Shedding Light on Stereoscopic Segmentation

Hailin Jin[†] hljin@cs.ucla.edu
Daniel Cremers[†] cremers@cs.ucla.edu
Anthony J. Yezzi[‡] ayezzi@ece.gatech.edu
Stefano Soatto[†] soatto@cs.ucla.edu

† Computer Science Department, University of California, Los Angeles, CA 90095
‡ School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, GA 30332

Abstract

We propose a variational algorithm to jointly estimate the shape, albedo, and light configuration of a Lambertian scene from a collection of images taken from different vantage points. Our work can be thought of as extending classical multi-view stereo to cases where point correspondence cannot be established, or extending classical shape from shading to the case of multiple views with unknown light sources. We show that a first naive formalization of this problem yields algorithms that are numerically unstable, no matter how close the initialization is to the true geometry. We then propose a computational scheme to overcome this problem, resulting in provably stable algorithms that converge to (local) minima of the cost functional. Although we restrict our attention to Lambertian objects with uniform albedo, extensions of our framework are conceivable.

1 Introduction

We are interested in recovering the geometry of a scene from multiple images taken from different vantage points. This is one of the classical problems of Computer Vision, and at this level of generality it does not admit a simple solution. Indeed, unless we impose appropriate priors, this problem does not admit a meaningful solution at all. In addition to scene geometry, images depend on scene reflection properties as well as illumination, and given any number of images, there exist infinite scene geometries and reflectance/illumination configurations that generate them. This structural lack of identifiability is normally approached by making assumptions on some of the unknowns (e.g. reflectance and/or illumination) to infer properties of the others (e.g. shape), as we illustrate below.

Most multi-view stereo reconstruction algorithms[1] rely on establishing point-to-point correspondence between multiple views of the same scene. Correspondence is usually based on various "feature descriptors," which are image statistics designed to be invariant or insensitive to local deformations (typically affine) of the domain of the image due to changes in the viewpoint, and to local deformations of the co-domain (intensity) of the image (also typically affine) due to changes in illumination or reflectance. Such descriptors range from the simple intensity of a window around a feature point detected using various corner detectors and compared using cross correlation or least squares [7, 17] to more elaborate descriptors that involve gradient histograms at various scales [16, 22]. Unfortunately, in general, given an object with arbitrary shape and arbitrary reflection, one can make the local appearance around any point arbitrary by acting on the illumination. This lack of identifiability of shape and reflectance/illumination is present even if one assumes that the scene is Lambertian [1]: local feature-based stereo relies on the tacit assumption that the appearance is independent of the viewpoint. This is tantamount to assuming that objects are self-luminous, in the sense that they radiate energy equally in all directions at a given point P according to a certain function $\rho(P)$, which can be thought of as the "texture map." The notion of illumination becomes superfluous, and all appearance information is encoded into ρ. Furthermore, in order to establish point-to-point correspondence for the entire scene, we need to assume that ρ is a "sufficiently exciting texture": it must have nowhere-zero gradient. In fact, for feature detectors and descriptors to work reliably, one needs ρ to be discontinuous, leading to the requirement of the gradient being everywhere-infinite. Naturally, in practice it is sufficient for the gradient of ρ to be sufficiently high at sufficiently many places on the scene, but then it is possible for feature descriptors to be ambiguous for scenes that have high-contrast repetitive patterns. So, for traditional stereo algorithms to work well, one needs not too much texture, and not too little texture. When this is not the case, one cannot reliably establish correspondence between individual points in different images.[2]

What happens when the assumptions that allow estab-

[1] This does not include photometric stereo since we consider a changing viewpoint.

[2] Recent variational algorithms for multi-view stereo ameliorate the situation by establishing a global correspondence between views in a way that allows filling in regions of low texture gradient with minimal surfaces [5]; however, the underlying assumptions remain the same.

lishing correspondence from image-to-image are not satisfied? Many approaches recently have followed a paradigm where there is no explicit correspondence from image to image; instead, all images are matched to an underlying model of the scene. Such a model must necessarily describe the geometry as well as the reflectance properties of the scene.

We will assume that the scene has ideal Lambertian reflection and is illuminated by a constant (but unknown) ambient term. In addition, we assume a finite number of point light sources of unknown intensity in unknown position. Therefore, all the variability of image appearance (shading) is generated by the interaction of light and surface geometry. This problem can be thought of as a multi-view version of the problem of Shape from Shading, when the number and direction of the light sources are unknown. From this viewpoint, this work can be considered as the natural extension of the problem of stereoscopic shading [11] to unknown illumination.

