HAL
open science

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

Kuk-Jin Yoon, Emmanuel Prados, Peter Sturm

To cite this version:

Kuk-Jin Yoon, Emmanuel Prados, Peter Sturm. Joint Estimation of Shape and Reflectance using Multiple Images with Known Illumination Conditions. International Journal of Computer Vision, 2010, 86 (2-3), pp.192-210. 10.1007/s11263-009-0222-4 . inria-00266293

HAL Id: inria-00266293

https://inria.hal.science/inria-00266293

Submitted on 21 Mar 2008

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.

Noname manuscript No.
(will be inserted by the editor)

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

Kuk-Jin YOON · Emmanuel PRADOS · Peter STURM

Received: date / Accepted: date

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. Contrary to previous work which considers and specializes in a specific scenario, our method applies indiscriminately with a number of classical scenarios; in particular it works for classical stereovision, multiview photometric stereo and multiview shape from shading. Moreover, unlike most previous methods dealing with only Lambertian surfaces, the proposed method considers general dichromatic surfaces. We verify the method using synthetic and real data sets containing specular reflection.

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

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

K.-J. YOON
Perception Team, INRIA Rhône Alpes, France
Tel.: +33 476 61 54 47
Fax: +33 476 61 54 54
E-mail: Kuk-Jin.Yoon@inrialpes.fr

E. PRADOS
Perception Team, INRIA Rhône Alpes, France
Tel.: +33 476 61 52 27
Fax: +33 476 61 54 54
E-mail: Emmanuel.Prados@inrialpes.fr

P. STURM
Perception Team, INRIA Rhône Alpes, France
Tel.: +33 476 61 52 32
Fax: +33 476 61 54 54
E-mail: Peter.Sturm@inrialpes.fr

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 the surface radiance/reflectance properties was somewhat 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 work such as the ones of Goesele et al (2006); Kolev et al (2007a,b); Pons et al (2005, 2007); Tran and Davis (2006); Zach et al (2006) compute the 3D shape without considering radiance estimation. However, the radiance/reflectance estimation has become a concern in multiview reconstruction scenarios in the last decade. For example, Jin, Soatto et al. estimate conjointly the 3D shape and radiance (tensors) (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)).

Roughly speaking, the radiance is the combination of the lighting, the reflectance, and the geometry of the scene. For example, radiance contains shading and shadows and, from raw radiance, it is impossible to correct them when changing the lighting. Therefore, recovering reflectance properties is required for realistic relighting, which is also fundamental, for example, in virtual reality as well as augmented reality where the lighting conditions when capturing the object are different from the ones where one resynthesizes it. *In this paper, we propose a method for jointly estimating the geometrical shape and the full reflectance of the surfaces of a scene from multiple images.* This can be also understood as a separation of geometry, reflectance, and illumination from the radiance.

In real life applications, perfect Lambertian surfaces do not exist. For that reason, multiview stereo algorithms have to be *robust to specularities*. A number of ideas have been exploited to improve the robustness of the algorithms, also being exhaustive is clearly impossible. A widespread idea consists in using some *similarity measures*, see for example 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). The weakness being then the similarity measures proposed are not generally valid under the general lighting conditions and/or *not physically motivated*. Another common strategy consists in modifying the input images in order to *remove highlights* as if the original surface was purely lambertian; see Mallick et al (2005); Yoon and Kweon (2006); 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 the specular reflection component is similar to the spectral energy distribution of the incident light. Nevertheless, these methods are strongly limited by the specific lighting configuration. Also, Yoon and Kweon (2006) is valid only for single (uniformly) colored illumination conditions, and, 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 work with up to two different illuminant colors the because images have three color channels in general. In addition, this approach is only valid when ambient light also has the same color. Similarly, some authors do not consider the image pixels where potentially there are highlights; in a sense they consider these points as outliers. As an example, Hernandez Esteban et al (2008) applies such a strategy. In such an approach, the idea is then to work only on data one is able to well model (and so to ignore what is too complicated to model). The authors have then to increase data (i.e. the number of input images) in order to compensate for the loss of information. For example, this last strategy cannot apply to stereo (with two cameras). 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 the non-Lambertian effects, it also worth to cite the work of Jin et al. (Jin et al (2005)) which considers *radiance tensor* and then does not need similarity measures. However, when some similarity measures such as normalized cross correlation (Faugeras and Keriven (1998); Pons et al (2005, 2007)) could help to be robust to some changes of illumination, the radiance tensor as presented in Jin et al (2005) is *not appropriated when the database contains images of the scene lighted by several conditions*.

In this paper, our goal is to provide a shape and reflectance estimation method that is global[1] 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 exploits prolifically all the photometric phenomena*, as it is explicitly done in photometric stereo methods. Also, it thus allows to naturally deal with a set of images taken under several lighting conditions.

Let us note that actually there already exist recent works that provide solutions in this direction. In particular Yu et al (2004, 2007) propose a model-based method for recovering the 3D shape and the reflectance properties of a non-Lambertian object. Nevertheless, in this last paper, the authors constrain the object to be made by a *single 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) who focuses on the Lambertian case. To our knowledge, the limitation to surfaces with single textureless material is in fact common to (almost) all the work going in same direction as ours; 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). To our knowledge, only two previous similar works are able to recover scenes with varying albedo: Birkbeck et al (2006) and Hernandez Esteban et al (2008). But in Birkbeck et al (2006); Hernandez Esteban et al (2008), they tried to filter out specular highlights by using a simple thresholding. As a result, they used only diffuse components to estimate the shape. Hernandez Esteban et al (2008) also used a thresholding to detect shadowed pixels that are not visible from light sources, which is however not working under multiple light sources. On the other hand, in Birkbeck et al (2006), the authors computed the light visibility using the surface normal and the light direction. Finally, let us emphasize that the method of Hernandez Esteban et al (2008) is specifically a (multiview) photometric stereo method (which needs several different lighting configurations); also their algorithm cannot perform classical stereo-vision.

