Stereoscopic Shading: Integrating Multi-frame Shape Cues in a Variational Framework*

Hailin Jin† Anthony Yezzi‡ Stefano Soatto†

† Washington University, St.Louis - MO 63130, {hljin,soatto}@essrl.wustl.edu
‡ Georgia Institute of Technology, Atlanta - GA 30332, Anthony.Yezzi@gatech.edu

Abstract

We address the problem of integrating multi-frame stereo and shading cues within the framework of optimization in the infinite-dimensional space of piecewise smooth surfaces. Cue integration then reduces to the determination of regions where prior assumptions on the reflectance of the surfaces can be enforced. By combining cues, our formulation allows defining a well-posed problem even when reconstruction from stereo or shading in isolation would be ill-posed. For a simplified model we prove the necessary conditions for optimality, and propose an iterative optimization algorithm, which we implement using ultra-narrowband level set methods.

1 Introduction

One of the central goals of Computer Vision is the reconstruction of the shape of the environment from images of its projection onto a two-dimensional sensor. To us, the environment is a collection of "objects", each bounded by a piecewise smooth surface radiating energy either directly or by reflection. A sensor provides a (noisy) reading of the amount of energy incident a discrete set of small regions on a two-dimensional surface (e.g. pixels on the CCD plane or receptors on the retina).

The sensory data depend upon the shape of the environment, but also upon its reflectance properties and the energy distribution. While, in general, neither one is known, the sensory data are not enough to allow a unique reconstruction of shape, reflectance and energy distribution. Consequently, research in Computer Vision has concentrated on recovering some properties *given* others (e.g. when shape is recovered from "shading" assuming given reflectance properties), or *independent* of others (e.g. when shape is recovered from "stereo" or "motion" regardless of reflectance properties and energy distribution).

While most of the research on cue integration is based upon merging the result of "Shape From X" modules each working in isolation, we propose an integrated formulation of the problem of reconstructing shape, where the integration of different cues amounts to determining where prior assumptions can and should be used. This results in a formulation of the correspondence problem which pertains to *regions*, rather than *points*. Due to the representation of the environment as a collection of surfaces, the resulting optimization problem is formulated in an infinite-dimensional space. We use the methods of calculus of variations, together with efficient numerical techniques to solve partial differential equations, to converge to a solution. However, the same problem could be formulated in a probabilistic (e.g. Bayesian) framework.

Our approach is geometric, and does not depend upon the parameterization chosen to represent surfaces in the scene. It can handle changes in the topology of the estimated surface due to the presence of multiple objects in the scene. Although thorough experimentation is under way we show some promising results on simulated scenes seen from several calibrated cameras.

Relation to previous work

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 of Horn and Brooks [9]. Other work on shading that is of generic relevance to this paper includes [11, 15, 10, 36, 6, 26, 2, 28, 27, 33, 25, 12, 17, 35, 21]. Using variational methods in shape from shading dates back to the eighties [8, 22], and even level set methods have been employed before [14, 13, 16]. The literature on stereo and motion is also extensive; we refer the reader to the book of Faugeras [4] for references. The most closely related work is that of Faugeras and Keriven [5], who cast the traditional multi-frame stereo in a variational framework and use level set methods to solve it. They address the correspondence problem by best approximating the brightness constancy assumption at every point in the image[1], thus obtaining in effect a dense correspon-

*Supported in part by NSF grant IIS-9876145 and ARO grant D AAD19-99-1-0139. We wish to thank A. Zisserman and P. Anandan for their comments.

[1] This is done by looking for corresponding patches that maximize a normalized cross-correlation criterion, the underlying assumption being that of brightness constancy of corresponding points.

dence wherever the brightness gradient is non-zero. Issues concerning the fusion of shading cues with parallax cues have been discussed in several works, including [1, 31, 18, 34, 20, 7, 3, 30].