Given the assumption that the scene is Lambertian, has constant albedo, and is viewed from multiple vantage points, we pose the problem in the framework of infinite-dimensional optimization in order to find the best shape (a surface) and light configuration that give rise to the images. This is done by setting up a cost functional and computing its first derivation, which is then used to define a gradient flow to evolve an initial surface towards a (local) minimum of the chosen cost functional. Unfortunately, as we will see in Section 3.1, a straightforward choice of a cost functional will result in an *unstable flow*, which prohibits convergence, even to a local minimum. A great deal of work in this paper is therefore devoted to computational models that allow for stable gradient flows.

1.1 Related work and our contributions

The literature on shape from shading is far too extensive for us to review here. A collection of earlier work can be found in the book edited by Horn and Brooks [10]. Zhang et al. wrote a nice survey on more recent methods [28]. The effect of changing lighting on the object appearance (for fixed viewpoint) was analyzed by Belhumeur et al. [26, 3].

Estimating the light direction first and subsequently reconstructing the shape (i.e. depth map) was performed by Zheng and Chellappa [29]. Recently, Samaras and Metaxas [21] suggested to instead account for the coupling of light and shape by reconstructing both in an alternating fashion. Our work differs from the latter mainly in two ways: Firstly, we use a variational framework, which implies that the interlaced estimation of lighting and geometry are both derived by minimizing a single cost functional. Secondly, we consider multiple views, which enables us to reconstruct a complete 3D object (rather than a depth map). In particular, this requires to also take into account the visibility of light source and camera for all points on the estimated surface. Using variational methods in shape from shading dates back to the eighties [10, 19], and even level set methods have been employed before [13, 15]. The literature on stereo and motion is also extensive; we refer the reader to the recent vision textbook [8] for references.

A closely related work is that of Faugeras and Keriven [5], who cast the traditional multi-frame stereo in a variational framework and use the level set method to solve it. This work differs from ours in that the authors perform image-to-image matching. They do not consider lighting, in fact, an incorporation of lighting in their framework is not straightforward. Issues concerning the fusion of shading cues with parallax cues have been discussed in several works, including [23, 14, 26, 18, 6, 4]. Recently in [27], it was suggested to combine ideas from shape-from-shading, photometric stereo and structure-from-motion. Given scene geometry and reflectance, the illumination configuration of the scene can be established, as Yu and Malik have shown [25]. Here, because we do not know the geometry, we cannot handle such complex reflectance model, so we restrict our attention to Lambertian scenes.

In this manuscript, we propose an algorithm to estimate shape, albedo and illumination configuration from a collection of images of a Lambertian scene. To the best of our knowledge, we are the first to integrate illumination into variational multi-view stereo reconstruction. Our algorithm is provably stable, and naturally extends existing shape from shading algorithms as well as multi-view stereo algorithms where point correspondence cannot be established.

2 Problem Formalization

Let $S \in \mathbb{R}^3$ be a smooth surface. We denote with $\mathbf{X} = [X, Y, Z]^T$ the coordinates of a generic point on S with respect to a fixed reference frame. The central goal of this paper is to reconstruct the surface S from a set of n images $I_i : \Omega_i \to \mathbb{R}, i = 1, 2, \dots, n$, where $\Omega_i \subset \mathbb{R}^2$ is the domain of each image. Each image is fully calibrated, i.e., intrinsic and extrinsic calibration parameters are assumed known [8]. After pre-processing, each camera can therefore be modeled as an ideal perspective projection $\pi_i : \mathbb{R}^3 \to \Omega_i; \mathbf{X} \mapsto \mathbf{x}_i \doteq \pi_i(\mathbf{X}) = \pi(\mathbf{X}_i) = [X_i/Z_i, Y_i/Z_i]^T$, where $\mathbf{X}_i = [X_i, Y_i, Z_i]^T$ are the coordinates of $\mathbf{X}$ in the i-th camera reference frame. $\mathbf{X}$ and $\mathbf{X}_i$ are related by a rigid body transformation, which can be represented in coordinates by a rotation matrix $R_i \in SO(3)$ and a translation vector $T_i \in \mathbb{R}^3$: $\mathbf{X}_i = R_i \mathbf{X} + T_i$. We assume that there is a background B which covers the field of view of each camera. Without loss of generality, we assume B to be a sphere with infinite radius. We define the foreground projection to be the region $Q_i = \pi_i(S) \subset \Omega_i$ and denote its complement in Ω_i by Q_i^c. Although the perspective projec-

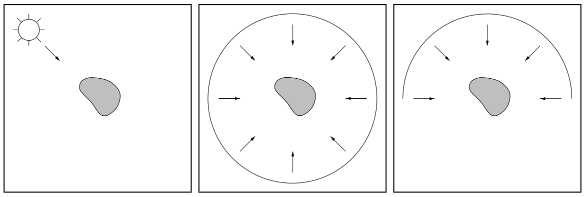


Figure 1: *We assume that the true light distribution illuminating a scene can be approximated by a superposition of the above components: a directional light* **(left)**, *an isotropic component* **(center)**, *and an isotropic component restricted to one hemisphere* **(right)**.

tion π_i is not one-to-one (and therefore not invertible), the back-projection $\pi_i^{-1} : \Omega_i \to \mathbb{R}^3$ of $\mathbf{x}_i$ onto S can be defined as the first intersection point with S of a ray starting from the i-th camera center and passing through $\mathbf{x}_i$.