In our work, we do not want to restrain ourself to a single textureless material: in other words, the reflectance properties of the object can spatially and strongly change. In effect, now a day, more and more objects are now printed and so it is fundamental to be able to recover textured and patterned objects. In return, of course, we will not be able to recover lighting conditions as done in Jin et al (2008), and we have to use a parallel process which

[1] in the sense that it simultaneously and consistently optimizes the shape and reflectance.

return them. In this work, we assume that lighting conditions are known in advance. Practically, we can use spherical objects with the reference white color to capture the directions and the colors of light sources (Powell et al (2001); Zhou and Kambhamettu (2002)).

More generally, one of the goal of this paper is to show that the joint computation of the shape and the reflectance is benefit from several point of views. In addition to provide 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 the 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 the changes of illumination, as in photometric stereo. Finally, this allows to easily incorporate some constraints on the reflectance and so in particular this allows to exploit naturally the shading effects in textureless regions (and thus 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).
Finally, let us emphasize that, contrary to previous works that consider a specific scenario, the method we propose here can be applied indiscriminately to a number of classical scenarios; in particular it works for classical stereovision, multiview photometric stereo and multiview shape from shading.

The paper is organized as follows. In section 2 we describe the modeling assumptions and we specify the notations. In section 3 we formulate the problem in the Bayesian framework; we then detail the asssociated cost functions in section 4. In section 5 we precisely explain how we are minimizing the global energy. We show some experimental results on synthetic and real images data sets in section 6. We conclude after a discussion about difficulties and future work (section 7).

2 Modeling Assumptions and Notations

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

2.1 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 the pixels). It is split into two parts: the pixels corresponding to the foreground, $\pi_{iF} = \pi_i \cap \Pi_i(S)$, and the other points $\pi_{iB} = \pi_i \setminus \pi_{iF}$ (associated to the background). $I_i : \pi_i \to \mathbb{R}^c$ is the image of the true scene, captured by the ith camera ($c = 1$ for a gray-scale image, and $c = 3$ for a color image). We denote I the set of input images: $I = \{I_1, I_2, \cdots, I_{n_c}\}$; I_{iF} and I_{iB} are the restrictions of the function I_i to π_{iF} and π_{iB}, respectively. In other respects, we consider the visibility function $v_S^i : \mathbb{R}^3 \to \mathbb{R}$ 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 it

is occluded by the foreground object from the same camera. 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 closest point on S along the ray joining $\mathbf{X}$ to the optical center of the ith camera.

2.2 Lighting Conditions

We assume that the scene is illuminated by a finite number of distant point light sources. We complete them by adding an ambient light term (which partially accounts for other complex phenomena), with constant energy radiated isotropically in all directions. Note that, based on Wiener's theorems, Vedaldi and Soatto (2006) shows that such a light distribution can approximate arbitrarily well any positive distribution on the sphere. Let n_{il} be the number of illuminants corresponding to the ith camera and $\mathbf{l}_{ij} \in \mathbb{S}^2$ and $L_{ij} \in \mathbb{R}^c$ be the direction and the intensity[2] of the jth illuminant of the ith camera, respectively. Similarly, $L_{ia} \in \mathbb{R}^c$ is the intensity[2] of the ambient illumination of the ith camera. As mentioned before, we can manage the set of images taken under several lighting conditions, while some cases with special lighting conditions will be also discussed later.

2.3 Modeling the Foreground Surface

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

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 unknown variables, we chose to represent the reflectance by a parametric model. Of course the chosen model directly depends on the applications aimed at; as an example, we consider the popular Blinn-Phong shading model. In this 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 then decomposed as

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

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

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

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

where $\rho_d(\mathbf{X}) \in \mathbb{R}^c$ is the diffuse albedo[2] at point $\mathbf{X}$, $\mathbf{n}(\mathbf{X})$ is the normal vector to the surface S at $\mathbf{X}$ and $v_{L_{ij}}$ represents the light visibility function: $S_{L_{ij}}$ being the part of S visible from the jth illuminant of the ith camera, we define $v_{L_{ij}}(\mathbf{X}) = 1$ if $\mathbf{X} \in S_{L_{ij}}$, $v_{L_{ij}}(\mathbf{X}) = 0$ otherwise.

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

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

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

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 in the scene and modeled as

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

where L_a is defined above.

By combining the diffuse, specular, and ambient reflection, 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

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

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

As suggested by Yezzi and Soatto (2003), to be sure that the estimated foreground surface does not shrink to an empty set (which is indeed the global optimum for most cost functionals used in other works) it is crucial to define and characterize the background. The choice of model is dictated by the scenario and the applications. For example, in Jin et al (2004); Yezzi and Soatto (2003), the background is characterized by its radiance which is constrained to be constant or strongly regular. At the opposite extreme, when the background is quite irregular, one can assume that one has at his disposal the background images, i.e. the images of the scene captured by the same cameras without foreground objects. In this work, we deal with the latter scenario. Therefore, in addition to the images I, we assume that we detain the background images $\tilde{I} = \{\tilde{I}_1, \cdots, \tilde{I}_{n_c}\}$. Finally, we define $\tilde{I}_{iF}$ and $\tilde{I}_{iB}$ analogously to I_{iF} and I_{iB}.