2 A generative model

Let S be a surface in space of class at least C^1; we indicate the tangent plane to the surface at a point P by T and the inward unit normal vector by N. At each $P \in S$ we can construct a reference frame with origin at P, $\mathbf{e}_3$-axis parallel to the normal vector N and $\langle \mathbf{e}_1, \mathbf{e}_2 \rangle$-plane parallel to T (see figure 1). The change of coordinates between the reference point at P and an inertial reference frame (the "world" frame) is indicated by $g_S(P, N)$; g_S maps points in the reference frame at P into points in the inertial frame. For instance, it maps the origin into the point P, $g_S 0 = P$, and the $\mathbf{e}_3$-vector into N, $g_{S*}\mathbf{e}_3 = N$. We recall that, if we represent the change of coordinates g with a rotation matrix $R \in SO(3)$ and a translation vector $\mathbf{t}$, then the action of g on a point P of coordinates $\mathbf{X} \in \mathbb{R}^3$ is given by $gP \doteq R\mathbf{X} + \mathbf{t}$, while the action of g on a *vector* of coordinates V is given by $g_* V \doteq RV$. In what follows we will not make a distinction between a change of coordinates g and its representation, and we will also consider interchangeably points P and their representation $\mathbf{X} \in \mathbb{R}^3$.

Consider then a distribution of energy dE over a compact region of a surface in space L (the light source). The portion of energy coming from a direction λ_P that is reflected onto a direction $\mathbf{x}_P$ is described by $\beta(\mathbf{x}_P, \lambda_P)$, the *bidirectional reflectance distribution function* (BRDF). The energy that P reflects onto $\mathbf{x}_P$ is therefore obtained by integrating the BRDF against the energy distribution

$$\mathcal{E}(\mathbf{x}_P, P) \doteq \int_L \beta(\mathbf{x}_P, \lambda_P) dE(\lambda_P) \tag{1}$$

which depends upon the direction $\mathbf{x}_P$ and the point $P \in S$, as well as on the energy distribution E of the light source L.

The geometry of the sensor is described by a central projection π (see figure 1). For a point P with coordinates $\mathbf{X} \in \mathbb{R}^3$ expressed in the camera coordinate frame, having the $\mathbf{e}_3$-axis parallel to the optical axis and the $\langle \mathbf{e}_1, \mathbf{e}_2 \rangle$-plane parallel to the lens, the projection can be modeled as

$$\pi : \mathbb{R}^3 \to \Omega; \quad \mathbf{X} \mapsto \mathbf{x} = \pi(\mathbf{X}) \tag{2}$$

where $\Omega \subset \mathbb{RP}^2$ with $\pi(\mathbf{X}) \doteq \mathbf{X}/Z$ in the case of planar projection (e.g. onto the CCD), or $\Omega \subset \mathbf{S}^2$ with $\pi(\mathbf{X}) \doteq \mathbf{X}/\|\mathbf{X}\|$ in the case of a spherical projection (e.g. onto the retina). We will not make a distinction between the two models, and indicate the projection simply by π.

In order to express the direction $\mathbf{x}_P$ in the camera frame, we consider the change of coordinates from the reference frame at the point P. For simplicity, we let the inertial frame coincide with the camera frame, so that $\mathbf{X} \doteq g_S(P, N)0 = P$ and $\mathbf{x} \sim g_{S*}(N)\mathbf{x}_P$ where[2] we note that g_{S*} is a rotation, so that $\mathbf{x}$ depends on N, while $\mathbf{X}$ depends on P. Once we substitute $\mathbf{x}$ for $\mathbf{x}_P$ into $\mathcal{E}$ in (1) we obtain the *radiance*

$$R_1(P, N) \doteq \mathcal{E}(g_{S*}^{-1}(N)\mathbf{x}, P) \quad \text{where} \quad \mathbf{x} = \pi(P). \tag{3}$$

Our (ideal) sensor can measure the amount of energy received along the direction $\mathbf{x}$

$$I_1(\mathbf{x}) = R_1(P, N) \quad \text{where} \quad \mathbf{x} = \pi(P). \tag{4}$$

Consider a change in the viewpoint, described by a change of coordinates g relative to the inertial reference frame. Assuming that the inertial frame coincides with the image I_1, we can obtain a new image I_2 by moving with g (see figure 1). Using the fact that $\mathbf{x} \sim g_{S*}\mathbf{x}_P$ and $\mathbf{x}_2 \sim g_*\mathbf{x}$, we have that $\mathbf{x}_{P_2} \sim g_{S*} g_* g_{S*}^{-1} \mathbf{x}_P$. Similarly, the coordinates of the point P in the first and second camera frames are related by $P_2 = gP$, and $\mathbf{x} = \pi(P)$, $\mathbf{x}_2 = \pi(P_2) = \pi(gP)$. Therefore, the (scene) radiance in the direction of the new viewpoint is given by $R_2(P, N, g; \beta, L, E) \doteq \mathcal{E}(g_{S*}^{-1}(N) g_* \mathbf{x}, gP)$ and the (image) *irradiance* is $I_2(\mathbf{x}_2) = R_2(P, N, g; \beta, L, E)$ where $\mathbf{x}_2 \doteq \pi(gP)$. So far, we have used the first image as a reference, so that $\mathbf{x}_1 = \mathbf{x}$, g_1 is the identity transformation and $P_1 = P$. This needs

