizes the cost

$$
E_{data} + E_{refl} = \int_S \left(\left(-\sum_{i=1}^{n_c} v_S^i \hat{e}_i \frac{\mathbf{d}_i \cdot \mathbf{n}}{z_i{}^3} \right) + \beta \left(1 - \frac{\|\nabla_S \rho_d\|^2}{M} \right) \eta(R_s) \right) d\sigma. \tag{50}
$$

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

$$
-\sum_{i=1}^{n_c} v_S^i \partial_2 \hat{e}_i \frac{\partial \bar{I}_i}{\partial \rho_d} \frac{\mathbf{d}_i \cdot \mathbf{n}}{z_i{}^3} + \frac{2\beta}{M} \eta(R_s) \Delta_S \rho_d = 0, \tag{51}
$$

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

$$
\frac{\partial \bar{I}_i}{\partial \rho_d} = \sum_{j=1}^{n_{il}} v_{L_{ij}} L_{ij} \left(\mathbf{v} \cdot \mathbf{l}_{ij} \right) + L_{ia}. \tag{52}
$$

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

$$
\frac{\partial \rho_d}{\partial t} = \left(-\sum_{i=1}^{n_c} v_S^i \partial_2 \hat{e}_i \frac{\partial \bar{I}_i}{\partial \rho_d} \frac{\mathbf{d}_i \cdot \mathbf{n}}{z_i{}^3} \right) + \left(\frac{2\beta}{M} \eta(R_s) \right) \Delta_S \rho_d. \tag{53}
$$

5.3.2 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

$$
\left(-\sum_{i=1}^{n_c} v_S^i \partial_2 \hat{e}_i \frac{\partial \bar{I}_i}{\partial \rho_s} \frac{\mathbf{d}_i \cdot \mathbf{n}}{z_i{}^3} \right) - 2\beta \left(\Delta_S \rho_s \right) \zeta(\rho_d) = 0, \tag{54}
$$

where $\frac{\partial \bar{I}_i}{\partial \rho_s}$ is given as

$$
\frac{\partial \bar{I}_i}{\partial \rho_s} = \sum_{j=1}^{n_{il}} v_{L_{ij}} L_{ij} \left(\mathbf{v} \cdot \mathbf{h}_{ij} \right)^{\alpha_s}. \tag{55}
$$

We again solve the PDE by performing gradient descent using the following PDE to get the solution of (54).

$$
\frac{\partial \rho_s}{\partial t} = -\sum_{i=1}^{n_c} \left(v_S^i \partial_2 \hat{e}_i \frac{\partial \bar{I}_i}{\partial \rho_s} \frac{\mathbf{d}_i \cdot \mathbf{n}}{z_i{}^3} \right) - 2\beta \left(\Delta_S \rho_s \right) \zeta(\rho_d). \tag{56}
$$

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

$$
\frac{\partial \alpha_s}{\partial t} = -\sum_{i=1}^{n_c} \left(v_S^i \partial_2 \hat{e}_i \frac{\partial \bar{I}_i}{\partial \alpha_s} \frac{\mathbf{d}_i \cdot \mathbf{n}}{z_i{}^3} \right) - 2\beta\gamma \left(\Delta_S \alpha_s \right) \zeta(\rho_d), \tag{57}
$$

where $\frac{\partial \bar{I}_i}{\partial \alpha_s}$ is given as

$$
\frac{\partial \bar{I}_i}{\partial \alpha_s} = \sum_{j=1}^{n_{il}} v_{L_{ij}} L_{ij} \rho_s \left(\mathbf{v} \cdot \mathbf{h}_{ij} \right)^{\alpha_s} \ln \left(\mathbf{v} \cdot \mathbf{h}_{ij} \right). \tag{58}
$$

5.3.3 Case of a Single-Material Surface

When dealing with a surface that has uniform specular reflectance R_s, it is possible to set $\rho_s(\mathbf{X}) = \rho_s$ and $\alpha_s(\mathbf{X}) = \alpha_s$

Fig. 6 Result for the "dino" image set (16 images)—Lambertian surface case (static illumination and varying viewpoint)

(a) input images   (b) synthesized images   (c) initial shape   (d) estimated shape
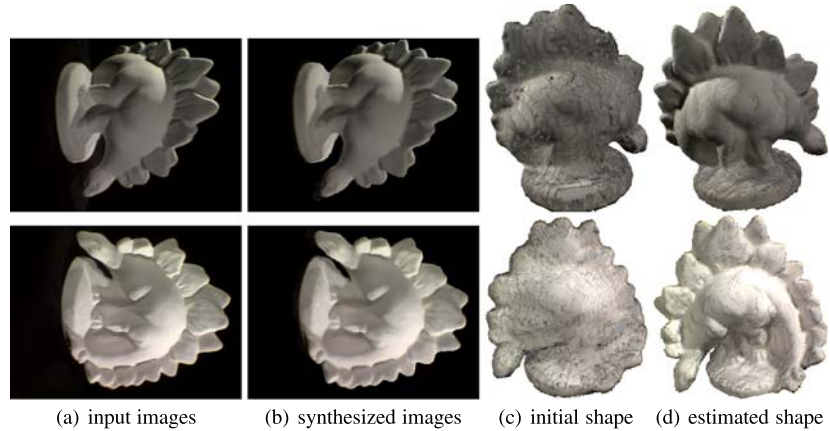
for all surface points. In this case, the discontinuity cost function of surface reflectance, $E_{refl}(R)$, can be excluded because $f(\mathbf{X})$ in (21) is zero everywhere on the surface. Hence, the gradient descent flow is then given by

$$S_t = \left(S_t|_{data} + S_t|_{dev} + S_t|_{shape} \right), \tag{59}$$

and the PDE used for the estimation of ρ_d, (53), is simplified to

$$\frac{\partial \rho_d}{\partial t} = -\sum_{i=1}^{n_c} v_S^i \partial_2 \hat{e}_i \frac{\partial \bar{I}_i}{\partial \rho_d} \frac{\mathbf{d}_i \cdot \mathbf{n}}{z_i{}^3}. \tag{60}$$