We assume that both the foreground and background are Lambertian. For gray-scale images their radiances can therefore be modeled as scalar-valued functions[3]:

$$\rho : S \to \mathbb{R}, \quad \text{and} \quad h : B \to \mathbb{R}, \tag{1}$$

where for simplicity we consider the background radiance h to be constant – for an extension to smooth background radiances, we refer to [12]. We assume that the surface has a constant albedo, which, without loss of generality, we assume to be 1. Therefore, the varying image appearance is solely generated by the lighting configuration and the scene geometry. In general, the whole world reflects light, which induces a very complicated space-varying lighting configuration. For simplicity, in this work we will assume that the true light distribution can be approximated by a superposition of three different components shown in Figure 1. These three components are:
(1) Distant point light sources – see Figure 1 left – induce a radiance of the form

$$\rho(\mathbf{X}) = \lambda \langle \mathbf{N}, \mathbf{L} \rangle \, \xi(\mathbf{X}), \tag{2}$$

where $\mathbf{L}$ denotes the unit vector pointing in the direction of the light, λ the light amplitude, $\mathbf{N}$ the surface unit outward normal and $\xi : S \to \{0, 1\}$ the visibility of the light. In the case of convex objects, the visibility is given by $\xi = \mathcal{H}(\langle \mathbf{N}, \mathbf{L} \rangle)$, where $\mathcal{H}$ denotes the Heaviside step function. For non-convex objects or the case of multiple objects, where there are cast shadows, there is no general expression.
(2) An ambient component, given by a constant λ_0, permits to approximate lighting effects induced by multiple interreflections from the surroundings
(3) Isotropic light restricted to one hemisphere – see Figure 1 (right) – is to account for the fact that the object of interest is placed on a ground plane which generally does not emit much light compared to the light from the upper hemisphere. In the case of a convex object, the integration of intensity λ over the hemisphere results in a radiance of the form [9]:

$$\rho = \frac{\lambda}{2} \left(\langle \mathbf{N}, \mathbf{L} \rangle + 1 \right), \tag{3}$$

where $\mathbf{L}$ denotes the unit vector from the object pointing to the center of the hemisphere. Mathematically, this component can be represented as

$$\begin{aligned} \rho &= \frac{\lambda}{2} \Big(\max\big(\langle \mathbf{N}, \mathbf{L} \rangle, 0\big) + \min\big(\langle \mathbf{N}, \mathbf{L} \rangle, 0\big) + 1 \Big) \\ &= \frac{\lambda}{2} \Big(\langle \mathbf{N}, \mathbf{L} \rangle \, \xi(\mathbf{L}) - \langle \mathbf{N}, -\mathbf{L} \rangle \, \xi(-\mathbf{L}) + 1 \Big), \end{aligned} \tag{4}$$

which corresponds to a superposition of an ambient light with two point light sources from opposite directions, the second one having a negative amplitude.[4] Combining the above components, therefore amounts to a lighting model of the form:

$$\rho = \sum_{j=1}^{\ell} \lambda_j \, \langle \mathbf{N}, \mathbf{L}_j \rangle \, \xi_j + \lambda_0 \tag{5}$$

where the amplitudes $\lambda_j \in \mathbb{R}$ account for positive and negative light sources and $\lambda_0 \in \mathbb{R}_+$ for the ambient component.

3 Variational Formulation

In order to optimally reconstruct surface, light and albedo from a set of views we propose to minimize a cost functional of the form:

$$E(S, \lambda_j, \mathbf{L}_j) = E_{data}(S, \lambda_j, \mathbf{L}_j) \; + \; \alpha E_{prior}(S), \tag{6}$$

where the first term enforces the similarity between the observed images and the corresponding projections and the second term imposes a prior which favors smooth surfaces and is given by:

$$E_{prior}(S) = \int_S dA. \tag{7}$$

3.1 A direct approach

A straightforward formulation for the data term is:

$$E_{data} = \sum_{i=1}^{n} \int_S \Big(I_i(\pi_i(\mathbf{X})) - \langle \lambda_j \xi_j \mathbf{L}_j, \mathbf{N} \rangle - \lambda_0 \Big)^2 dA, \tag{8}$$

where, following the Einstein summation convention, we assume summation over the light sources j. To simplify the exposition, we neglected the background fitting term in (8).