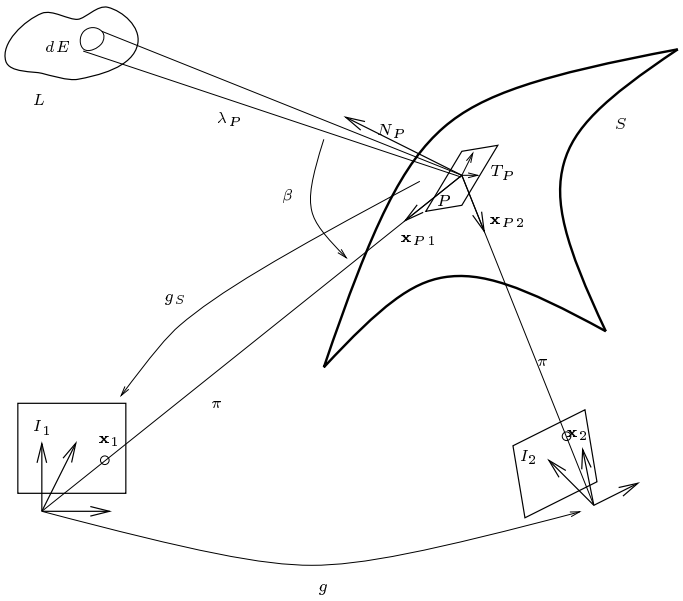

Figure 1: **Generative model**

[2] The symbol $\sim$ indicates projective equivalence, that is equality up to a scalar. Strictly speaking, $\mathbf{x}$ and $\mathbf{x}_P$ do not represent the same vector, but only the same direction (they have opposite sign and different lengths). However, they do represent the same point in the projective plane, and therefore we will regard them as one and the same. In order to obtain the same embedded representation (i.e. a vector in $\mathbb{R}^3$ with the same coordinates), we would have to write $\mathbf{x} = \pi(-g_{S*}(N)\mathbf{x}_P)$. The same holds if we model the projective plane using the sphere with antipodal points identified.

not be the case. If we choose an arbitrary inertial reference and indicate with g_k the change of coordinates to the reference of camera k, P_k the coordinate of the point P in this frame and $\mathbf{x}_k$ the corresponding direction, then for each of $k = 1 \dots M$ images we can write $I_k(\mathbf{x}_k) = R_k(P, N, g_k; \beta, L, E)$.

The irradiance I_k of image k can be measured only up to noise. Since we assume that energy transport phenomena are additive, such a noise will be additive, but will have to satisfy the constraint that both radiance and irradiance must be *positive*. This constraint is not satisfied if, for instance, we choose to model noise as a Gaussian process. We will defer the discussion on the choice of the noise model until next section. In the meantime, we write the generative model as

$$\begin{cases} I_k(\mathbf{x}_k) = R_k(P, N, g_k; \beta, L, E) + n_k(\mathbf{x}_k) \\ \text{subject to} \quad \mathbf{x}_k = \pi(g_k P) \quad \text{and} \quad R_k \geq 0 \end{cases} \tag{5}$$

for $k = 1 \dots M$.

3 Formalization of the problem in a variational framework

In order to reconstruct shape, we want to "invert" the generative model (5) and infer P and N from measurements of I_k, $k = 1 \dots M$. However, such an inversion is not possible since we only have noisy measurements of I_k available. Therefore, we convert the inversion task into an optimization task where we wish to find the "cause" (the shape of S) that explains the "effects" (the data I_k) "best". What is "best" in this context depends on how we model the noise n_k.