In addition, ρ_s and α_s are computed by performing gradient descent using the following PDEs.

$$\frac{\partial \rho_s}{\partial t} = \int_S \left(-\sum_{i=1}^{n_c} v_S^i \partial_2 \hat{e}_i \frac{\partial \bar{I}_i}{\partial \rho_s} \frac{\mathbf{d}_i \cdot \mathbf{n}}{z_i{}^3} \right) d\sigma, \tag{61}$$

$$\frac{\partial \alpha_s}{\partial t} = \int_S \left(-\sum_{i=1}^{n_c} v_S^i \partial_2 \hat{e}_i \frac{\partial \bar{I}_i}{\partial \alpha_s} \frac{\mathbf{d}_i \cdot \mathbf{n}}{z_i{}^3} \right) d\sigma. \tag{62}$$

6 Experiments

6.1 Implementation

We have implemented the gradient descent surface evolution in the level set framework in which the topological changes of surfaces are handled automatically (Osher and Sethian 1988; Sethian 1999; Osher and Fedkiw 2002). 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. The camera and light visibility were computed using OpenGL z-

buffering.[7] In all experiments, we detected saturated pixels by thresholding the intensity against $I_{th} = 253$, and ignored them in all further computations.

For synthetic data sets, the estimated shape is quantitatively evaluated in terms of accuracy and completeness as in Seitz et al. (2006). We used 95% for accuracy and the 1.0 mm error for completeness. For easy comprehension, the size of a target object is normalized so that it is smaller than [100 mm 100 mm 100 mm]. Here, beside the shape evaluation, we also evaluated the estimated reflectance in the same manner. For each point on an estimated surface, we found the nearest point on the true surface and compute the distance and reflectance differences, and vice versa. In addition, we computed the average difference between input images and synthesized images as

$$e_{image} = \frac{1}{n_c} \sum_{i=1}^{n_c} \frac{1}{A} \int_{\pi_i} \left\| \left(I_i(\mathbf{x}) - \bar{I}_i(\mathbf{x}) \right) \right\| d\sigma_i, \tag{63}$$

where $A = \int_{\pi_i} d\sigma_i$.

6.2 Experimental Results

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 other words, when images of Lambertian surfaces are taken under static illumination, the proposed method can be applied even without lighting information, assuming that there is only an ambient illumination. In this case, we do not need to take care of surface/photometric normals and only $E_{data}(\Omega)$ and

---

[7]Light visibility is computed by using virtual cameras located at the positions of light sources.

Fig. 7 Result for the "bimba" image set (18 images)—textureless Lambertian surface case (varying illumination and viewpoint). 95% accuracy (shape, $\rho_{dr}, \rho_{dg}, \rho_{db}$) = (2.16 mm, 0.093, 0.093, 0.093), 1.0 mm completeness (shape, $\rho_{dr}, \rho_{dg}, \rho_{db}$) = (82.63%, 0.104, 0.104, 0.104), $e_{image} = 1.44$
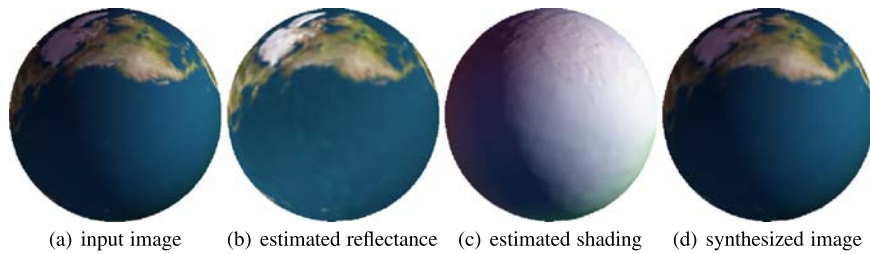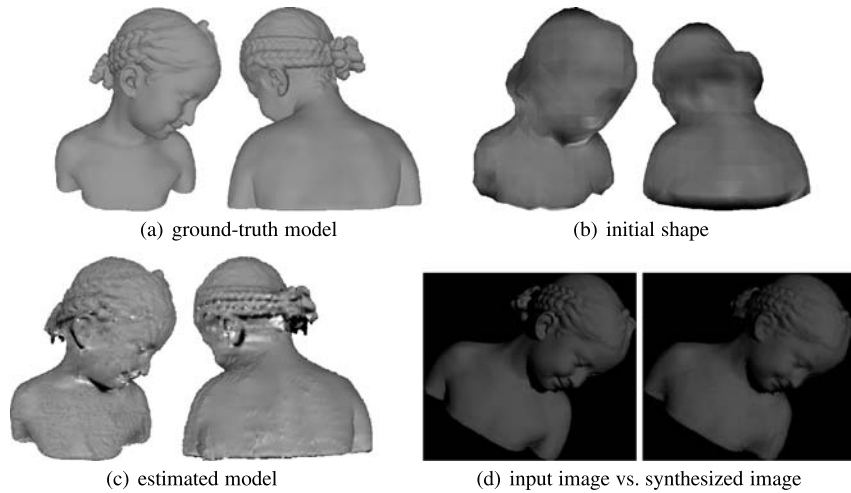

(a) ground-truth model

(b) initial shape

(c) estimated model

(d) input image vs. synthesized image

(a) input image (b) estimated reflectance (c) estimated shading (d) synthesized image

Fig. 8 Result for the "sphere" image set (32 images)—textured Lambertian surface case (static illumination and varying viewpoint). 95% accuracy (shape, $\rho_{dr}, \rho_{dg}, \rho_{db}$) = (1.06 mm, 0.025, 0.019, 0.017), 1.0 mm completeness (shape, $\rho_{dr}, \rho_{dg}, \rho_{db}$) = (99.74%, 0.023, 0.017, 0.016), $e_{image} = 0.60$

$E_{shape}(S)$ are computed. The intensity conservation assumption (ICA) is valid in all images and the proposed method works much like the conventional multiview stereo methods and estimates the shape and radiance of Lambertian surfaces. Figure 6 shows the result for the dino image set (Seitz et al. 2006), 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 Figs. 7 to 12. Figure 7 shows the ground-truth shape of the "bimba" image set (18 images) of a textureless object, and the estimation result. The surface has uniform diffuse reflectance and input images were taken under different illuminations. In this case, the proposed method works as a multiview photometric stereo method and recovers the shape and the diffuse reflectance of each surface point. Here, black points in the estimated model correspond to points that were not visible from any camera and/or any light source.

