

Towards Robust and Physically Plausible Shaded Stereoscopic Segmentation

Dejun Wang, Emmanuel Prados, Stefano Soatto

► **To cite this version:**

Dejun Wang, Emmanuel Prados, Stefano Soatto. Towards Robust and Physically Plausible Shaded Stereoscopic Segmentation. Computer Vision and Pattern Recognition Workshop (CVPRW), Jun 2006, New York, United States. 10.1109/CVPRW.2006.204 . inria-00377430

HAL Id: inria-00377430

<https://hal.inria.fr/inria-00377430>

Submitted on 21 Apr 2009

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

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

Towards Robust and Physically Plausible Shaded Stereoscopic Segmentation

Dejun Wang[†]
wangdj@cs.ucla.edu

Emmanuel Prados[‡]
Emmanuel.Prados@inrialpes.fr

Stefano Soatto[†]
soatto@ucla.edu

[†] Vision Lab, Computer Science Department, University of California, Los Angeles, CA 90095
[‡] Perception lab, INRIA Rhone-Alpes, 655, avenue de l'Europe, 38330 Montbonnot, France.

Abstract

We address the multi-view shape from shading problem, that is the recovery of 3-D shape, lighting configuration and surface albedo from multiple calibrated views. Previous approaches to this problem relied on physically impossible illumination models (negative light) and only work on constant albedo and resulted in biased estimates of shape and lighting positions. Furthermore, since the solution involves infinite-dimensional optimization, existing approaches were quite slow. We develop a new model that explicitly enforces positivity in the light sources with the assumption that the object is Lambertian and its albedo is piecewise constant and show that the new model significantly improves the accuracy and robustness relative to existing approaches. Furthermore, we show that the most computationally expensive step in the optimization can actually be solved in closed form. This significantly improves speed of convergence over existing schemes. We illustrate the behavior of our algorithm directly on the same data used by previous authors, so direct comparison is possible.

1 Introduction

We propose a method to estimate the shape, albedo and illumination direction from a collection of multiple views of a Lambertian scene. This problem has been named multi-view shape from shading, although it also relates to stereo reconstruction. It is usually assumed that the views are calibrated, that is the mutual position and orientation of the cameras is known. This problem has been introduced by Jin et al. [1], who have shown that the optimal inference at its root is ill-posed. Following classical techniques of the shape from shading (SFS) literature [2], they have proposed an auxiliary vector field, that can be interpreted as a relaxation of the model constraints, and they have demonstrated their approach on a collection of real and synthetic images. Unfortunately, the illumination model of [1] is not

physically plausible, in that it contains light sources with negative radiance. Consequently, the recovered illumination has no relation to the actual illumination in the scene. We build on their work and go beyond it in a number of ways. First, we formulate the multi-view shading problem in a way that guarantees that the recovered illumination is physically plausible, i.e. has positive radiance. This model is more complex than [1] because it enforces inequality constraints. Second, we show that one of the steps of the optimization, indeed the most computationally intensive one where the (infinite-dimensional) auxiliary vector field is estimated using a gradient flow, can actually be solved in closed form. This results in significant improvements in speed and robustness. Third, we assume that albedo is piecewise constant, whereas [1] could only handle constant albedo. We illustrate our results on the same dataset of [1], available through the Web, and show radical improvements both in the estimation of shape, radiance, and the position of the light, all with significantly reduced computational complexity. Also our work differs from [3], which works on region segmentation of piecewise constant albedo, in that our image formation model accounts for the light sources and their variation explicitly.

While one may argue that our work is an improvement over existing work, the improvement is significant: on accuracy, robustness and speed at the same time, all while guaranteeing physical plausibility of the model. While existing approaches could not be easily integrated with other reconstruction modalities, for instance multi-view stereo, because of the presence of the auxiliary vector field and the non-physical nature of the light, that impinges on the estimate of reflectance, our model has the potential of being integrated with multi-view stereo, at least for Lambertian scenes, since the auxiliary vector field can be solved for in closed-form and therefore can be factored out of the reconstruction process. To the best of our knowledge, we are the first to propose the analysis of recovering shape, illumination and albedo under the piecewise constant albedo assumption.

1.1 Prior Work and our contribution

For general work on shape from shading we refer the reader to [4, 5, 6, 7, 8, 9]. Belhumer et al.[10, 11] analyzed the effect of changing lighting on the object appearance for fixed viewpoint. Estimating the light direction and shape in an alternating way was done by Samaras and Metaxas [12], which differs from our work mainly in that we consider multiple views, which enable us to recover the whole shape instead of a depth map only. Our work also differs from their work which combines shape from shading and stereo [13], because we do not use correspondences. Using variational methods in shape from shading dates back to the eighties [5, 14]. Yu and Malik showed how to recover illumination with known scene geometry and reflectance. Jin et al [3] presented the work on recovering shape for piecewise constant albedo without considering the light effects. In this article, we also restrict our attention to Lambertian scenes.

A closely related work is that of Jin et al. [1], who combine the reconstruction of shape and light configuration together and solve the variational problems via level set methods. The assumption of this work differs from ours in two points: 1) Jin et al.[1] allows the light, either a point light source or an ambient term, to be negative, which is not physically possible. 2) Jin et al. [1] deal with the case of constant albedo. They introduced an auxiliary vector field to replace the normal direction for increasing stability.