In a probabilistic framework we would assume a statistical model for n_k and then optimize a criterion (for instance the total likelihood) with respect to the unknown shape of S, possibly weighted by a prior density on S. In a variational framework we assume that the data I_k live in a function space, and then optimize a criterion (for instance the L_2 norm, the total variation, the information-divergence, or normalized cross-correlation) with respect to the unknown shape S. Although conceptually these are very different approaches, in practice they all reduce to an optimization problem on an infinite-dimensional function space. For the purpose of this study ,we choose a variational approach and select a norm $\|\cdot\| : C^1(\Omega) \to \mathbb{R};\ n \mapsto \|n\|$ for instance $\|n\| = \int_\Omega |n(\mathbf{x})| d\mu(x)$, where μ is an appropriate metric defined on the surface S. Once a metric is chosen we look for the shape of the surface S, represented by the function P and its derivative N, that minimizes $\sum_{k=1}^{M} \|n_k\|$ subject to (5) for all $\mathbf{x}_k \in \Omega$ and all $k = 1 \dots M$.

The principled solution to the problem of inferring shape from images obtained with the model (5) would be to set up the optimization problem and solve simultaneously for all unknown quantities, that is S, β, L, E.

$$\hat{S}, \hat{\beta}, \hat{L}, \hat{E} = \arg \min_{P,\beta,L,E} \sum_{k=1}^{M} \|n_k\| \quad \text{subject to (5)}. \tag{6}$$

Unfortunately, a simple counting of the orders of infinity involved suggests that, from a countable number of images M, we can recover at most a countable number of parameters[3] in β, L and E.

Therefore, our approach will be that of imposing a finite parameterization a of β, L and E, each corresponding to a *case study*, and then simultaneously recovering the shape S and the unknown parameters a. We therefore solve for $\hat{S}, \hat{a}$ that minimize $\sum_{k=1}^{M} \|n_k\|$ subject to (5) for all $\mathbf{x}_k \in \Omega$ and all $k = 1 \dots M$. Once we substitute the expression of n_k and $\mathbf{x}_k$, omitting the positivity constraints, we obtain

$$\min_{P,a} \sum_{k=1}^{M} \|I_k(\pi(g_k P)) - \int_L \beta(g_{S*}^{-1}(N)\pi(g_k P), \lambda_P; a) dE(\lambda_P; a)\| \tag{7}$$

The problem can be further generalized (and rendered significantly more difficult) by including as unknown the pose of the camera in each view, g_k, $k = 1 \dots M$.

3.1 A bare-bone model of reflection

In this section we describe a simple parametric model for the reflection function $\beta(\mathbf{x}_P, \lambda_P)$. This is an extension of what is commonly used in Computer Graphics (see for instance [32]). To this end, let the incident angle the light makes with the surface normal be represented by its cosine: $\gamma_P \doteq \langle \lambda_P, N \rangle$ and the reflection angle towards the k-the camera by $\xi_{P_k} \doteq \langle \mathbf{x}_{P_k}, N \rangle$. Furthermore, call H_k the "halfway" vector between the incident and viewing directions $H \doteq \frac{\mathbf{x}_{P_k} + \lambda_P}{2}$ and let δ represent the (cosine of the) angle H makes with the surface normal $\delta_k \doteq \langle H_k, N \rangle$. Finally, let θ_x and θ_y be the cosine of the angles between the coordinate axes $\mathbf{e}_1, \mathbf{e}_2$ in the reference frame at the point P and the corresponding axes in the inertial reference frame $\theta_x \doteq \langle \mathbf{e}_1, g_{S*}\mathbf{e}_1 \rangle$, $\theta_y \doteq \langle \mathbf{e}_2, g_{S*}\mathbf{e}_2 \rangle$. We write the BRDF as the sum of a diffuse (Lambertian) component whose intensity $\rho_r(P)$ depends upon the point P on the surface and a specular component whose intensity depends upon the roughness of the surface along the two coordinate directions on the tangent plane, represented by two coefficients α_x, α_y. The overall strength of the specular component is weighted by a coefficient ρ_s which

[3] That is if our primary interest is the shape S. If we are interested in the illumination L, then we can recover at most L and a countable number of parameters in S, β, E, and similarly for other unknowns.

we will assume to be constant. We then have

$$\beta = \rho_r(P)\gamma_P + \rho_s \frac{e^{-\frac{1-\delta_k^2}{\delta_k^2}\left(\frac{\theta_x^2}{\alpha_x^2}+\frac{\theta_y^2}{\alpha_y^2}\right)}}{4\pi\alpha_x\alpha_y\sqrt{\gamma_P\xi_{P_k}}}. \quad (8)$$