Figure 8 shows one of 32 textured input images and the synthesized image generated using the estimated shape (i.e., shading) and reflectance. Based on this result, we can also synthesize images of the scene for different lighting conditions, as shown in Fig. 9. Results for a more complex object
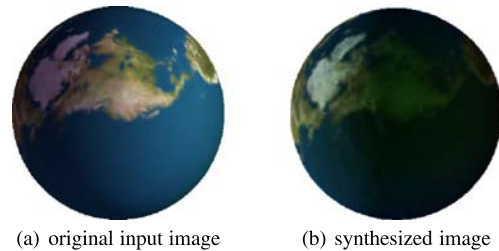

(a) original input image

(b) synthesized image

Fig. 9 Image synthesis—an image of the same scene, for different lighting conditions

are shown in Figs. 10 and 11. The images synthesized using the estimation closely resemble input images while the shading and the reflectance are successfully separated. Furthermore, it is possible to synthesize images under different lighting conditions, even from different viewpoints. The proposed method also recovers concave parts well as shown in Fig. 12.

We then applied our method to the images of textureless/textured non-Lambertian surfaces showing specular reflection. Note that, unlike previous methods (Birkbeck et al. 2006; Hernández Esteban et al. 2008), we do not use any thresholding to filter out specular highlight pixels. The result for the smoothed "bimba" data set is shown in Fig.

Fig. 10 Result for the "dragon" image set (32 images)—textured Lambertian surface case (static illumination and varying viewpoint). 95% accuracy (shape, $\rho_{dr}, \rho_{dg}, \rho_{db}$) = (1.28 mm, 0.090, 0.073, 0.066), 1.0 mm completeness (shape, $\rho_{dr}, \rho_{dg}, \rho_{db}$) = (97.11%, 0.064, 0.056, 0.052), $e_{image} = 1.25$
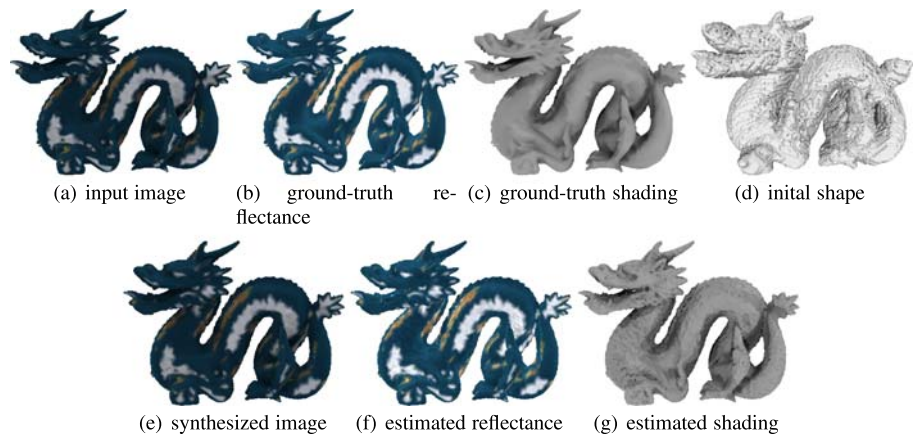

(a) input image (b) ground-truth reflectance (c) ground-truth shading (d) inital shape

(e) synthesized image (f) estimated reflectance (g) estimated shading

Fig. 11 Synthesized result for different lighting conditions and viewed from a viewpoint that is different from all input viewpoints. A comparison with the ground-truth is possible because this is synthetic data
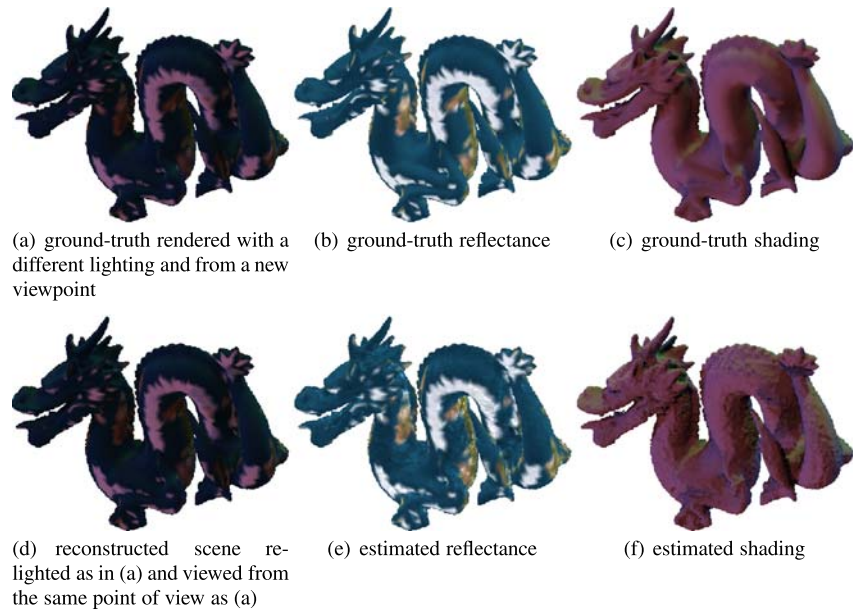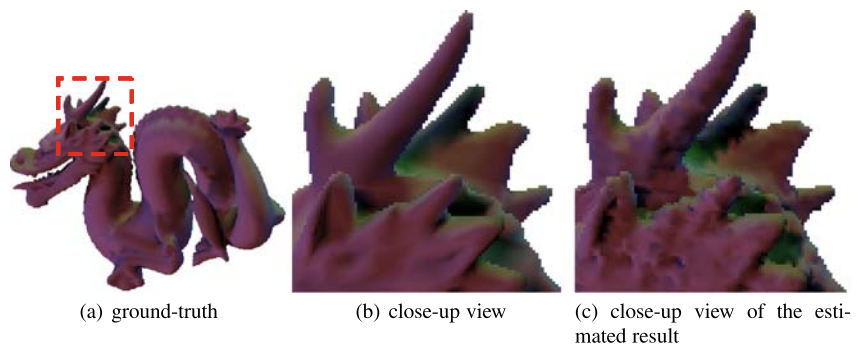