In this article, our approach mainly differs from [1] in the following aspects: First, we add the positivity constraint for the light sources of both ambient light and point source; second, albedo is assumed to be piecewise constant; third, we give a closed-form solution for the auxiliary vector field, which is provably correct. We demonstrate its performance on experiments with both synthesized data and real images and illustrate the stability, accurateness, robustness.

2 Problem Formalization

Let $S \in \mathbb{R}^3$ be a smooth surface. We denote with $X = [x, y, z]^T$ the coordinates of a generic point on S with respect to a fixed reference frame. The goal is to reconstruct the surface S and the light sources from a set of n images $I_i : \Omega_i \rightarrow \mathbb{R}, i = 1, \dots, n$, where $\Omega_i \subset \mathbb{R}^2$. The intrinsic and extrinsic calibration parameters for each image are assumed to be known [15]. Thus each camera can be modeled as a perspective projection $\pi_i : \mathbb{R}^3 \rightarrow \Omega_i; X_i \rightarrow x_i$, where X_i is the coordinates of X in the i -th camera reference frame. X and X_i are related by a rigid body transformation, that is $X_i = R_i X + T_i$. We assume that there is a background B covering the field of view of each camera. We also assume B to be a sphere with infinite radius and 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 . We also define the back-projection $\pi_i^{-1} : \Omega_i \rightarrow \mathbb{R}^3$ of x_i onto S , which is the first intersection of the ray through x_i in the i -th camera with S .

We assume that both the foreground and background are Lambertian. The radiances are modeled as positive scalar-valued functions: $r : S \rightarrow \mathbb{R}^+$, and $h : B \rightarrow \mathbb{R}^+$. The surface is assumed to have piecewise constant albedo. We assume that the true light configuration could be approximated by a superposition of two different components: 1) an ambient term E_0 that approximates inter-reflections, diffuse illumination and other short-range effects, and 2) a number of distant positive point light sources. For simplicity, we deal here with a single distance point light source. The extension to multiple point light sources is straightforward.

For an arbitrary point X , the radiance is $r(X) = \rho(X)(\langle N(X), L \rangle \xi(X) + E_0)$, where L characterizes the distance light source, i.e. the intensity of the point light source times the unit vector pointing in the direction of the light, N is the surface unit outward normal, $\xi : S \rightarrow \{0, 1\}$ is the visibility of the light and $\rho(\cdot)$ is the albedo of the point X . In the case of convex objects, visibility is given by $\xi = \mathcal{H}(\langle N, L \rangle)$, where \mathcal{H} denotes the Heaviside step function.

Thus we can write the image formation model as:

$$I(\pi(X)) = \rho(X)(\langle N(X), L \rangle \xi(X) + E_0). \quad (1)$$

Our model differs from [1] mainly in two aspects: 1) the albedo here is assumed to be piecewise constant, while [1] simply let it be 1; 2) we add a positive constraint on both ambient term and intensity of point light source for the sake of guaranteeing physical constraints.

3 Formalization of Energy Minimization

We assume that the albedo ρ is piecewise constant and that it partitions S into two regions D_1 and D_2 , which can have several disconnected components. We denote by C the union of the curves dividing D_1 and D_2 . We assume that these curves are smooth and closed. Also, we define ρ as follows: $\rho(X) = \rho_i \in \mathbb{R}$ for $X \in D_i, i = 1, 2$.

In order to recover the 3D shape of the scene S , the 2D curves C on the surface S , the albedos ρ_1 and ρ_2 of the scene, the intensity of the background h and the light sources, L and E_0 , we propose to minimize a energy func-

tional with respect to $\rho_1, \rho_2, E_0, L, C, S, h$ as follows:

$$\begin{aligned}
E &= E_{data} + \alpha E_{prior} + \beta E_{curv} + \gamma E_{coupling}. \\
E_{data} &= \sum_{i=1}^n \left(\int_{\pi_i(D_1)} (I_i(x_i) - \rho_1(\langle V, L \rangle \xi(\pi_i^{-1}(x_i)) + E_0))^2 d\Omega_i \right. \\
&+ \sum_{i=1}^n \left(\int_{\pi_i(D_2)} (I_i(x_i) - \rho_2(\langle V, L \rangle \xi(\pi_i^{-1}(x_i)) + E_0))^2 d\Omega_i \right. \\
&\quad \left. \left. + \sum_{i=1}^n \int_{Q_i^c} (I_i(x_i) - h)^2 d\Omega_i; \right. \right. \\
E_{prior} &= \int_S dA; \quad E_{coupling} = \int_S (1 - \langle V(X), N(X) \rangle) dA. \\
E_{curv} &= \int_C ds.
\end{aligned}$$

α, β, γ are coefficients that weigh the relative strengths of each regularizer, and are design parameters for the user to choose. π_i is the projection function from the 3D surface to 2D i th image. ξ is the light visibility function at the surface point $\pi_i^{-1}(x_i)$. χ_i is the visibility function of a point X on surface S with respect to the i -th camera, i.e. $\chi_i(X) = 1$ for points on S that are visible from the i -th camera and $\chi_i(X) = 0$ otherwise. σ_i represents the transformation of coordinates from $d\Omega_i$ to dA , i.e. $\sigma_i = \frac{d\Omega_i}{dA} = \langle X_i, N \rangle / Z_i^3$. h is the estimation of background intensity. E_{curv} and E_{prior} are regularization term in order to well-pose the problem. E_{prior} minimizes the area of the surface and favors smooth surfaces. E_{curv} minimizes the length of curve C and then favors smooth edge. E_{data} minimizes the error between image intensities and radiances on the 3D surfaces.