Unfortunately, this direct approach leads to an unstable gradient flow for the surface S. We will detail this for the case of a single light source $\lambda \mathbf{L}$ and no visibility constraint, i.e., $\xi{=}1 \;\; \forall \mathbf{X} \in S$. The curvature dependent terms in the total flow are given by:

$$\sum_{i=1}^{n} \Big(2H \left((I_i - \lambda_0)^2 + 2\lambda^2 - 3 \, \langle \lambda \mathbf{L}, \mathbf{N} \rangle^2 \right) - 2\lambda^2 \Pi(\mathbf{N} \times \mathbf{L}) \Big)$$

[3] The formulation can easily be generalized to color images by using vector-valued radiance functions.

[4] The term “negative light” facilitates the treatment of hemispherical illumination.

where H denotes the mean curvature and Π the second fundamental form. In regions where the surface faces the light (and therefore $\langle \mathbf{L}, \mathbf{N} \rangle \to 1$) and where the modeled directional intensity λ^2 exceeds the measured one $(I_i - \lambda_0)^2$, the coefficients of H or Π are negative. Therefore, the resulting flow is numerically unstable in those regions. This poses a problem for estimating shape, albedo and light using the direct approach.

3.2 "Soft" shape from shading

The above instability arises due to the strong coupling between surface appearance and its normal in (8). In the presence of measurement noise, the surface will bend and ripple to fit the data. This behavior can be suppressed by increasing the surface regularization [11], yet this results in over-smoothed reconstructions. To circumvent this instability, we propose a relaxed cost functional in which the normal is decoupled from the surface through an auxiliary unit vector field

$$\mathbf{V} : S \to S^2, \mathbf{X} \mapsto \mathbf{V}(\mathbf{X}). \tag{9}$$

$\mathbf{V}$ will take the place of the unit normal in modeling the shading effects. We will show in Section 4 that the induced surface flow lacks the potentially unstable curvature-based diffusion terms. We will now be left with the problem of minimizing this new energy function jointly with respect to both S and $\mathbf{V}$. The data fitness can be measured in the sense of $\mathcal{L}^2$ for background and foreground as:

$$\begin{aligned} E_{data} = &\sum_{i=1}^{n} \int_{Q_i} \big(I_i(\mathbf{x}_i) - \big\langle \mathbf{V}(\pi_i^{-1}(\mathbf{x}_i)), \lambda_j \mathbf{L}_j \xi_j \big\rangle - \lambda_0\big)^2 d\Omega_i \\ &+ \sum_{i=1}^{n} \int_{Q_i^c} \big(I_i(\mathbf{x}_i) - h\big)^2 d\Omega_i. \end{aligned} \tag{10}$$

To stay faithful to the physically motivated interaction between the surface normal and the light source direction, we will introduce an indirect coupling between the unit normal and the modeled surface radiance by adding a second term to our energy which penalizes the average deviation between the true unit normal of the surface and the unit vector field $\mathbf{V}$ which takes its place in the new radiance model. The constraint for $\mathbf{V}$ is given by a penalty on the $\mathcal{L}^2$ distance between $\mathbf{V}$ and the unit normal field $\mathbf{N}$ on S:

$$E_{coupling} = \frac{1}{2} \int_S \|\mathbf{V} - \mathbf{N}\|^2 dA = \int_S \big(1 - \langle \mathbf{V}, \mathbf{N} \rangle\big) dA. \tag{11}$$

The overall cost functional is simply a weighted average of the three costs:

$$\begin{aligned} &E_{total} \;=\; E_{data} \;+\; \alpha E_{prior} \;+\; \beta E_{coupling} \\ &= \sum_{i=1}^{n} \int_{Q_i} (I_i - \langle \mathbf{V}, \lambda_j \mathbf{L}_j \xi_j \rangle - \lambda_0)^2 d\Omega_i + \sum_{i=1}^{n} \int_{Q_i^c} \big(I_i - h\big)^2 d\Omega_i \\ &+ \;\alpha \int_S dA \;+\; \beta \int_S \big(1 - \langle \mathbf{V}, \mathbf{N} \rangle\big) dA. \end{aligned}$$

When minimizing E_{total}, we need to guarantee that $\mathbf{V}$ is always a unit vector field, i.e. $\|\mathbf{V}(\mathbf{X})\|^2 = 1 \;\; \forall \mathbf{X} \in S$. Rather than imposing this constraint by augmenting the cost functional with a Lagrange multiplier, we will revert to a *projection method* detailed in Section 4.2.