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 a scene surface Ω, that maximize $P(\Omega|I)$ for given 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, $P(I|\Omega) = P(I|S,R)$ is a likelihood and $P(S)$ and $P(R)$ are priors on the shape and reflectance respectively.

3.1 Likelihood

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

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

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 typical and reasonable prior for the surface shape S is about the area[3] or about the smoothness of a surface. When using the surface area for the prior on S, it is expressed as

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

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, $R = (R_d, R_s)$. We express our prior as $P(R) = P(R_d)P(R_s)$ under the assumption that R_d and R_s are independent. Here, $P(R_d)$ and $P(R_s)$ can be assumed uniform in general so that $P(R)$ is constant. However, unfortunately, estimating reliable specular reflectance for all surface points with the uniform prior is very difficult unless there are enough observations exhibiting specular reflection at 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 observations[4]. It is physically valid to assume that specular reflectance varies smoothly within each homogeneous material surface patch. It is, however, also very difficult to partition Ω according to the types of materials. Instead, we use the diffuse reflectance of a surface as a soft constraint to partition Ω and define the prior on the surface reflectance as

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

where $\omega(R)$ is a function of the intrinsic gradient of the diffuse and specular reflectance of a surface. This function will be defined below.

[3] In this case, a minimal surface that may be characterized as the surface of minimal surface area for 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 section 5.3.3.

4 Description of the Cost Functions

Based on the derivations in section 3, the problem is formulated as

$$\begin{aligned} P(\Omega|I) &\propto P(I|\Omega)P(\Omega) \\ &= P(I|S,R)\ P(S,R) \\ &= P(I|S,R)\ P(S)\ P(R) \\ &\propto \prod_{i=1}^{n_c} \exp\big(-\xi_i(\Omega)\big) \times \Big(\exp\big(-\psi(S)\big)\Big) \\ &\quad \times \Big(\exp\big(-\omega(R)\big)\Big), \end{aligned} \tag{11}$$

and it can be expressed in terms of cost functions as

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

This shows that maximizing the probability (Eq. (11)) is equal to minimizing the total cost (Eq. (12)).

4.1 Data Cost Function

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

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

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

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

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

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

4.1.1 Similarity Measure

When computing ξ, any statistical correlation among color or intensity patterns such as the sum of squared differences (SSD), cross correlation (CC), and mutual information (MI) can be used[6]. In any case, ξ can be expressed as the integral over the image plane as

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

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

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

where $\hat{e}_i(\mathbf{x}) = e_i\big(I_i(\mathbf{x}), \bar{I}_i(\mathbf{x})\big) - e_i\big(I_i(\mathbf{x}), \tilde{I}_i(\mathbf{x})\big)$. 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 Eq. (5), surface appearance (i.e., the data cost function) is dependent on both the surface normal and position, and this makes the problem hard to solve and unstable. To resolve this problem, we introduce a 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. In this case, Eq. (6) is written in terms of $\mathbf{v}$ as

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

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

$$\begin{aligned} E_{dev}(\Omega) &= \tau \int_S \chi(\mathbf{X}) d\sigma = \frac{\tau}{2} \int_S \|\mathbf{n}(\mathbf{X}) - \mathbf{v}(\mathbf{X})\|^2 d\sigma \\ &= \tau \int_S (1 - (\mathbf{n}(\mathbf{X}) \cdot \mathbf{v}(\mathbf{X}))) d\sigma, \end{aligned} \tag{18}$$

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

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

where λ is a control constant.

[6] In fact, we do not need to use any sophisticated measure because we also recover the surface reflectance. Instead, we can use the simple pixel-wise measure.

4.3 Reflectance Discontinuity Cost Function

Based on the assumption on surface reflectance in section 3.3, we define a discontinuity cost function of surface reflectance as

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

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

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

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

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

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

with a pre-defined constant M[7]. ∇_S denotes the intrinsic gradient defined on S.

The proposed discontinuity cost function of surface reflectance makes the discontinuities of specular reflectance generally coincide with the discontinuities of diffuse reflectance, which is physically valid in general. Accordingly, surface points that do not have enough specular observations get assigned specular reflectance inferred from the specular reflectance of neighboring surface points.

4.4 Total Cost Function

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

$$\begin{aligned} E_{total}(\Omega) &= E_{data}(\Omega) + E_{dev}(\Omega) + E_{shape}(S) + E_{refl}(R) \\ &= C + \sum_{i=1}^{n_c} \int_{\pi_{iF}} \hat{e}_i(\mathbf{x}) d\sigma_i + \tau \int_S \chi(\mathbf{X}) d\sigma \\ &\qquad + \lambda \int_S d\sigma + \beta \int_S f(\mathbf{X}) d\sigma. \end{aligned} \tag{24}$$

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

$$\begin{aligned} E_{data}(\Omega) &= C - \sum_{i=1}^{n_c} \int_{S_i} \left(\hat{e}_i(\Pi_i(\mathbf{X})) \frac{\mathbf{d}_i(\mathbf{X}) \cdot \mathbf{n}(\mathbf{X})}{z_i(\mathbf{X})^3}\right) d\sigma \\ &= C - \int_S \left(\sum_{i=1}^{n_c} v_S^i(\mathbf{X}) \hat{e}_i(\Pi_i(\mathbf{X})) \frac{\mathbf{d}_i(\mathbf{X}) \cdot \mathbf{n}(\mathbf{X})}{z_i(\mathbf{X})^3}\right) d\sigma. \end{aligned} \tag{26}$$