Following [1], we have introduced the auxiliary vector field V as a relaxed version of normal vector N [1]. V takes the place of the *unit* normal in modeling the shading effects and is supposed to resolve the instability that arises when using the normal direction in modeling shading. $E_{coupling}$ restricts the relaxation of normal direction by minimizing the integral of the distance between auxiliary vector and normal direction. Here we propose a closed-form solution for V in section 3.3 and we verify its impact on the energy minimization. Our minimization strategy is based on an iterative process minimizing successively the energy with respect to various variables.

3.1 Updating the Surface S , Segmenting Curve C and Albedo ρ_1, ρ_2

By fixing V, L, E_0, h , we first propose to update the shape S . It is quite straightforward, although lengthy, to derive the gradient flow that evolves S_t and C_t towards the solution S

and C :

$$\begin{aligned}
S_t &= \left(\sum_{i=1}^n \frac{\Gamma_i}{Z_i^3} \langle \chi_i X, R_i^T X_i \rangle - \chi_i \frac{\delta(\phi)}{Z_i^3} ((I_i - \rho_1(\langle V, L \rangle \xi + E_0))^2 \right. \\
&\quad \left. - (I_i - \rho_2(\langle V, L \rangle \xi + E_0))^2 \right) \langle \nabla_S \phi, R_i^T X_i \rangle + 2\alpha k + \\
&\quad \beta \frac{\delta(\phi)}{\|\nabla_S \phi\|} \Pi(\nabla_S \phi \times N) N, \quad (2)
\end{aligned}$$

where

$$\begin{aligned}
\Gamma_i &= H(\phi) (I_i - \rho_1(\langle V, L \rangle \xi + E_0))^2 \\
&+ (1 - H(\phi)) (I_i - \rho_2(\langle V, L \rangle \xi + E_0))^2 - (I_i - h)^2.
\end{aligned}$$

and

$$\begin{aligned}
C_t &= \sum_{i=1}^n ((I_i - \rho_2(\langle V, L \rangle \xi + E_0))^2 \\
&\quad - (I_i - \rho_1(\langle V, L \rangle \xi + E_0))^2) \sigma_i + \beta k_g \vec{n}. \quad (3)
\end{aligned}$$

In the above, S_t represents the gradient descent flow for the surface geometric prior. C_t is the gradient descent flow for the segmenting curve, which is related to the smoothness of the curve. \vec{n} is the normal direction of a point in the segmenting curve between regions of different albedo. k is the curvature. k_g is the geodesic curvature. χ is the visibility function of surface point X for the i th camera.

The motion of the curve is defined on the surface. To facilitate computing the gradient descent flow for curve C , we define the level set function $\phi : S \rightarrow \mathbb{R}$, which describes the location of C for the given surface S , such that $D_1 = \{X | \phi(X) > 0\}$, $D_2 = \{X | \phi(X) < 0\}$ and $C = \{X | \phi(X) = 0\}$. $H(\phi)$ is the Heaviside function of the level set function ϕ . We can now report the evolution of the level set function as

$$\begin{aligned}
\phi_t &= \|\nabla_S \phi\| \sum_{i=1}^n ((I_i - \rho_2(\langle V, L \rangle \xi + E_0))^2 \\
&\quad - (I_i - \rho_1(\langle V, L \rangle \xi + E_0))^2) \sigma_i + \beta \left(\delta_S \phi - \frac{\nabla_S^T \phi \nabla_S^2 \phi \nabla_S \phi}{\|\nabla_S \phi\|^2} \right). \\
\rho_1 &= \sum_{i=1}^n \int_{\pi_i(D_1)} (I_i(x_i)) d\Omega_i / \sum_{i=1}^n \int_{\pi_i(D_1)} (\langle V, L \rangle \xi + E_0) d\Omega_i; \\
\rho_2 &= \sum_{i=1}^n \int_{\pi_i(D_2)} (I_i(x_i)) d\Omega_i / \sum_{i=1}^n \int_{\pi_i(D_2)} (\langle V, L \rangle \xi + E_0) d\Omega_i;
\end{aligned} \quad (4)$$

The next step consists in fixing S, C, ρ_1, ρ_2 and in minimizing energy E with respect to V, L, E_0, h . Note here that the step is divided into two parts and positivity constraints are considered.

3.2 Updating of L , E_0 and h

In this step, we fix S , V , ρ_1 and ρ_2 and we minimize E with respect to L , E_0 and h . For the sake of minimizing E over h , h should satisfy

$$h = \left(\sum_{i=1}^n \int_{Q_i^c} I_i d\Omega_i \right) / \left(\sum_{i=1}^n \int_{Q_i^c} d\Omega_i \right). \quad (5)$$

Remind that Q_i^c represents the complement of the region $Q_i = \pi_i(S) \subset \Omega_i$. We constrain $E_0 \geq 0$ in the minimization with respect to (E_0, L) . The Lagrangian is thus as follows:

$$\begin{aligned} & \sum_{i=1}^n \left(\int_{\pi_i(D_1)} (I_i(x_i) - \rho_1(\langle V, L \rangle \xi + E_0))^2 d\Omega_i \right. \\ & + \sum_{i=1}^n \left(\int_{\pi_i(D_2)} (I_i(x_i) - \rho_2(\langle V, L \rangle \xi + E_0))^2 d\Omega_i \right. \\ & + \sum_{i=1}^n \int_{Q_i^c} (I_i(x_i) - h)^2 d\Omega_i + \alpha \int_S dA + \\ & \left. \left. \beta \int_S (1 - \langle V, N \rangle) dA + \gamma \int_C ds + \lambda E_0 = 0 \right) \right) \end{aligned}$$