(a) ground-truth rendered with a different lighting and from a new viewpoint (b) ground-truth reflectance (c) ground-truth shading

(d) reconstructed scene re-lighted as in (a) and viewed from the same point of view as (a) (e) estimated reflectance (f) estimated shading

Fig. 12 Close-up view of the concave part of the "dragon" model

(a) ground-truth (b) close-up view (c) close-up view of the estimated result

13. In this case, the surface has uniform diffuse/specular reflectance and each image was taken under a different illumination. Here, we used the method described in Sect. 5.3.3 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 ($\rho_s = 0.7, \alpha_s = 50$) while the estimated one is ($\rho_s = 0.61, \alpha_s = 41.8$).[8]

---

[8]Note that small errors in estimated surface normals can cause large errors in specular reflectance because of its sensitivity to the surface normal. For instance, $0.7 \times (0.98)^{50}(= 0.255) \approx 0.61 \times (0.979)^{41.8}(= 0.251)$.
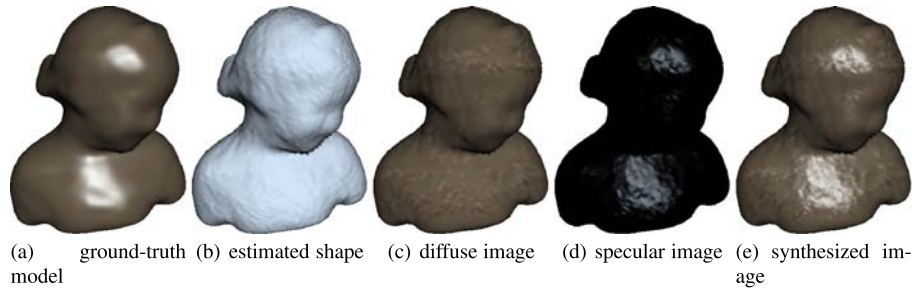
(a)   ground-truth (b) estimated shape (c) diffuse image   (d) specular image (e) synthesized im-
model   age

Fig. 13 Result for the smoothed "bimba" image set (36 images)—textureless non-Lambertian surface case (uniform specular reflectance, varying illumination and viewpoint). 95% accuracy (shape,

$\rho_{dr}, \rho_{dg}, \rho_{db}, \rho_s, \alpha_s) = (0.33 \text{ mm}, 0.047, 0.040, 0.032, 0.095, 8.248)$, 1.0 mm completeness (shape, $\rho_{dr}, \rho_{dg}, \rho_{db}, \rho_s, \alpha_s) = (100\%, 0.048, 0.041, 0.032, 0.095, 8.248)$, $e_{image} = 1.63$

Fig. 14 Result comparison using the "ellipse" image set (16 images)—textureless non-Lambertian surface case (uniform specular reflectance, static illumination and varying viewpoint)

(a) two input images

(b) results using Pons et al. (2007) (MI and CCL)   (c) our result
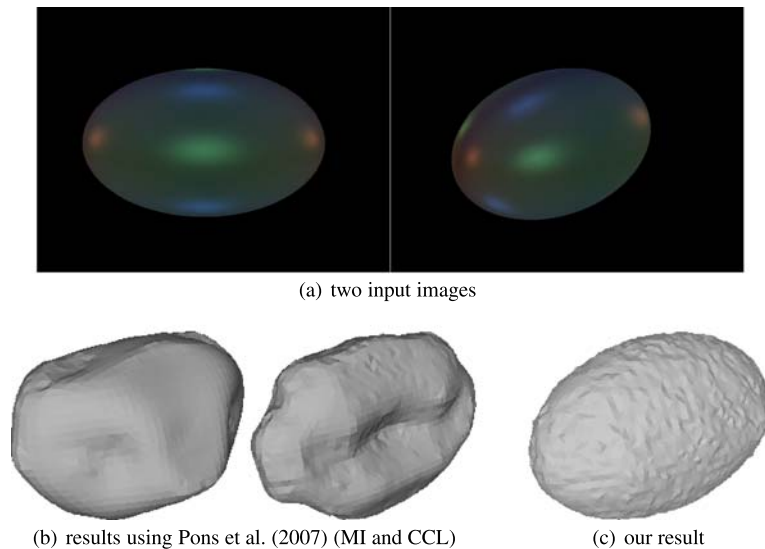
Note that most previous methods do not work for image sets taken under varying illumination and, moreover, they have difficulties to deal with specular reflection even if the images are taken under static illumination. For example, Figs. 14 and 15 show some results obtained by the method of Pons et al. (2007) 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 in Figs. 14 and 15, the method of Pons et al. (2007) 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 Birkbeck et al. (2006) and Hernández Esteban et al. (2008) (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 Birkbeck et al. (2006),

Hernández Esteban et al. (2008)—in this case, each image has a different illumination and observes specular reflections. The light position and color were measured using a white sphere placed in the scene. Figure 16 shows one image among 59 input images, the initial shape obtained using silhouettes, and the final result. Here, we simply assumed a single-material surface (i.e. uniform specular reflectance, but varying albedo). More results using real image sets are shown in Figs. 17 to 18. ($72 \times 72 \times 72$) grids were used for the "saddog" and "duck" image sets and ($64 \times 64 \times 64$) grids for "bunny". Although sparse grid volumes were used, the proposed method successfully estimated the shape of the glossy object even under specular reflection, while estimating the latter. Here, we can see that, although the estimated specular reflectance may not be highly accurate because of the inaccuracy of lighting calibration, saturation, and unmodeled photometric phenomena such as interreflections that often occur on glossy surfaces, it really helps to recover the shape well.