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

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

$$E_{total}(\Omega) = C + \int_S \left(-\sum_{i=1}^{n_c} \left(v_S^i \hat{e}_i \frac{\mathbf{d}_i \cdot \mathbf{n}}{{z_i}^3} \right) + \tau\chi + \lambda + \beta f \right) d\sigma. \tag{27}$$

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

$$g(\mathbf{X}, \mathbf{n}(\mathbf{X})) = \left(-\sum_{i=1}^{n_c} \left(v_S^i \hat{e}_i \frac{\mathbf{d}_i \cdot \mathbf{n}}{{z_i}^3} \right) + \tau\chi + \lambda + \beta f \right), \tag{28}$$

Eq. (24) is simply rewritten as

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

Here, although the total cost function is an integral over the surface, it does not suffer from the usual minimal surface bias: most functionals used in multiple stereo have an empty set as globally optimal surface, since they do not "explain" all pixels in the input images. Our approach, like Yezzi and Soatto (2003), takes into account all pixels in the cost function, using both the estimated foreground and the available background information.

5 Scene Recovery

Recently, via graph cuts or convexity, some authors have proposed some global optimization methods for the classical multiview stereovision problem; see Kolev et al (2007a,b); Paris et al (2006); Snow et al (2000); Vogiatzis et al (2007). 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 previous section. 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 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 corresponding to the cost functions respectively. The final gradient descent flow is then given by

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

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

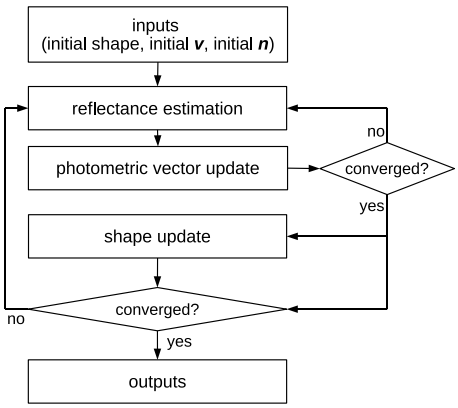


Fig. 1 Overall procedure of the proposed method

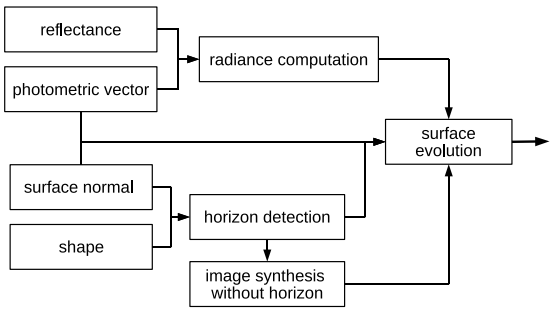


Fig. 2 Shape update

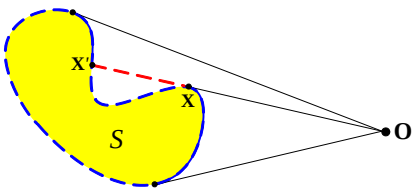


Fig. 3 Horizon point **X** and its terminator point **X**′

5.1.1 Gradient Descent Flow for the Data Cost

As shown in Eq. (26), the data cost is a function of the visibility of a surface point, which is dependent on the whole surface shape. According to Gargallo et al (2007) which extends of the work of Solem and Overgaard (2005) and the one of Yezzi and Soatto (2003) for correctly dealing with the visibility of non-convex objects, $S_t\big|_{data}$ is given by

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

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 $\mathbf{X}$ as shown

in Fig. 3 (for more details see Gargallo (2008); Gargallo et al (2007)). 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. (5) as

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

where

$$\nabla \mathbb{L}_{ij} = \nabla \mathbb{L}^d_{ij} + \nabla \mathbb{L}^s_{ij}, \tag{33}$$

and

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

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

This gradient descent flow includes both the variation related to the camera visibility changes (the first term in Eq. (31)) and the variation related to the image changes (the second term in Eq. (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 Jin et al (2004, 2008), the gradient descent flows for the normal deviation cost $S_t\big|_{dev}$ (originating from $E_{dev}(\Omega)$) is

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

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

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

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.

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

Here,

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

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

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

where $\mathrm{II}(\mathbf{t})$ is the second fundamental form for a tangent vector $\mathbf{t}$ with respect to $\mathbf{n}$.