constrained by $\lambda \geq 0$. The Kuhn-Tucker conditions are:

$$\begin{cases} 1) E_0 \geq 0, \\ 2) \lambda E_0 = 0, \\ 3) \sum_{i=1}^n \int_{\pi_i(D_1)} (I_i(x_i) - \rho_1(\langle V, L \rangle \xi + E_0)) \rho_1 V \xi d\Omega_i \\ \quad + \sum_{i=1}^n \int_{\pi_i(D_2)} (I_i(x_i) - \rho_2(\langle V, L \rangle \xi + E_0)) \rho_2 V \xi d\Omega_i = 0, \\ 4) - \sum_{i=1}^n \int_{\pi_i(D_1)} (I_i(x_i) - \rho_1(\langle V, L \rangle \xi + E_0)) \rho_1 d\Omega_i \\ \quad - \sum_{i=1}^n \int_{\pi_i(D_2)} (I_i(x_i) - \rho_2(\langle V, L \rangle \xi + E_0)) \rho_2 d\Omega_i + \lambda = 0, \end{cases} \quad (6)$$

λ is non-negative. We now focus on the definition field of the above problem. If the optimization is retrieved within the field of $\{E_0 \mid E_0 > 0\}$. Then $\lambda = 0$. Thus according to the third and fourth equations of (6) we have

$$\begin{cases} E_0 = \left(\sum_{i=1}^n \int_{\pi_i(D_1)} (I_i(x_i) - \rho_1 \langle V, L \rangle \xi) \rho_1 d\Omega_i \right. \\ \quad \left. + \sum_{i=1}^n \int_{\pi_i(D_2)} (I_i(x_i) - \rho_2 \langle V, L \rangle \xi) \rho_2 d\Omega_i \right) \\ \quad / \left(\sum_{i=1}^n \int_{\pi_i(D_1)} \rho_1^2 d\Omega_i + \sum_{i=1}^n \int_{\pi_i(D_2)} \rho_2^2 d\Omega_i \right), \\ L = \left(\sum_{i=1}^n \int_{\pi_i(D_1)} V V^T \rho_1^2 \xi d\Omega_i \right. \\ \quad \left. + \sum_{i=1}^n \int_{\pi_i(D_2)} V V^T \rho_2^2 \xi d\Omega_i \right)^{-1} \\ \quad \left(\sum_{i=1}^n \int_{\pi_i(D_1)} (I_i(x_i) - \rho_1 E_0) \rho_1 V \xi d\Omega_i \right. \\ \quad \left. + \sum_{i=1}^n \int_{\pi_i(D_2)} (I_i(x_i) - \rho_2 E_0) \rho_2 V \xi d\Omega_i \right). \end{cases} \quad (7)$$

In order to simplify the above equations, we let

$$\begin{cases} A = \sum_{i=1}^n \int_{\pi_i(D_1)} I_i(x_i) \rho_1 d\Omega_i + \sum_{i=1}^n \int_{\pi_i(D_2)} I_i(x_i) \rho_2 d\Omega_i, \\ V^* = \sum_{i=1}^n \int_{\pi_i(D_1)} \rho_1^2 V \xi d\Omega_i + \sum_{i=1}^n \int_{\pi_i(D_2)} \rho_2^2 V \xi d\Omega_i, \\ M = \sum_{i=1}^n \int_{\pi_i(D_1)} V V^T \rho_1^2 \xi d\Omega_i + \sum_{i=1}^n \int_{\pi_i(D_2)} V V^T \rho_2^2 \xi d\Omega_i, \\ p^* = \sum_{i=1}^n \int_{\pi_i(D_1)} I_i(x_i) \rho_1 V \xi d\Omega_i \\ \quad + \sum_{i=1}^n \int_{\pi_i(D_2)} I_i(x_i) \rho_2 V \xi d\Omega_i, \\ B = \sum_{i=1}^n \int_{\pi_i(D_1)} \rho_1^2 d\Omega_i + \sum_{i=1}^n \int_{\pi_i(D_2)} \rho_2^2 d\Omega_i. \end{cases}$$

By using the above abbreviations, which are all known, we simplify the equations as follows:

$$\begin{cases} E_0 = (A - \langle V^*, L \rangle) / B, \\ L = M^{-1}(p^* - E_0 V^*). \end{cases} \quad (9)$$

By combining the two equations above, we have

$$L = M^{-1}(p^* - (A - \langle V^*, L \rangle) V^* / B).$$

So

$$L = \langle V^*, L \rangle (M^{-1} V^*) / B + M^{-1}(p^* - A V^* / B). \quad (10)$$

In order to simplify this last equation, we introduce two new vector variables, \vec{a} and \vec{b} :

$$\begin{aligned} \vec{a} &= (a_1, a_2, a_3) = (M^{-1} V^*) / B, \\ \text{and } \vec{b} &= (b_1, b_2, b_3) = M^{-1}(p^* - A V^* / B). \end{aligned} \quad (11)$$

Thus we have

$$\begin{aligned} (L_1, L_2, L_3)^T &= (V_1^* L_1 + V_2^* L_2 + V_3^* L_3)(a_1, a_2, a_3)^T \\ &\quad + (b_1, b_2, b_3)^T. \end{aligned} \quad (12)$$