Finally, we applied our method to the most general case—images of textured non-Lambertian surfaces with

Fig. 15 Result comparison using the smoothed "bimba" image set (16 images)—textured non-Lambertian surface case (uniform specular reflectance, varying illumination and viewpoint)
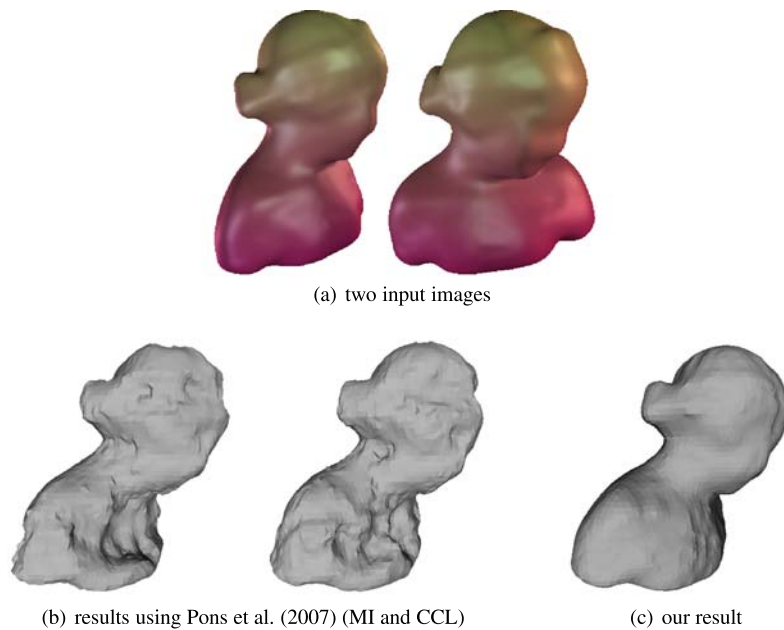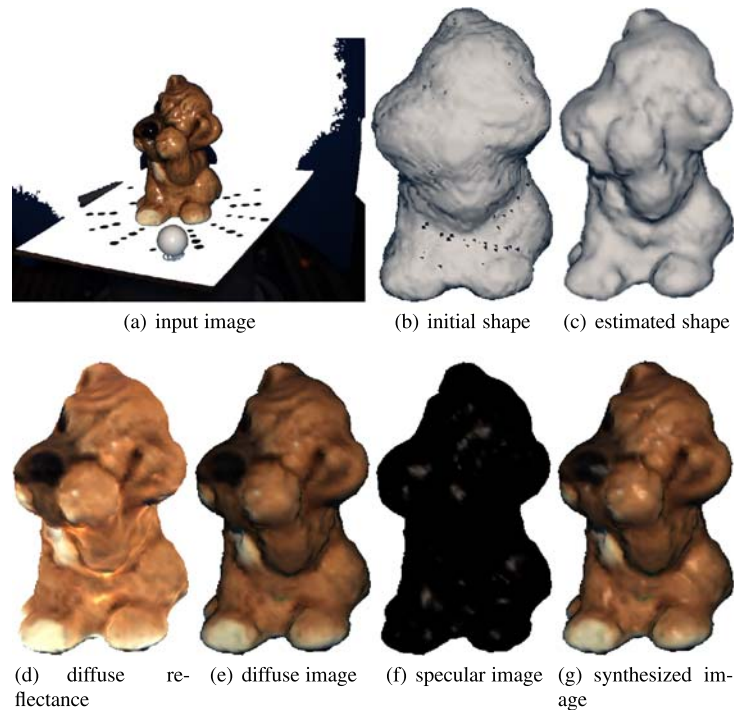
(a) two input images

(b) results using Pons et al. (2007) (MI and CCL)   (c) our result

Fig. 16 Result for the "saddog" image set (59 images)—textured non-Lambertian surface case (uniform specular reflectance, varying illumination and viewpoint). $e_{image} = 2.45$

(a) input image   (b) initial shape   (c) estimated shape

(d) diffuse re-   (e) diffuse image   (f) specular image   (g) synthesized im-
flectance   age

spatially varying diffuse and specular reflectance and shininess, cf. Fig. 19. Input images were generated under static illumination (with multiple light sources) while changing the viewpoint. Figure 19 shows one image among 36 input images, one ground-truth diffuse image, one ground-truth specular image, ground-truth shading, and our results. ($64 \times 125 \times 64$) 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 because of sparse specular reflection observation. 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 observation of specular reflections for every surface point.

Fig. 17 Result for the "bunny" image set (26 images)—textured non-Lambertian surface case (uniform specular reflectance, varying illumination and viewpoint). $e_{image} = 5.34$