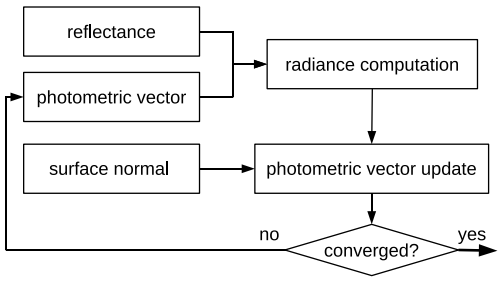


Fig. 4 Photometric unit vector update

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 Eq. (44) directly. Since $\mathbf{v} \in \mathbb{S}^2$, $\mathbf{v}$ can be expressed 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 $\mathbf{v}$ in the spherical coordinates. 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

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

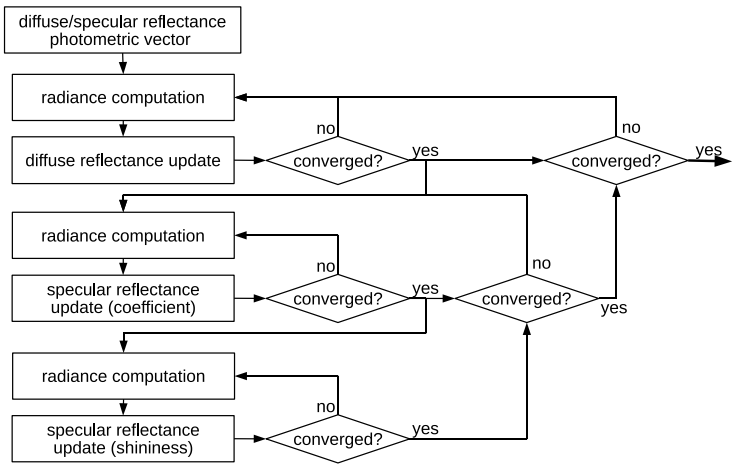


Fig. 5 Reflectance estimation

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 a fixed S, still minimizing the total cost function. Since E_{dev} and E_{shape} do not depend on R at all, we seek an optimal R by minimizing $(E_{data}(\Omega) + E_{refl}(R))$. 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. We repeat the procedure until they no longer change. Fig. 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

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

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

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

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:

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

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 given as

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

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 Eq. (54).

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

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

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

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 Single-Material Surface

When dealing with a single-material surface that has a single specular reflectance R_s, it is possible to set $\rho_s(\mathbf{X}) = \rho_s$ and $\alpha_s(\mathbf{X}) = \alpha_s$ 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 Eq. (21) is zero everywhere on the surface. Hence, the gradient descent flow is then given by

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

and the PDE used for the estimation of ρ_d , Eq. (53), is simplified as

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

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 Fedkiw (2002); Osher and Sethian (1988); Sethian (1999). In the level set formulation, hypersurfaces are represented as the zeros of a continuous function[8].

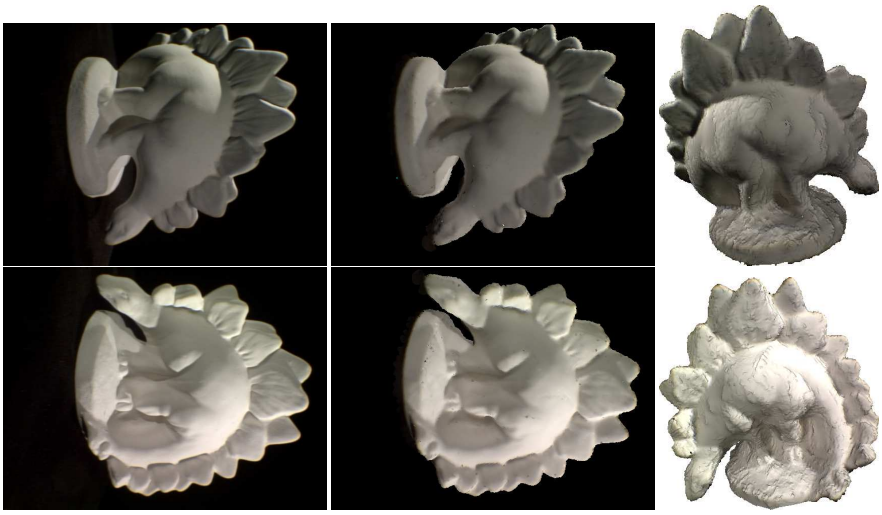
The proposed method starts with the visual hull obtained by rough silhouette images to reduce computational time and to avoid local minima, and we also adopt the multi-scale strategy. 640×480 or 800×600 images were used as inputs and and the simple L^2-norm was used to compute the image similarity, e. The camera and light visibility were computed by using the OpenGL z-buffering[9]. In experiments using real images, we detected saturated pixels by thresholding the intensity as $I_i(\mathbf{x}) > I_{th}$, where we set $I_{th} = 253$, and ignored them.

6.2 Experimental results

Due to the generality of the proposed method, it can be applied to various types of images sets with different cameras/lights configurations. First, when the images of Lambertian surfaces were taken under the fixed illumination condition while changing viewpoints, the proposed method can be applied even without lighting information, assuming that there is only ambient lighting. Therefore, we do not need to take care of the surface normal and only

[8] $\phi(\mathbf{X},t) > 0$ for the outside and $\phi(\mathbf{X},t) < 0$ for the inside.

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



(a) input images (b) synthesized images (c) estimated shape

Fig. 6 Result for the "dino" image set (16 images) – textureless Lambertian surface case (fixed lighting)

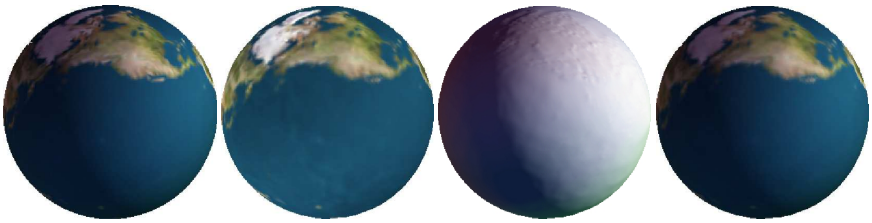
$E_{data}(\Omega)$ and $E_{shape}(S)$ are computed. In this case, the ICA (intensity conservation assumption) is valid in all images, and the proposed method works as the conventional multiview stereo methods and estimates the shape and the radiance. The result for the "dino" image set (16 images) is given in Fig. 6. As shown in Fig. 6, the proposed method successfully recovers the shape as well as the radiance.