4 Energy Minimization

4.1 Surface evolution

To facilitate finding the variation of the data fitness term with respect to S, we need to introduce two more terms. Let $\chi_i : S \to \mathbb{R}$ be the surface visibility function with respect to the i-th camera, i.e. $\chi_i(\mathbf{X}) = 1$ for points on S that are visible from the i-th camera and $\chi_i(\mathbf{X}) = 0$ otherwise. Let σ_i account for the change of coordinates from $d\Omega_i$ to dA, i.e, $\sigma_i = \frac{d\Omega_i}{dA} = \langle \mathbf{X}_i, \mathbf{N}_i \rangle / Z_i^3$, where $\mathbf{N}_i$ the unit normal $\mathbf{N}$ expressed in the i-th camera reference frame. We can now express the data term as follows

$$\begin{aligned} &\sum_{i=1}^{n} \int_{Q_i} \Big(\big(I_i - \langle \mathbf{V}, \lambda_j \mathbf{L}_j \xi_j \rangle - \lambda_0\big)^2 - \big(I_i - h\big)^2\Big) d\Omega_i \\ &\quad + \sum_{i=1}^{n} \int_{\Omega_i} \big(I_i - h\big)^2 d\Omega_i \\ &= \sum_{i=1}^{n} \int_S \chi_i \Big(\big(I_i - \langle \mathbf{V}, \lambda_j \mathbf{L}_j \xi_j \rangle - \lambda_0\big)^2 - \big(I_i - h\big)^2\Big) \sigma_i dA \\ &\quad + \sum_{i=1}^{n} \int_{\Omega_i} \big(I_i - h\big)^2 d\Omega_i \end{aligned}$$

It can be shown that the gradient descent flow minimizing the total energy is given by:

$$\begin{aligned} S_t = &\Big(\sum_{i=1}^{n} \frac{1}{Z_i^3} \big((I_i - (\langle \mathbf{V}, \lambda_j \mathbf{L}_j \xi_j \rangle + \lambda_0))^2 - (I_i - h)^2\big) \\ &\cdot \Big\langle \nabla \chi_i, R_i^T \mathbf{X}_i \Big\rangle - \sum_{i=1}^{n} 2\chi_i \big(I_i - \langle \mathbf{V}, \lambda_j \mathbf{L}_j \xi_j \rangle - \lambda_0\big) \\ &\big(\xi_j \lambda_j \mathbf{L}_j^T \nabla_S \mathbf{V} R_i^T \mathbf{X}_i + \sum_{j=1}^{l} \langle \mathbf{V}, \lambda_j \mathbf{L}_j \rangle \langle \nabla \xi_j, R_i^T \mathbf{X}_i \rangle\big) \\ &+ (2H(\alpha + \beta) - \beta \nabla_S \cdot \mathbf{V})\Big) \mathbf{N} \end{aligned} \tag{12}$$

Note that the only second order term (curvature term) in the flow (12) is $2H(\alpha + \beta)\mathbf{N}$, therefore the flow is always numerically stable (with a properly chosen time step). Another advantage of flow (12) is that it depends only upon the image values, *not the image gradients*. This property greatly increases the robustness of the resulting algorithm to image noise.

The numerical implementation of the flow (12) is carried out in the standard level set framework [20]. For more details on shape estimation using level set methods, we refer the reader to [5, 12].

4.2 Updating the auxiliary field

In alternation with the surface evolution, we minimize the cost functional with respect to the auxiliary field $\mathbf{V}$. The corresponding negative energy gradient is given by

$$U=-\frac{dE}{dV}=\sum_{i=1}^{n}2\chi_i(I_i-\langle\mathbf{V},\lambda_j\mathbf{L}_j\xi_j\rangle-\lambda_0)\lambda_j\mathbf{L}_j\xi_j\sigma_i+\beta\mathbf{N}.$$

In order to constrain the evolving vector field to the space of unit vector fields, we restrict its evolution to the tangent space of S^2 (which is the plane perpendicular to $\mathbf{V}$), yielding the following updating equation:

$$\mathbf{V}_t=U-\langle U,\mathbf{V}\rangle\,\mathbf{V}. \tag{13}$$

Once $\mathbf{V}$ is updated, we have to restrict it to the unit vector field, which can be achieved by simply normalizing the norm of the vector field to 1. The numerical implementation of the update equation (13) is also carried out in the level set framework following the practice of [2, 12]. The basic idea is to first extend the supporting domain of the vector field $\mathbf{V}$ from the surface S to $\mathbb{R}^3$ and then use standard finite difference schemes to implement equation (13). For more details, we refer the reader to [2, 12].