We let

$$W_{3*3} = \begin{bmatrix} (V_1^* a_1 - 1) & V_2^* a_1 & V_3^* a_1 \\ V_1^* a_2 & (V_2^* a_2 - 1) & V_3^* a_2 \\ V_1^* a_3 & V_2^* a_3 & (V_3^* a_3 - 1) \end{bmatrix}. \quad (13)$$

thus we have

$$W_{3*3} L = -\vec{b}. \quad (14)$$

It is a simple matter of linear algebra to retrieve L from the equation above, since all other terms are known. Now, from the representation of E_0 with respect to L , we can derive E_0 . If it is greater than zero, then it is what we are searching for and thus L is also fixed. Otherwise, we have to look for the minima within the field $\{E_0 \mid E_0 = 0\}$. This problem is significantly easier: Because $E_0 = 0$, we can derive L from the third equation of (6) as follows

$$L = \left(\sum_{i=1}^n \int_{Q_i} V V^T \xi d\Omega_i \right)^{-1} \left(\sum_{i=1}^n \int_{Q_i} I_i V \xi d\Omega_i \right). \quad (15)$$

3.3 Updating the Auxiliary Vector Field V

Having fixed S , C , ρ_1 , ρ_2 , L , E_0 and h we now minimize energy E with respect to V . Let us remind that we need to guarantee that $\|V\| = 1$. [1] proposed to update V with an iterated process. Unfortunately there is no strict proof that this strategy would guarantee the convergence of V .

We propose to give a closed-form representation for auxiliary vector field V . Without loss of generality, we let

$V = (p, q, \sqrt{1-p^2-q^2})$. Thus the minimization problem of E over vector field V is transformed into the minimization over the vector field (p, q) . The gradient of E over (p, q) is $\nabla E(p, q) = (\frac{dE}{dp}, \frac{dE}{dq})$. For each point X on shape S , the minimization condition is :

$$\begin{cases} \frac{dE}{dp} = 0, \\ \frac{dE}{dq} = 0. \end{cases} \quad (16)$$

Besides, according to the derivation over compound function, we have

$$\begin{cases} 0 = \frac{dE}{dp} = \frac{dE}{dV} \Big|_{V=(p,q,\sqrt{1-p^2-q^2})} \cdot \frac{dV}{dp} \\ = \frac{dE}{dV} \Big|_{V=(p,q,\sqrt{1-p^2-q^2})} \cdot (1, 0, -\frac{p}{\sqrt{1-p^2-q^2}})^T, \\ 0 = \frac{dE}{dq} = \frac{dE}{dV} \Big|_{V=(p,q,\sqrt{1-p^2-q^2})} \cdot \frac{dV}{dq} \\ = \frac{dE}{dV} \Big|_{V=(p,q,\sqrt{1-p^2-q^2})} \cdot (0, 1, -\frac{q}{\sqrt{1-p^2-q^2}})^T. \end{cases} \quad (17)$$

Accordingly, there exists a real scalar ν , such that

$$\frac{dE}{dV} \Big|_{V=(p,q,\sqrt{1-p^2-q^2})} = \nu \cdot (p, q, \sqrt{1-p^2-q^2})^T. \quad (18)$$

According to the definition of E , we have

$$\begin{aligned} \frac{dE}{dV} \Big|_{V=(p,q,\sqrt{1-p^2-q^2})} \\ = - \sum_{i=1}^n 2\chi_i (I_i(x_i) - \rho_1(\langle V, L \rangle \xi + E_0)) \rho_1 L \xi \sigma_i \\ - \sum_{i=1}^n 2\chi_i (I_i(x_i) - \rho_2(\langle V, L \rangle \xi + E_0)) \rho_2 L \xi \sigma_i - \beta N. \end{aligned} \quad (19)$$

By combining the above two equations, we have:

$$\begin{aligned} \nu (p, q, \sqrt{1-p^2-q^2}) \\ = - \sum_{i=1}^n 2\chi_i (I_i(x_i) - \rho_1(\langle V, L \rangle \xi + E_0)) \rho_1 L \xi \sigma_i \\ - \sum_{i=1}^n 2\chi_i (I_i(x_i) - \rho_2(\langle V, L \rangle \xi + E_0)) \rho_2 L \xi \sigma_i - \beta N. \end{aligned}$$

In order to simplify the above nonlinear system of equations we introduce an equivalent simpler form as follows:

$$\begin{aligned} \nu V &= \sum_{i=1}^n \chi_{i,D_1} (I_i(x_i) - \rho_1(\langle V, L \rangle \xi + E_0)) \rho_1 L \xi \sigma_i \\ &+ \sum_{i=1}^n \chi_{i,D_2} (I_i(x_i) - \rho_2(\langle V, L \rangle \xi + E_0)) \rho_2 L \xi \sigma_i + \beta N. \end{aligned}$$

If $\xi = 0$, obviously the solution is $V = N/\|N\|$; else we let

$$\begin{cases} G = (L\xi) (\sum_{i=1}^n 2\chi_i \sigma_i), \\ H = \beta N + L\xi \sum_{i=1}^n 2\chi_i (I_i - E_0) \sigma_i. \end{cases} \quad (20)$$

Thus we could get that

$$\begin{cases} \|(\nu \cdot I_{3*3} + G \cdot L^T)^{-1} H\|_2 = 1, \\ V = (\nu \cdot I_{3*3} + G \cdot L^T)^{-1} H. \end{cases} \quad (21)$$

In the above Equations, G , L^T , H are known, thus we could simply get ν via Newton-methods with the initial value ν determined by the previous V and other parameters in the above equation on ν . Actually as for the initialization of ν , since we have known the previous V^* , we could simply transfer it to the following linear minimization problem

$$\arg \min_{\nu} \left\| \sum_{i=1}^n \chi_i (2I_i - 2\langle V^*, kL\xi \rangle - 2E_0) kL\xi \sigma_i + \beta N - \nu V^* \right\|. \quad (22)$$

After calculating ν , we could get V from the second equation of (21), which minimizes E and fulfill the requirement of $\|V\| = 1$. One may argue that this is not a strictly closed-form solution because of the three parameters. Nevertheless, our experiments show that the above solution for ν takes two Newton steps to converge and costs less than ten percent of the time using a general iterative solution for auxiliary vector field.