The proposed method can be also applied to images taken under varying illumination conditions. The results using the simple images of textureless/textured Lambertian surfaces are shown in Fig. 7 – Fig. 12. Figure 7 shows the ground-truth shape of the "bimba" image set (18 images) and the estimation result. The surface is with the uniform diffuse reflectance and input images were taken under different lighting configurations. 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, the points with black diffuse reflectance are because they are not visible from any cameras and/or any light sources. On the other hand, Fig. 8 shows one of 32 textured input images and the synthesized image generated by using the estimated shape (i.e., shading) and reflectance. Based on this result, we can also generate the images of a scene with different lighting conditions as shown in Fig. 9. The result for a more complex object is shown in Fig. 10 and Fig. 11. The images synthesized by using the estimation closely resemble input images while the shading and the reflectance are successfully separated, and it is possible to synthesize an image with different lighting conditions even at a different viewpoint. 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); Hernandez Esteban et al (2008), we do not use any thresholding to filter out specular highlight pixels when dealing with images with specular reflection. The result for the smoothed "bimba" data set is shown in Fig. 13. In this case, the surface is with the uniform diffuse/specular reflectance and each images was taken with different illumination conditions. Here, we used the method described in section 5.3.3 to estimate the specular reflectance. Al-

(a) ground-truth model (b) estimated model

Fig. 7 Result for the "bimba" image set (18 images) – textureless Lambertian surface case



(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

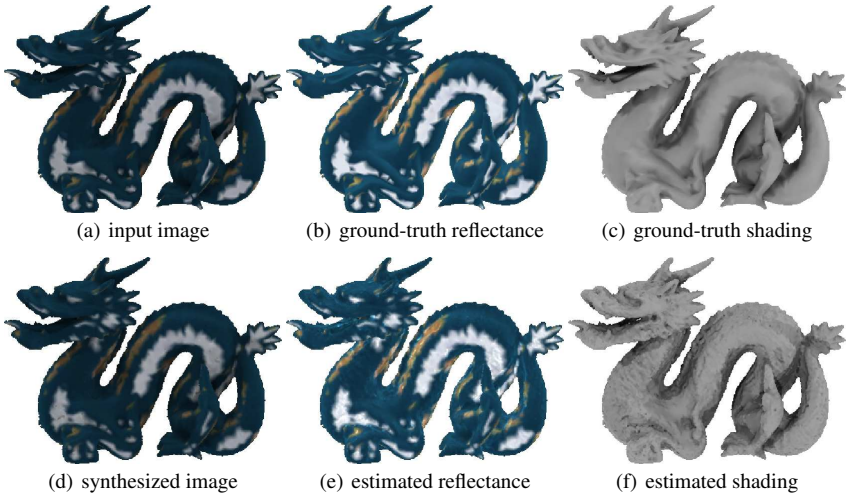
(a) original input image (b) synthesized image

Fig. 9 Scene synthesis under different lighting conditions

though 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 estimated is (ρ_s=0.61, α_s=41.8)[10].

We also used the real image sets of textured glossy objects, which were taken by using fixed cameras/light sources with rotating objects as in Birkbeck et al (2006); Hernandez Esteban et al (2008) — in this case, each image has different lighting conditions and has specular reflection. The light position and color were measured by using a white sphere. Figure 14 shows one image among 59 input images and the initial shape obtained by using silhouette and the estimation result. Here, we assumed that the single-material surface. More results using real image sets are also shown in Fig. 15 – Fig. 16. $(72 \times 72 \times 72)$ grids were used for the "saddog" and "duck" image sets while $(64 \times 64 \times 64)$ grids were used for the "bunny" image set. Although sparse grid volumes were used, the proposed method

[10] In fact, the small error of surface normals can cause the large of specular reflectance because of its sensitivity to surface normals. For instance, $0.7 \times (0.98)^{50} (= 0.255) \approx 0.61 \times (0.979)^{41.8} (= 0.251)$.



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

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

Fig. 10 Result for the "dragon" image set (32 images) – textured Lambertian surface case with the same lighting conditions for all the input images

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

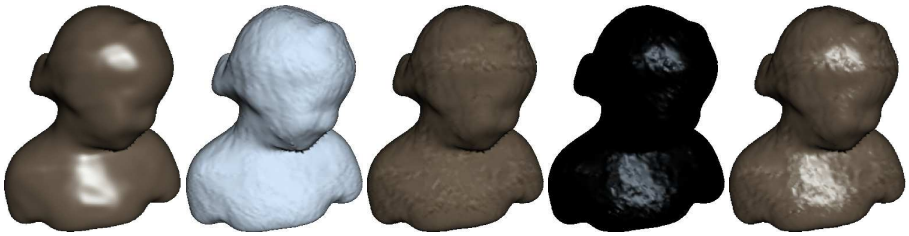
(d) reconstructed scene relighted as in (a) and viewed from the same point of view as (a) (e) estimated reflectance (f) estimated shading

Fig. 11 Synthesized result with different lighting conditions and viewed from a viewpoint different of all the input image viewpoint. A comparison with the ground-truth is possible because it is synthetic data

successfully estimated the shape of the glossy object even under specular reflection while estimating specular reflectance. Here, we can see that, 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 that often occurs on glossy surfaces, it really helps to recover the shape well.