(a) input image   (b) initial shape   (c) estimated shape

(d) diffuse reflectance   (e) diffuse image   (f) specular image   (g) synthesized image
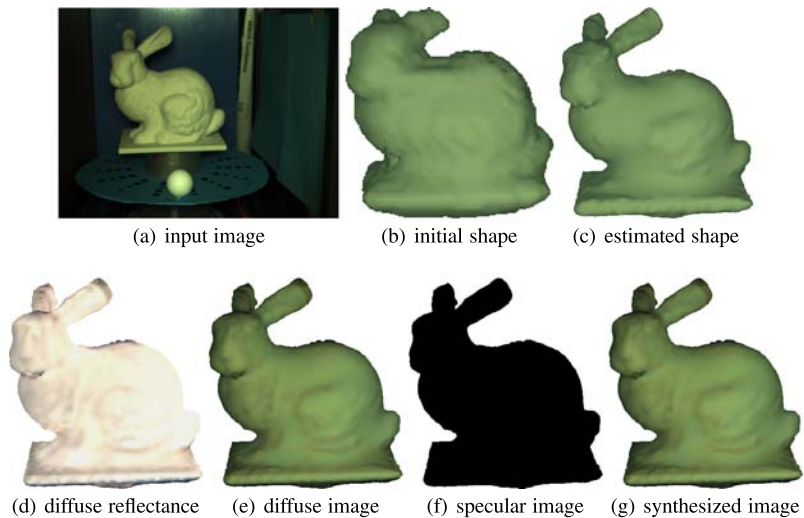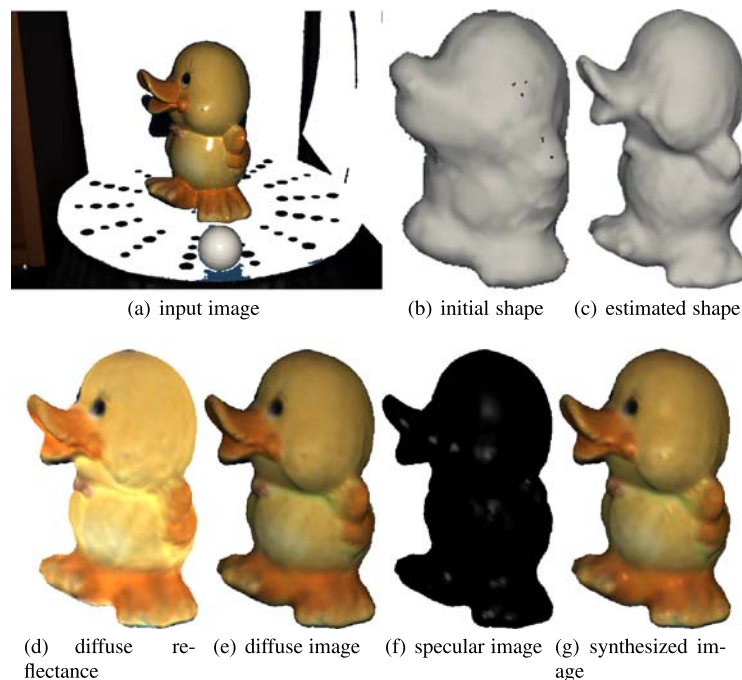
Fig. 18 Result for the "duck" image set (28 images)—textured non-Lambertian surface case (uniform specular reflectance, varying illumination and viewpoint). $e_{image} = 2.79$

(a) input image   (b) initial shape   (c) estimated shape

(d) diffuse reflectance   (e) diffuse image   (f) specular image   (g) synthesized image
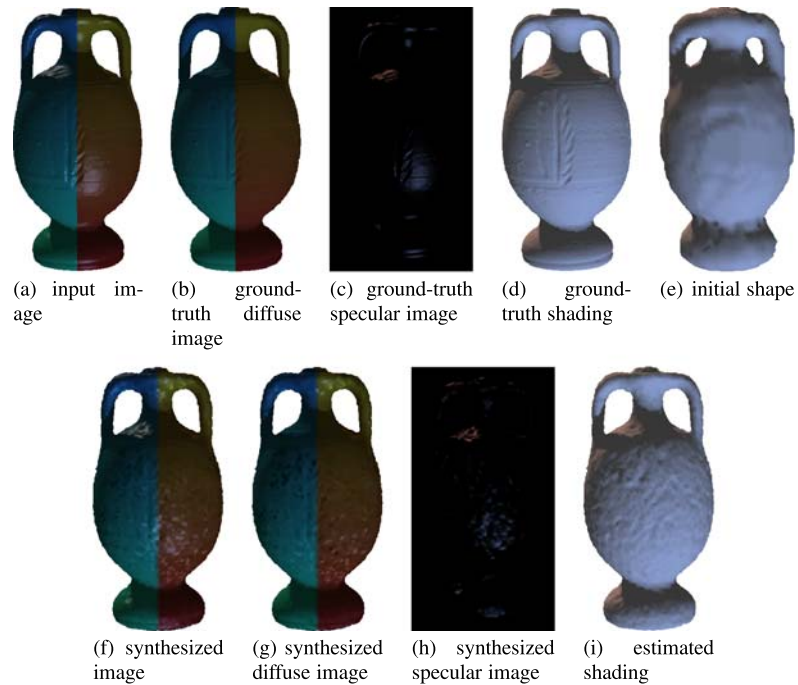
7 Discussion and Further Work

When considering non-Lambertian surfaces under varying illumination, the core question is how to deal with specular reflection in shape estimation. As mentioned in the introduction, one common way is to decouple shape and reflectance estimation using reflectance invariants or a specular-free image representation as in Yoon and Kweon (2006), Mallick et al. (2005), Zickler et al. (2008). In this approach, input images are transformed so that they are free from specular reflection and the resultant images are then used for surface shape recovery. Reflectance can then be recovered using the estimated shape. This approach might be less computationally intensive and more robust/stable than the proposed

method because it is not required to consider the complex specular reflection during shape estimation. However, it is strongly limited by the specific lighting configuration—the transformation is valid only when the illumination conditions have specific properties. While it is also useful to detect specular highlights first and to treat them as outliers as in Hernández Esteban et al. (2008), Birkbeck et al. (2006), detecting specular highlights is a hard problem in itself.

The proposed method uses the Blinn-Phong shading model to describe specular reflectance, but it is also possible to use any other parametric reflectance model. The parametric reflectance model is directly incorporated in the problem formulation. As a result, the proposed method can be applied to various data sets, and is not limited by the specific lighting

Fig. 19 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) = (0.59 mm, 0.041, 0.047, 0.042, 0.226, 12.69), 1.0 mm completeness (shape, ρ_{dr}, ρ_{dg}, ρ_{db}, ρ_s, α_s) = (89.73%, 0.042, 0.047, 0.042, 0.226, 12.65), $e_{image} = 1.99$

(a) input image (b) ground-truth image (c) ground-truth specular image (d) ground-truth shading (e) initial shape

(f) synthesized image (g) synthesized diffuse image (h) synthesized specular image (i) estimated shading

configuration. This is one of the main contributions of the proposed method. However, some recent studies (Ngan et al. 2005; Stark et al. 2005) have shown conclusively that, although parametric reflectance models are widely used in the literature, they are often unable to capture important visual effects. This is especially true for specular reflectance. As a result, the proposed method may produce rather inaccurate results for real images.