3.4 Ambiguity in recovering albedo and illumination

When minimizing the energy function with respect to ρ_1, ρ_2, L, E_0 , we have the following ambiguity: Suppose $(\rho_1, \rho_2, E_0, L, C, S)$ is a solution, then $(\delta\rho_1, \delta\rho_2, E_0/\delta, L/\delta, C, S)$ ($\delta \in \mathbb{R}^+$) is another solution for our minimization problem, which is quite straightforward.

Although the ambiguity makes it impossible to extract the albedo and illumination exactly, we could still recover $\rho^* E_0; \|L\|/E_0; L/\|L\|$, i.e., reconstruct the parameters up to a scalar transformation, or recover ρ_1/ρ_2 . When we know the maximum albedo on the object, we would still expect to generate the exact albedos and the amplitude of ambient term and point light source as well. In practice, in each iterative step, we fix E_0 to a constant.

4 Experiments

In this section we report the experimental results of our algorithm tested on several data sets: doll dancer data, synthesized constant albedo sphere illuminated by a directional light, synthesized piecewise constant albedo sphere illuminated by ambient light and an additional directional light source, fish model data. Figure 1 shows four views of the doll dancer data. Figure 6 shows the synthesized constant albedo sphere. Figure 9 shows the third and fourth data sets.

For the doll dancer data, we took 28 calibrated images of a doll model 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 [1]. We ran with the algorithm using two directional light : a positive light and negative light [1] and show the result in Figure 2. 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.

Also for the sake of giving a more physically plausible image formation model, we add the positivity constraint on light and use the closed-form of auxiliary vector field and show the reconstructed shape in Figure 4. The improvement on shape details extraction are obtained especially at the hand and buttress area because the closed-form of auxiliary vector field recovers more accurate auxiliary vector, which is noticeable from Figure 5 in comparison to Figure 3. On the synthesized constant albedo sphere data shown in Figure 6, which is a constant albedo sphere illuminated by an ambient light and a point light source, we ran the algorithm with positivity constraint of light and show the reconstructed shape and the ground truth in Figure 7. The table in Figure 8 lists the reconstructed shape, light errors on synthesized data.

In further, we apply our algorithm to the piecewise constant albedo objects as Figure 9 shows. This data set includes a fish model painted with green band and white band and is illuminated by standard fluorescent lamps and a strong directional light. Noticing the ambiguity problem in our approach as proposed in the article, we rescale the terms ρ_1, ρ_2, E_0, L in each evolution step as follows : $\delta = E_0/100; E_0 = E_0/\delta; \|L\| = \|L\|/\delta; \rho_1 = \rho_1 * \delta; \rho_2 = \rho_2 * \delta;$ In the above rescaling, 100 is just an arbitrary number, we could also choose a number other than 100. Figure 10 shows the reconstructed shape and partitioning curve of different albedo for the synthesized painted ball illuminated by directional light. Figure 11 shows the reconstructed shape for the Fish model. We also show the evolution process of the shape and partitioning curve in Figure 12.

5 Conclusion

We have introduced an image-formation model for Lambertian scenes viewed from multiple calibrated viewpoints under a collection of point light sources. Unlike previously proposed models, we explicitly enforce positivity of the light sources, and therefore our model is physically plausible. Also we deal with piecewise constant albedo, rather than constant as previously done in the literature. We develop a novel algorithm for the estimation of shape, albedo and lighting position. This algorithm significantly improves the prior art on all grounds: accuracy, robustness and speed.



Figure 1: Example views of the input data set consisting 28 views of a dancer model on a table. The head is much brighter than the rest because the head is facing the spotlight.

More specifically, we develop a novel solution for the auxiliary vector field that represents the normal to the surface in closed form. This was the most computationally expensive step of existing algorithms. Furthermore, the positivity constraint that we enforce allows us to recover an illumination model that is physically plausible, and we show that it in turn allows for fairly accurate estimation of the lighting position. Of course, in laboratory environments the light is never ideal, as there are inter-reflections, diffuse illumination etc., but our model approximates the real illumination in the best possible way, as measured by the discrepancy between the real images and those generated by our model.

Naturally, the model we propose is limited by the assumptions of Lambertian reflection and point/diffuse illumination. The assumption of piecewise constant albedo is not very restrictive in theory, as many man-made objects can be well approximated by this model, but in practice when objects become complex the topology and geometry of the albedo boundaries become so difficult that our algorithm may fail to capture subtle variations. One of the potential advantages of our approach is its potential to be integrated with multi-view stereo algorithms, by providing additional constraints where objects with simple albedo do not provide enough constraints to establish point-to-point correspondence and therefore local feature-based approaches fail to provide a dense reconstruction.

Acknowledgment

This work is supported by ONR N00014-03-1-0850:P0001 and AFOSR E-16-V91-G2. Also the authors thank Dr. Hailin Jin for the great help on algorithm implementation.