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

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



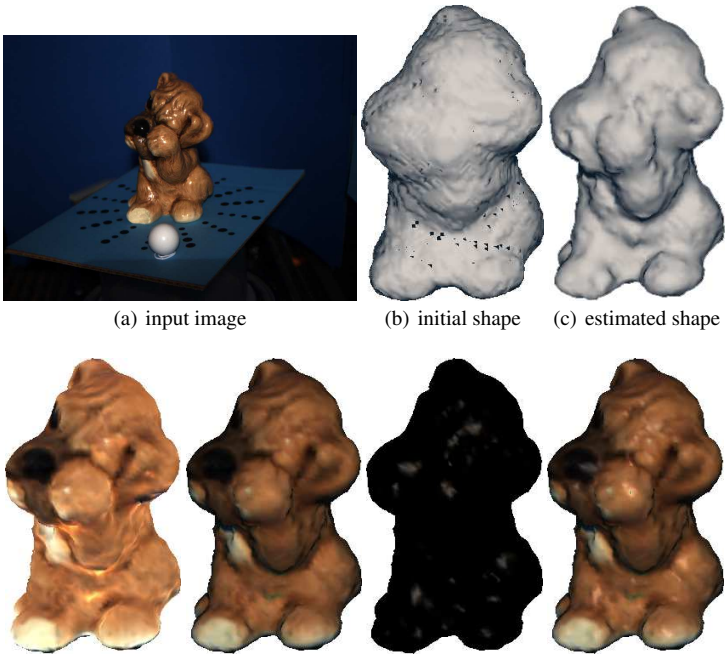
(a) ground-truth model (b) estimated shape (c) diffuse image (d) specular image (e) synthesized image

Fig. 13 Result for the smoothed "bimba" image set (36 images) - textureless non-Lambertian surface case

7 Discussion and Further Work

As shown in section 6, the proposed method can be applied to various cases thanks to the generality. This is one of the main contributions of the proposed method. However, at the same time, the proposed method can become more unstable and inaccurate as the case gets more complex. For example, the proposed method works very well for diffuse images taken under fixed illumination conditions, 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 estimations of the shape and the reflectance are rather less accurate. In some aspect, this is natural because the proposed method deals with many unknowns and estimates the unknowns alternatively. As described, the proposed method is consists of many sub-parts, which also have alternative loops in them. Also, it suffers from local minima and sometimes in practice the alternative scheme fails. In addition, for non-Lambertian cases, the proposed method may produce inaccuracy results because specular reflection can be extremely sensitive to the surface normal depending to the surface shininess.

Another difficulty for non-Lambertian surfaces is the computational time. All sub-loops should converge at each iteration as shown in Fig. 1, Fig. 2, Fig. 4, and Fig. 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 number of grids used in the level set framework, the number of input images, and the number of light sources.



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

(d) diffuse reflectance (e) diffuse image (f) specular image (g) synthesized image

Fig. 14 Result for the "saddog" image set (58 images)

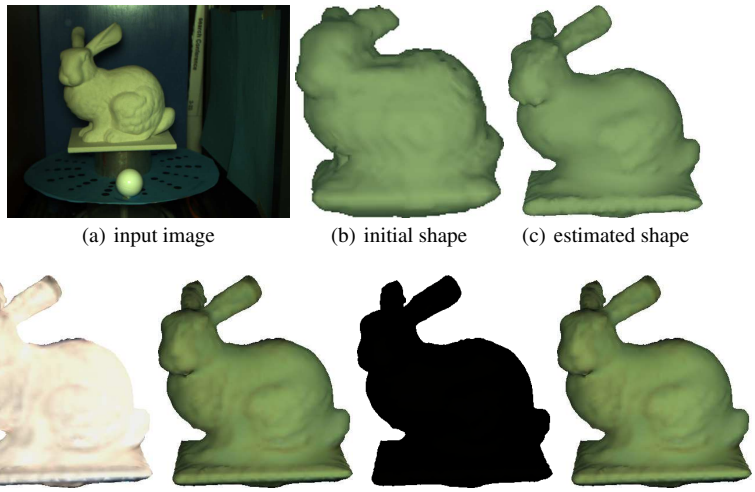
In other respects, the proposed method needs several user-specified parameters such as τ, λ, and β in Eq. (27). These parameters control the contribution rates of cost functions according to the types of image sets. As a result, the overall performance may be under a bias toward specific prior according to these parameters.

As a future work, we would like to develop faster and more stable schemes to overcome these limitations. Also, there is much room for improvement. For example, in this work, we adopted the Blinn-Phong model to describe the specular reflection, but it should be relevant to adopt a more realistic model. In addition, we used the simple L^2-norm for 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 finally also possible to change the inner product structure as propose by Charpiat et al (2007).

8 Conclusion

In this paper, we have presented a variational method that recovers both the shape and the reflectance of scene surfaces using multiple images, assuming that illumination conditions and cameras calibration are known in advance. Scene recovery was achieved by minimizing the global cost functional alternatively. As a result, the proposed method produced the complete description of a scene surface.

Contrary to any previous work that considers a specific scenario, our method can be applied indiscriminately to a number of classical scenarios; in particular it works for classical



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

(d) diffuse reflectance (e) diffuse image (f) specular image (g) synthesized image

Fig. 15 Result for the "bunny" image set (26 images)

stereovision, (multiview) photometric stereo and multiview shape from shading. Moreover, unlike most previous methods dealing with only Lambertian surfaces, the proposed method considers general dichromatic surfaces.