Moreover, the proposed method can become less stable and less accurate as the scenario gets more complex. For example, it works very well for diffuse images taken under static illumination, which is the simplest case, because we do not need to consider the surface normal and the photometric normal of each point. However, when dealing with non-Lambertian surfaces, the estimates of shape and reflectance are rather less accurate. In some aspects, this is natural because the proposed method deals with many spatially varying unknowns and estimates them alternatively. As described, the proposed method consists of many sub-parts, which also have alternative loops in them. It suffers from local minima and the alternative scheme sometimes fails in practice. In addition, for non-Lambertian cases, the proposed method may produce inaccurate results because specular reflection can be extremely sensitive to the surface normal depending on the surface shininess. On the other hand, when specular reflectance occurs broadly and its magnitude is small compared to that of diffuse reflectance, the problem can be ambiguous and the proposed method may return poor results because it is hard to distinguish the specular reflectance from the diffuse one in this case. The accuracy of specular reflectance estimation depends on the number of

specular reflection observations as well. However, by combining multiple cues, the dependency on the initial shape can be reduced.

Two other difficulties of the proposed method are computational time and huge memory requirement. In fact, the proposed method is more computationally expensive than shape and reflectance decoupling approaches. All sub-loops should converge at each iteration as shown in Figs. 1, 2, 4, and 5, and the shape can not evolve much at each iteration because of the stability. Therefore, the proposed method takes from a few hours to a few days according to the image sets and the initial conditions—the computational time is in proportion to the resolution of the grid used in the level set framework, the number of input images, and the number of light sources. In addition, contrary to previous works, we rigorously consider camera and light visibility and self-occlusion. This incurs huge memory requirements, precluding the use of a dense level set grid. Therefore, we have some difficulties to deal with complex shaped objects in experiments, which is why we used the smoothed bimba data set in Fig. 13. In addition, because of the sparseness of the level set grid, the initial shape given by rough silhouette images is not accurate, so the initially computed specular reflectance is not close to the ground truth. However, even when we use a small number of images, we estimate the specular reflectance at each iteration and this improves the accuracy of the method.

In other respects, the proposed method needs several user-specified parameters such as τ, λ, and β in (27). These parameters control the contribution of the individual cost functions, according to the types of image sets. As a result,

the overall performance may be biased towards specific priors according to these parameters.

As a future work, we would like to develop faster, more stable, and more accurate schemes to overcome these limitations. Also, there is some other room for improvement. For example, we used the simple L^2-norm for computing e, while it is possible to use other global/robust measures such as cross correlation or mutual information as in Pons et al. (2007). It is also possible to change the inner product structure as proposed by Charpiat et al. (2007). In addition, in this work, we adopted the Blinn-Phong model to describe specular reflectance, but it should be relevant to adopt a more realistic model.

8 Conclusion

In this paper, we have presented a variational method that recovers both the shape and the reflectance of surfaces using multiple images, assuming that illumination conditions and camera calibration are known in advance. Scene recovery is achieved by minimizing 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 non-Lambertian surfaces showing strong specular reflection, which is difficult even in some other state of the art methods using complex similarity measures.

References

Bhat, D. N., & Nayar, S. K. (1998). Stereo and specular reflection. *International Journal of Computer Vision*, 26(2), 91–106.

Birkbeck, N., Cobzas, D., Sturm, P., & Jägersand, M. (2006). Variational shape and reflectance estimation under changing light and viewpoints. In *European conference on computer vision* (Vol. 1, pp. 536–549).

Blinn, J. F. (1997). Models of light reflection for computer synthesized pictures. In *SIGGRAPH '77: Proceedings of the 4th annual conference on computer graphics and interactive techniques* (pp. 192–198). New York: ACM. http://doi.acm.org/10.1145/563858.563893.

Charpiat, G., Maurel, P., Pons, J. P., Keriven, R., & Faugeras, O. (2007). Generalized gradients: Priors on minimization flows. *International Journal of Computer Vision*, 73(3), 325–344.

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

Gargallo, P. (2008). *Contributions to the Bayesian approach to multiview stereo*. PhD thesis, Institut National Polytechique de Grenoble, France.

Gargallo, P., Prados, E., & Sturm, P. (2007). Minimizing the reprojection error in surface reconstruction from images. In *IEEE international conference on computer vision*.

Georghiades, A. S. (2003). Incorporating the Torrance and Sparrow model of reflectance in uncalibrated photometric stereo. In *IEEE international conference on computer vision* (Vol. 02, pp. 816–823).

Goesele, M., Curless, B., & Seitz, S. M. (2006). Multi-view stereo revisited. In *IEEE conference on computer vision and pattern recognition* (Vol. 2, pp. 2402–2409).

Goldman, D. B., Curless, B., Hertzmann, A., & Seitz, S. M. (2005). Shape and spatially-varying BRDFs from photometric stereo. In *IEEE international conference on computer vision* (pp. 341–348).

Hernández Esteban, C., Vogiatzis, G., & Cipolla, R. (2008). Multiview photometric stereo. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(3), 548–554.

Hertzmann, A., & Seitz, S. M. (2005). Example-based photometric stereo: Shape reconstruction with general, varying BRDFs. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(8), 1254–1264.

Jin, H., Yezzi, A., & Soatto, S. (2002). Variational multiframe stereo in the presence of specular reflections. In *International symposium on 3D data processing visualization and transmission* (pp. 626–630).

Jin, H., Yezzi, A. J., Tsai, Y. H., Cheng, L. T., & Soatto, S. (2003). Estimation of 3D surface shape and smooth radiance from 2D images: A level set approach. *Journal of Scientific Computing*, 19(1–3), 267–292.