References

- [1] H.Jin, D.Cremers, A.Yezzi, S.Soatto: Shedding light on stereoscopic segmentation. In: Proc. IEEE Conf. on Comp. Vision and Pattern Recogn. (2004) 36–42
- [2] Horn, B.: Relative orientation. Int. J. of Computer Vision **4** (1990) 59–78

Reconstruction Result	positive light algorithm	Ground Truth.
L	(0.000,0.001,0.999)	(0.000,0.000,1.000)
E_0	98.4	100.00
Intensity of point light	98.5	100.00
Shape Error	0.9 %	0(Note: Radius of ball is 10)

Figure 8: The table lists the results of the reconstruction results on synthesized data: A ball with radius 10 illuminated by a point light source $100 * (0, 0, 1)$ and an ambient light 100. We use the proposed algorithm here and add the positivity constraint of light and closed-form of auxiliary vector field

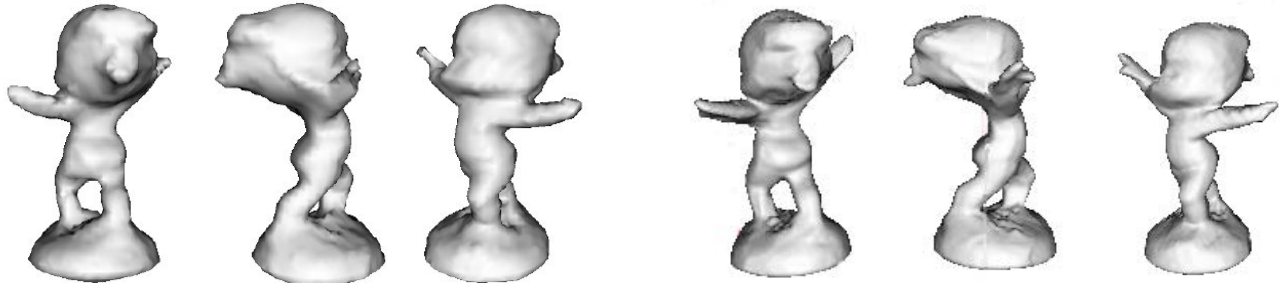


Figure 2: Final shape estimated using the proposed algorithm when we use a positive light and negative light. The algorithm reconstructs the 3D object much more accurately, because the simultaneously estimated light configuration allows to account for the shading effects in the input data.

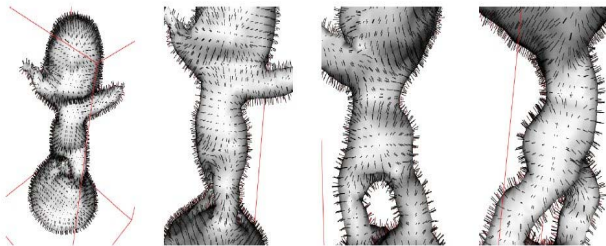


Figure 3: Visualization of the auxiliary vector field V via the proposed algorithm when we use two directional lights, a positive and a negative

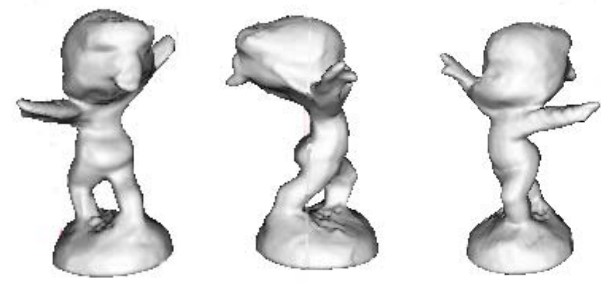


Figure 4: Final shape estimated via the proposed algorithm when we add the positivity constraint of light and closed-form of auxiliary vector field. The further improvement on 3D shape details, especially at hand, buttress and head, was obtained because the closed-form of auxiliary vector field is more accurate, which helps to estimate the details of shape.

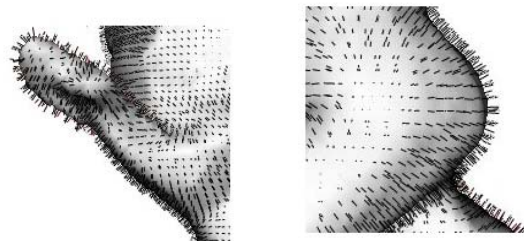


Figure 5: Visualization of auxiliary vector field via the proposed algorithm when we add the positivity constraint of light and closed-form of auxiliary vector field.

[3] Jin, H., Yezzi, A.J., Soatto, S.: Region-based segmentation on evolving surfaces with application to 3d shape and radiance estimation. In: Proc. of the Eur. Conf. on Comp. Vision. (2004) 114–125

[4] Zhang, R., Tsai, P., Cryer, J., Shah, M.: Analysis of shape from shading techniques. In: Proc. of the Intl. Conf. on Computer Vision and Pattern Recognition. (1994) 377–384

[5] Horn, B., (eds.), M.B.: Shape from Shading. MIT Press (1989)

[6] Koenderink, J., van Doorn, A.: Photometric invariants related to solid shape. Optica Acta 27(7) (1980) 981–996

[7] Nayar, S., Ikeuchi, K., Kanade, T.: Surface reflection: physical and geometrical perspectives. IEEE Trans. Pattern Anal. Mach. Intell. 13(7) (1991) 611–634

[8] Robles-Kelly, A., E.R.Hancock: Surface radiance correction for shape-from-shading. Pattern Recognition (2005) to appear