4.3 Updating the illumination

The ambient components for object and background can be determined in closed forms as follows. For the background, we get

$$h=\frac{\sum_{i=1}^{n}\int_{Q_i^c}I_i d\Omega_i}{\sum_{i=1}^{n}\int_{Q_i^c}d\Omega_i}, \tag{14}$$

which corresponds to the mean intensity estimated over the area outside the object. For the object, we get

$$\lambda_0=\frac{\sum_{i=1}^{n}\int_{Q_i}\big(I_i-\langle\mathbf{V},\lambda_j\mathbf{L}_j\xi_j\rangle\big)d\Omega_i}{\sum_{i=1}^{n}\int_{Q_i}d\Omega_i}, \tag{15}$$

which corresponds to the intensity not explained by the directional components averaged over the area of the object.

The optimal directional lighting can be found by gradient descent with respect to $\tilde{\mathbf{L}}_j=\lambda_j\mathbf{L}_j\in\mathbb{R}^3$, where the gradient is given by:

$$\frac{dE_{total}}{d\tilde{\mathbf{L}}_j}=\sum_{i=1}^{n}\int_{Q_i}\big(I_i-\langle\mathbf{V},\tilde{\mathbf{L}}_k\xi_k\rangle-\lambda_0\big)\mathbf{V}\xi_j d\Omega_i \tag{16}$$

As is well known, the above iterative solution is suboptimal: In order to gradually evolve the light to its optimal configuration, one needs to select sufficiently small time steps and define appropriate convergence criteria. Fortunately there exists a closed-form solution for the light configuration. The light gradient in (16) vanishes for:

$$\sum_{i=1}^{n}\int_{Q_i}(I_i-\lambda_0)\mathbf{V}\xi_j d\Omega_j=\sum_{i=1}^{n}\int_{Q_i}\mathbf{V}\mathbf{V}^T\tilde{\mathbf{L}}_k\xi_k\xi_j d\Omega_i\quad\forall j.$$

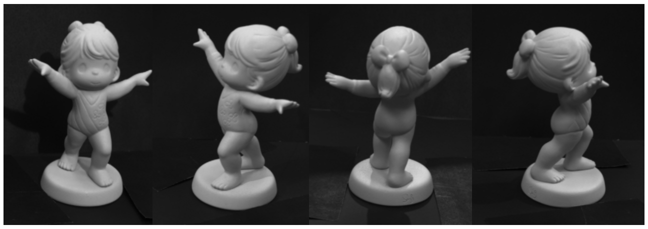


Figure 2: *Example views of the input data set consisting of 28 views of a doll on a table, illuminated by normal over-headed fluorescent lamps and an additional strong directional spot light. Despite a fairly uniform albedo, the appearance of the doll is strongly modulated by the light. For instance, the doll's head is much brighter than the rest because it is facing the spot light, whereas the back of the doll is almost as dark as the background.*

Assuming that the visibility ξ of the lights does not change much from one iteration to the next, i.e. $\xi(t+1)\approx\xi(t)$, the solution for $\tilde{\mathbf{L}}=(\tilde{\mathbf{L}}_1,\ldots,\tilde{\mathbf{L}}_\ell)$ is given by:

$$\tilde{\mathbf{L}}=M^{-1}p, \tag{17}$$

with the matrix $M\in\mathbb{R}^{3\ell\times3\ell}$ made up of the sub-matrices

$$M_{jk}=\sum_{i=1}^{n}\int_{Q_i}\mathbf{V}\,\mathbf{V}^T\,\xi_j\xi_k\,d\Omega_i,\quad j,k=1,\ldots,\ell, \tag{18}$$

and the vector $p\in\mathbb{R}^{3\ell}$ containing the sub-vectors

$$p_j=\sum_{i=1}^{n}\int_{Q_i}(I_i-\lambda_0)\mathbf{V}\,\xi_j\,d\Omega_i,\quad j=1,\ldots,\ell. \tag{19}$$

5 Experiments

We took 28 calibrated images of a doll figure of approximately uniform albedo standing on a table. The background is dark, and the doll is illuminated both by standard fluorescent overhead lamps and by an additional strong spotlight. Figure 2 shows 4 representative views which show how the light configuration modulates the object intensity from a very bright front of the head to very dark regions in the upper back.

Subsequently, we ran our implementation of stereoscopic segmentation [24], which corresponds to iterating the surface evolution in Section 4.1 with no directional lighting, i.e. $\lambda_j=0,_{j=1,\ldots,\ell}$, in alternation with the update of the ambient light components according to Section 4.3. Upon minimization, the surface converges to the result shown (from several viewpoints) in Figure 3: The object surface cannot be reconstructed correctly, since the assumption of constant object intensity is strongly violated.