Acknowledgements We would like to thank Neil Birkbeck for providing real data sets for our experiments and Jean-Philippe Pons and CERTIS for providing their C++ level-set library. This work was supported by the Korea Research Foundation Grant funded by the Korean Government(MOEHRD) (KRF-2006-352-D00087) and by the Flamenco project ANR-06-MDCA-007.

References

Bhat DN, Nayar SK (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

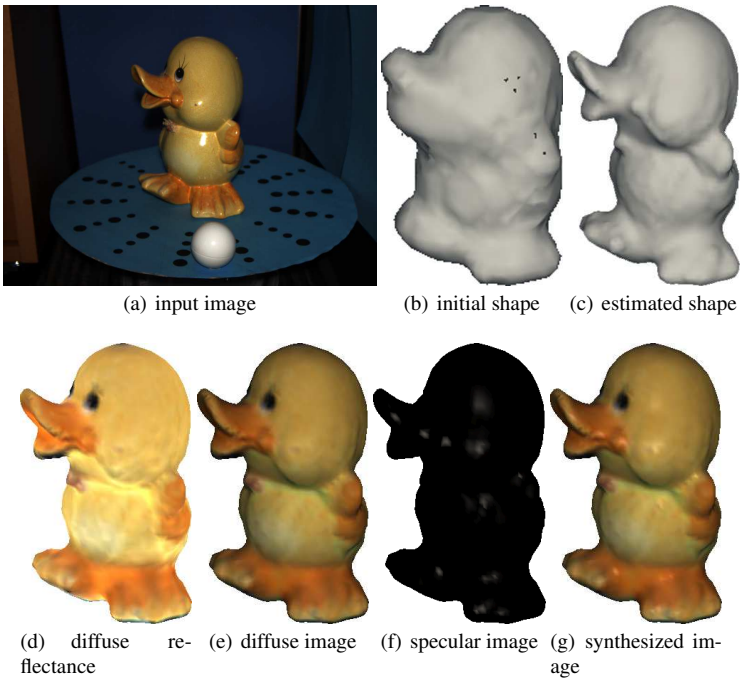
Charpiat G, Maurel P, Pons JP, Keriven R, Faugeras O (2007) Generalized gradients: Priors on minimization flows. Int J Comput Vision 73(3):325–344, DOI http://dx.doi.org/10.1007/s11263-006-9966-2

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 multi-view 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 AS (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



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

(d) diffuse reflectance (e) diffuse image (f) specular image (g) synthesized image

Fig. 16 Result for the "duck" image set (28 images)

Goesele M, Curless B, Seitz SM (2006) Multi-view stereo revisited. In: CVPR'06, IEEE Computer Society, vol 02, pp 2402–2409

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

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 AJ, Tsai YH, Cheng LT, 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 AJ, Soatto S (2004) Shedding light on stereoscopic segmentation. IEEE Conference on Computer Vision and Pattern Recognition 01:36–42

Jin H, Soatto S, Yezzi AJ (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: Energy Minimization Methods in Computer Vision and Pattern Recognition (EMMCVPR), Springer, EZhou, China, LNCS, vol 4679, pp 441–452

Lee HC, Breneman EJ, Schulte CP (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, Zickler T, Kriegman D, Belhumeur P (2005) Beyond lambert: reconstructing specular surfaces using color. Computer Vision and Pattern Recognition, 2005 CVPR 2005 IEEE Computer Society Conference on 2:619–626 vol. 2

Osher S, Fedkiw R (2002) The Level Set Method and Dynamic Implicit Surfaces. Springer-Verlag

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

Paris S, Sillion FX, Quan L (2006) A surface reconstruction method using global graph cut optimization. Int J Comput Vision 66(2):141–161

Pons JP, 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 JP, 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 MW, 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

Seitz SM, 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 University Press

Snow D, Viola P, Zabih R (2000) Exact voxel occupancy with graph cuts. CVPR'00 01:1345

Soatto S, Yezzi AJ, Jin H (2003) Tales of shape and radiance in multi-view stereo. In: IEEE International Conference on Computer Vision, pp 974–981

Solem JE, Overgaard NC (2005) A geometric formulation of gradient descent for variational problems with moving surfaces. In: Scale-Space, pp 419–430

Tran S, Davis L (2006) 3d surface reconstruction using graph cuts with surface constraints. In: ECCV06, pp II: 219–231

Vedaldi A, Soatto S (2006) Viewpoint invariance for non-planar scenes. Tech. Rep. TR050012, UCLA CSD

Vogiatzis G, Favaro P, Cipolla R (2005) Using frontier points to recover shape, reflectance and illumunation. In: IEEE International Conference on Computer Vision, vol 1, pp 228–235

Vogiatzis G, Esteban CH, Torr PHS, Cipolla R (2007) Multiview stereo via volumetric graph-cuts and occlusion robust photo-consistency. IEEE Trans Pattern Anal Mach In-

tell 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 KJ, Kweon IS (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, IEEE Computer Society, vol 0, pp 113–120

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 PN, Kriegman DJ (2002) Helmholtz stereopsis: Exploiting reciprocity for surface reconstruction. International Journal of Computer Vision 49(2-3):215–227

Zickler T, Mallick SP, Kriegman DJ, Belhumeur P (2008) Color subspaces as photometric invariants. To appear in the International Journal of Computer Vision