[9] Prados, E.: Application of the theory of the viscosity solutions to the Shape From Shading problem. PhD thesis, Univ. of Nice-Sophia Antipolis (2004)

[10] Yuille, A.L., Snow, D., Epstein, R., Belhumeur, P.: Deter-

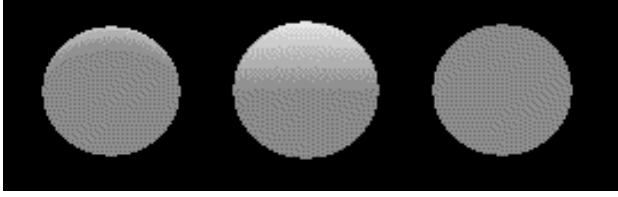


Figure 6: three views of the synthesized constant albedo sphere illuminated by a point light source $100 \times (0,0,1)$ and an ambient light of intensity 100.

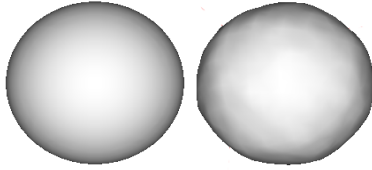


Figure 7: Final shape estimated by the algorithm adding positivity of light and closed-form of auxiliary vector field(right). The ground truth shape is shown(left).

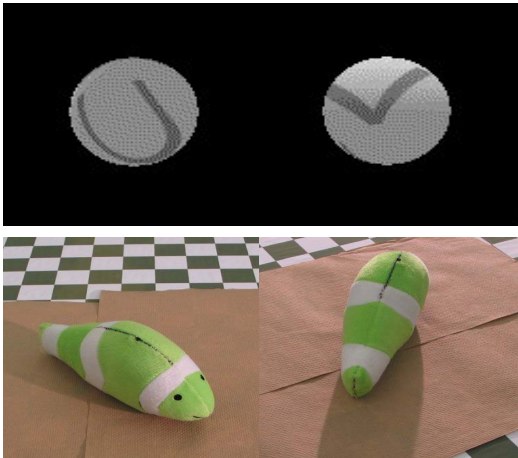


Figure 9: COLOR: The upper two images are 2 out of 26 views of a synthetic scene. The scene consists of one sphere painted with a word 'CV' with grey and the rest part is white. And the sphere is illuminated by ambient light of 100 and point light source of $100 \times (0, 0.717, 0.717)^T$; The lower two images are 2 out of 30 views of a real scene, which is a model of fish. Each image size is 390×400 and calibrated with a rig. The fish is green with white stripes. It is illuminated by a point light source from the head of the fish.

mining generative models of objects under varying illumination: shape and albedo from multiple images using svd and integrability. *Int. J. of Computer Vision* **35** (1999) 203–222

- [11] Chen, H.F., Belhumeur, P.N., Jacobs, D.W.: In search of illumination invariants. In: *Proc. IEEE Conf. on Comp. Vision and Pattern Recogn.* (2000)

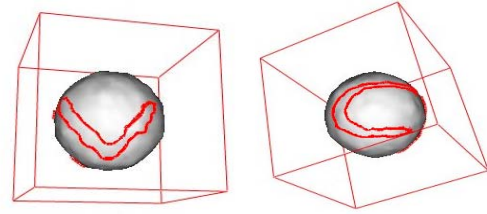


Figure 10: COLOR: The reconstructed shape and partitioning curve of different albedo for the synthesized data in Figure 9. The two images show the final shape result of the proposed algorithm together with the red segmenting curve. The red straight lines show the coordinate frame for the spheres.

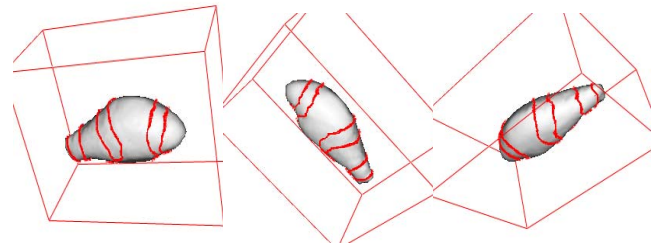


Figure 11: COLOR: recovered shape via the algorithm with adding positivity constraint of light, closed-form of auxiliary vector field and piecewise constant albedo, for the real scene shown in Figure 9. The red curves represent the segmenting curves for different albedo.

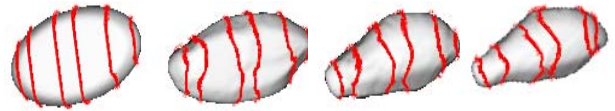


Figure 12: COLOR: The evolution process of shape and the segmenting curve for different albedo with the algorithm of adding the positivity constraint of light, closed-form auxiliary vector field and dealing with piecewise constant albedo. The initial surface is an ellipsoid.

- [12] D.Samaras: Illumination constraints in deformable models for shape and light direction estimation. *IEEE Trans. Pattern Anal. Mach. Intell.* (2003) 247–264
- [13] Samaras, D., Metaxas, D., Fua, P., Leclerc, Y.: Variable albedo surface reconstruction from stereo and shape from shading. In: *Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition.* (2000) I: 480–487
- [14] Oliensis, J., Dupuis, P.: A global algorithm for shape from shading. In: *Proc. of the Intl. Conf. on Computer Vision.* (1993) 692–710
- [15] Hartley, R., Zisserman, A.: *Multiple view geometry in computer vision.* Cambridge University Press (2000)