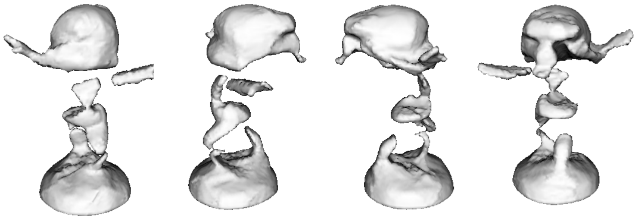
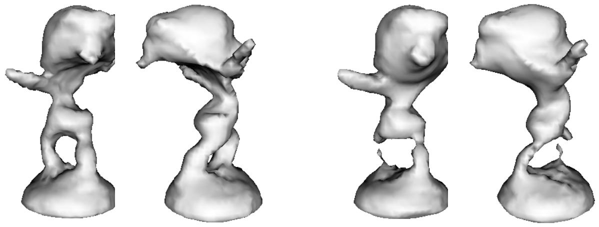


Figure 3: *Final shape estimated using [24]. The algorithm fails to reconstruct the doll, notably the legs and the back, because the assumption of constant radiance of the object is strongly violated by shading effects – see the input data in Figure 2.*



Positive spot light Negative spot light

Figure 4: *Final shape estimated using the enhanced stereoscopic segmentation with one directional light, either positive (left) or negative (right). The reconstruction is improved, because the simultaneously estimated directional light source accounts for some of the shading effects.*

In particular, the parts of the legs and back which are most strongly affected by shading are missing: due to their dark intensity they are ascribed to the background.

We then introduced one directional light source, which we randomly initialize. Subsequently, we run the enhanced stereoscopic segmentation, evolving the surface as detailed in Section 4.1 in alternation with an evolution of the auxiliary normal field given in Section 4.2 and an update of the ambient and directional lighting according to Section 4.3. It turns out that if the light is initialized close to the front of the doll, it is positive light and if the light is initialized close to the back of the doll, it is actually a negative light. Views of the final segmentation are shown (from various viewpoints) in Figure 4: The object is reconstructed more accurately, because the simultaneously estimated directional light source accounts for some of the shading effects on the surface.

Introducing two light sources (to account for the spot light and the lack of illumination from the ground plane) improves the reconstruction even more. In this case, the algorithm automatically returns one positive light, which faces the front of the doll and a negative light, which faces the back of the doll. Figure 5 shows 4 views of the reconstructed object generated by evolving the surface in alternation with updating ambient light and a positive and negative directional light.

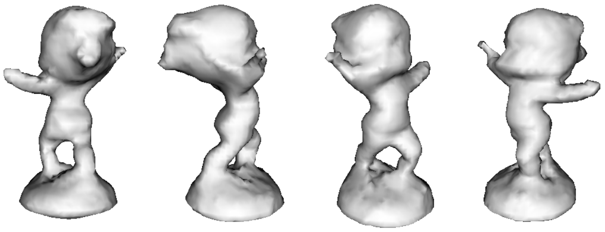


Figure 5: *Final shape estimated using the enhanced stereoscopic segmentation with two directional lights, a positive and a negative one. The algorithm reconstructs the 3D object much more accurately, because the simultaneously estimated light configuration allows to account for the shading effects in the input data.*

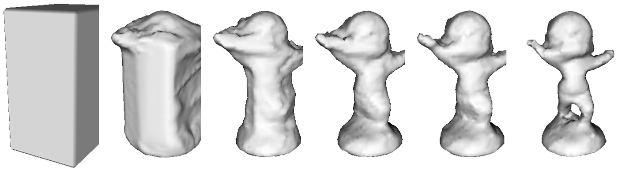


Figure 6: *Rendered views of the surface evolution, which starts from a cube containing the object and converges a solid model.*

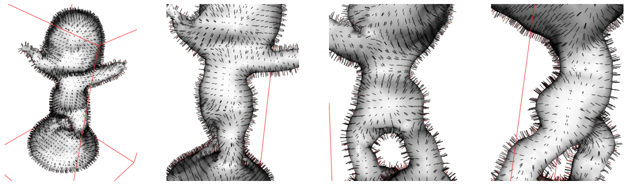


Figure 7: *Visualization of the auxiliary vector field* $\mathbf{V}$.

For completeness, we show in Figure 6 several steps indicating the evolution of the estimated surface from the initialization to the final segmentation and in Figure 7 a close-up on the estimated auxiliary normal field $\mathbf{V}$.

6 Conclusions

We proposed a variational framework to simultaneously estimate a 3D surface, the albedo and a light configuration from a set of calibrated views of a Lambertian object with uniform albedo. We extend the standard stereoscopic segmentation scheme with an explicit model of the the interaction of light with the object surface. In contrast to the extension of stereoscopic segmentation to piecewise smooth radiance functions [12], the smooth radiances in our formulation are induced by the physical interaction of light and surface.

Acknowledgements

This work is supported by ONR N00014-03-1-0850 and N00014-02-1-0720, NSF IIS-0208197 and CCR-0133736, AFOSR F49620-03-1-0095, and E-16-V91-G2.