Jin, H., Cremers, D., Yezzi, A. J., & Soatto, S. (2004). Shedding light on stereoscopic segmentation. In *IEEE conference on computer vision and pattern recognition* (Vol. 1, pp. 36–42).

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

Jin, H., Cremers, D., Wang, D., Prados, E., Yezzi, A., & Soatto, S. (2008). 3-d reconstruction of shaded objects from multiple images under unknown illumination. *International Journal of Computer Vision*, 76(3).

Kim, J., Kolmogorov, V., & R Zabih, R. (2003). Visual correspondence using energy minimization and mutual information. In *IEEE international conference on computer vision* (pp. 1033–1040).

Kolev, K., Klodt, M., Brox, T., & Cremers, D. (2007a). Propagated photoconsistency and convexity in variational multiview 3d reconstruction. In *Workshop on photometric analysis for computer vision*, Rio de Janeiro, Brazil.

Kolev, K., Klodt, M., Brox, T., Esedoglu, S., & Cremers, D. (2007b). Continuous global optimization in multiview 3d reconstruction. In *LNCS: Vol. 4679. Energy minimization methods in computer vision and pattern recognition (EMMCVPR)*, Ezhou, China (pp. 441–452). Berlin: Springer.

Lee, H. C., Breneman, E. J., & Schulte, C. P. (1990). Modeling light reflection for computer color vision. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(4), 402–409.

Lu, J., & Little, J. (1995). Reflectance function estimation and shape recovery from image sequence of a rotating object. In *IEEE international conference on computer vision* (pp. 80–86).

Mallick, S. P., Zickler, T., Kriegman, D. J., & Belhumeur, P. N. (2005). Beyond Lambert: Reconstructing specular surfaces using color. In *IEEE conference on computer vision and pattern recognition* (Vol. 2, pp. 619–626).

Ngan, A., Durand, F., & Matusik, W. (2005). Experimental analysis of BRDF models. In *Eurographics symposium on rendering* (pp. 117–126).

Osher, S., & Fedkiw, R. (2002). *The level set method and dynamic implicit surfaces*. Berlin: Springer.

Osher, S., & Sethian, J. A. (1988). Fronts propagating with curvature-dependent speed: Algorithms based on Hamilton-Jacobi formulations. *Journal of Computational Physics*, *79*, 12–49. citeseer.ist.psu.edu/osher88fronts.html.

Paris, S., Sillion, F. X., & Quan, L. (2006). A surface reconstruction method using global graph cut optimization. *International Journal of Computer Vision*, *66*(2), 141–161.

Pons, J. P., Keriven, R., & Faugeras, O. (2005). Modelling dynamic scenes by registering multi-view image sequences. In *IEEE conference on computer vision and pattern recognition* (Vol. 2, pp. 822–827).

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

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

Schultz, H. (1994). Retrieving shape information from multiple images of a specular surface. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *16*(2), 195–201. http://dx.doi.org/10.1109/34.273732.

Seitz, S. M., Curless, B., Diebel, J., Scharstein, D., & Szeliski, R. (2006). A comparison and evaluation of multi-view stereo reconstruction algorithms. In *IEEE conference on computer vision and pattern recognition* (pp. 519–528).

Sethian, J. (1999). *Level set methods and fast marching methods: evolving interfaces in computational geometry, fluid mechanics, computer vision, and materials sciences. Cambridge monograph on applied and computational mathematics*. Cambridge: Cambridge University Press.

Snow, D., Viola, P., & Zabih, R. (2000). Exact voxel occupancy with graph cuts. In *IEEE conference on computer vision and pattern recognition* (Vol. 1, pp. 345–352).

Soatto, S., Yezzi, A. J., & Jin, H. (2003). Tales of shape and radiance in multi-view stereo. In *IEEE international conference on computer vision* (pp. 974–981).

Solem, J. E., & Overgaard, N. C. (2005). A geometric formulation of gradient descent for variational problems with moving surfaces. In *Scale-space* (pp. 419–430).

Stark, M., Arvo, J., & Smits, B. (2005). Barycentric parameterizations for isotropic BRDFs. *IEEE Transactions on Visualization and Computer Graphics*, *11*(2), 126–138.

Tran, S., & Davis, L. (2006). 3d surface reconstruction using graph cuts with surface constraints. In *European conference on computer vision* (Vol. 2, pp. 219–231).

Vogiatzis, G., Favaro, P., & Cipolla, R. (2005). Using frontier points to recover shape, reflectance and illumination. In *IEEE international conference on computer vision* (Vol. 1, pp. 228–235).

Vogiatzis, G., Hernández Esteban, C., Torr, P. H. S., & Cipolla, R. (2007). Multiview stereo via volumetric graph-cuts and occlusion robust photo-consistency. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *29*(12), 2241–2246.

Yang, R., Pollefeys, M., & Welch, G. (2003). 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* (pp. 576–583).

Yezzi, A., & Soatto, S. (2003). Stereoscopic segmentation. *International Journal of Computer Vision*, *53*(1), 31–43.

Yoon, K. J., & Kweon, I. S. (2006). Correspondence search in the presence of specular highlights using specular-free two-band images. In *Asian conference on computer vision* (pp. 761–770).

Yu, T., Xu, N., & Ahuja, N. (2004). Recovering shape and reflectance model of non-lambertian objects from multiple views. In: *IEEE conference on computer vision and pattern recognition* (pp. 226–233).

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

Zach, C., Sormann, M., & Karner, K. (2006). High-performance multi-view reconstruction. In *3DPVT* (pp. 113–120). IEEE Computer Society.

Zhou, W., & Kambhamettu, C. (2002). Estimation of illuminant direction and intensity of multiple light sources. In *European conference on computer vision* (pp. 206–220).

Zickler, T. (2006). Reciprocal image features for uncalibrated Helmholtz stereopsis. In *IEEE conference on computer vision and pattern recognition* (pp. 1801–1808).

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

Zickler, T., Mallick, S. P., Kriegman, D. J., & Belhumeur, P. (2008). Color subspaces as photometric invariants. *International Journal of Computer Vision*, *79*(1), 13–30.