References

[1] P. N. Belhumeur, D. J. Kriegman, and A. L. Yuille. The bas-relief ambiguity. *Int. J. of Computer Vision*, 35(1):33–44, 1999.

[2] M. Bertalmio, L. Cheng, S. J. Osher, and G. Sapiro. Variational problems and partial differential equations on implicit surfaces. *J. Comput. Phys.*, 174(2):759–780, December 2001.

[3] H. F. Chen, P. N. Belhumeur, and D. W. Jacobs. In search of illumination invariants. In *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition*, volume 1, pages 254–261, June 2000.

[4] J. Cryer, P.-S. Tsai, and M. Shah. Integration of shape from shading and stereo. *Pattern Recognition*, 28(7):1033–1043, July 1995.

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

[6] P. Fua and Y. G. Leclerc. Object-centered surface reconstruction: Combining multi-image stereo and shading. *Int. J. of Computer Vision*, 16(1):35–56, September 1995.

[7] C. Harris and M. Stephens. A combined edge and corner detector. In *Proc. 4th Alvey Vision Conference*, pages 189–192, 1988.

[8] R. Hartley and A. Zisserman. *Multiple view geometry in computer vision*. Cambridge University Press, 2000.

[9] B. K. P. Horn. *Robot Vision*. MIT Press, 1986.

[10] B. K. P. Horn and M. J. Brook. *Shape From Shading*. MIT Press, 1989.

[11] H. Jin, A. J. Yezzi, and S. Soatto. Stereoscopic shading: integrating multi-frame shape cues in a variational framework. In *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition*, volume 2, pages 169–176, June 2000.

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

[13] R. Kimmel, K. Siddiqui, B. B. Kimia, and A. M. Bruckstein. Shape from shading: level set propagation and viscosity solutions. *Int. J. of Computer Vision*, 16(2):107–133, October 1995.

[14] Y. G. Leclerc and A. F. Bobick. The direct computation of height from shading. In *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition*, pages 552–558, 1991.

[15] P. L. Lions, E. Rouy, and A. Tourin. Shape-from-shading, viscosity solutions and edges. *Numerische Mathematik*, 64(3):323–353, 1993.

[16] D. G. Lowe. Object recognition from local scale-invariant features. In *Proc. of Intl. Conf. on Computer Vision*, volume 2, pages 1150–1157, September 1999.

[17] B. D. Lucas and T. Kanade. An iterative image registration technique with an application to stereo vision. In *Int. J. Conf. on Artificial Intell.*, pages 674–679, 1981.

[18] N. Mukawa. Estimation of shape, reflection coefficients and illuminant direction from image sequences. In *Proc. of Intl. Conf. on Computer Vision*, pages 507–512, 1990.

[19] J. Oliensis and P. Dupuis. A global algorithm for shape from shading. In *Proc. of Intl. Conf. on Computer Vision*, pages 692–701, 1993.

[20] S. J. Osher and J. A. Sethian. Fronts propagating with curvature dependent speed: Algorithms based on hamilton-jacobi formulations. *J. Comput. Phys.*, 79(1):12–49, 1988.

[21] D. Samaras and D. Metaxas. Illumination constraints in deformable models for shape and light direction estimation. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 25(2):247–264, 2003.

[22] C. Schmid and R. Mohr. Local grayvalue invariants for image retrieval. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 19(5):530–535, May 1997.

[23] T. Simchony, R. Chellappa, and M. Shao. Direct analytical methods for solving poisson equations in computer vision problems. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 12(5):435–446, May 1990.

[24] A. J. Yezzi and S. Soatto. Stereoscopic segmentation. In *Proc. of Intl. Conf. on Computer Vision*, volume 1, pages 59–66, 2001.

[25] Y. Yu and J. Malik. Recovering photometric properties of architectural scenes from photographs. In *Proc. of ACM SIGGRAPH*, pages 207–217, July 1998.

[26] A. L. Yuille, D. Snow, R. Epstein, and P. N. Belhumeur. Determining generative models of objects under varying illumination: Shape and albedo from multiple images using svd and integrability. *Int. J. of Computer Vision*, 35(3):203–222, December 1999.

[27] L. Zhang, B. Curless, A. Hertzmann, and S. M. Seitz. Shape and motion under varying illumination: Unifying structure from motion, photometric stereo, and multi-view stereo. In *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition*, pages 618–625, October 2003.

[28] R. Zhang, P.-S. Tsai, J. E. Cryer, and M. Shah. Shape-from-shading: a survey. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 21(8):690–706, August 1999.

[29] Q. Zheng and R. Chellappa. Estimation of illuminant direction, albedo, and shape from shading. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 13(7):680–702, July 1991.