In the experiments described in section 6 we restrict our attention to the simplest case of a source of light located at a point in space with coordinates L_0 relative to the inertial reference frame. The coordinates in the reference frame at point P are indicated by L_P. Therefore, we have that $L \in \mathbb{R}^3$ and $dE(\lambda_P) = E_0\delta(\lambda_P - L_P)$. If we call a the reflectance parameters $a \doteq \begin{bmatrix} \rho_r & \rho_s & \alpha_x & \alpha_y \end{bmatrix}^T$ then we have that $R(P, N, g_k; \beta, L, E)$ depends on a, L_0, E_0 as follows

$$R(P, N, g_k; a, L_0, E_0) \doteq \beta(\mathbf{x}_{P_k}, L_P; a)E_0 \quad (9)$$

which, for isotropic Lambertian materials, is equal to $\rho_r(P)\langle L_P, N\rangle E_0$, and the residual to be minimized is given by[4]

$$\sum_{k=1}^{M}\int_\Omega |I_k(\pi(g_kP(\mathbf{x}))) - \rho_r(P(\mathbf{x}))\langle\frac{P-L_0}{\|P-L_0\|}, N\rangle E_0|d\mu_P(\mathbf{x}) \quad (10)$$

where the area form is $d\mu_P(\mathbf{x}) \doteq |P_{\mathbf{e}_1}(\mathbf{x}) \times P_{\mathbf{e}_2}(\mathbf{x})|d\mathbf{x}$.

4 Stereoscopic shading

Even in the simplest case of a single direct light source and a Lambertian surface, counting the orders of infinity involved in equation (10) will convince the reader that it is not possible to reconstruct a unique shape S and reflectance function ρ_r from images alone, without making prior assumptions.

Since the second term of the norm in (10) *does not depend on* k one could substitute it with $I_j(\pi(g_kP(\mathbf{x})))$ for any $j = 1 \dots M$ and obtain[5]

$$\int_\Omega \sum_{j,k=1}^{M} |I_k(\pi(g_kP(\mathbf{x}))) - I_j(\pi(g_jP(\mathbf{x})))|d\mu_P(\mathbf{x}) \doteq$$
$$\doteq \int_\Omega \phi_{SFM}(\mathbf{x}, P)d\mathbf{x}. \quad (11)$$

However, we know that when the gradient of I_k is small (zero, in the limit), and noise is present, the localization of P becomes ill-posed (I_k does not depend on P in the limit).

[4] Notice that $L_P = g_S^{-1}L_0$ and $N = \mathbf{e}_3$ in the reference frame at P. In the inertial frame we have that $N = g_{S*}\mathbf{e}_3$ and $L_P = g_{S*}g_S^{-1}L_0 = L_0 - P$. The expression in (10) follows by reducing L_P to be unit-norm.

[5] As an alternative to $|I_k - I_j|$ one can consider other discrepancy measures, $\phi(I_k, I_j)$, such as the information-divergence or normalized cross-correlation. This does not change the substance of our argument.

If, on the other hand, we have some prior information on the reflectance function ρ_r, for instance that it is locally constant, the expression (10) could be minimized with respect to S and ρ_r, therefore leading to a multi-frame version of shape from shading

$$\int_\Omega \sum_{k=1}^{M} |I_k(\pi(g_kP(\mathbf{x}))) - \rho_r\langle\frac{P-L_0}{\|P-L_0\|}, N\rangle E_0|d\mu_P(\mathbf{x}) \doteq$$
$$\doteq \int_\Omega \phi_{SFS}(\mathbf{x}, P, \rho_r)d\mathbf{x}. \quad (12)$$

However, if ρ_r were assumed to be constant while, in fact, it is not, this procedure would lead to gross errors. Therefore, the crucial problem remains of determining the subset $\mathcal{E}$ of Ω where to enforce a prior assumption on ρ_r.

4.1 The correspondence problem revisited

Substituting $I_j(\pi(g_kP(\mathbf{x})))$ for $\rho_r(P(\mathbf{x}))\langle\frac{P-L_0}{\|P-L_0\|}, N\rangle E_0$ in (10) is always legitimate, although it may lead to an ill-posed problem (11) in regions of Ω of uniform intensity. On the other hand, enforcing $\rho_r(P) = \rho_r$ leads to a well-posed problem (12) provided that $\mathbf{x}$ is in a region where $\rho_r(P(\mathbf{x}))$ is constant. Therefore, we are left with the problem of finding the region $\mathcal{E} \subset \Omega$ where to enforce the prior on shading. Once this is done, merging shading and stereo will be achieved by finding the shape S the minimizes the following cost functional

$$\int_{\mathcal{E}} \phi_{SFS}(\mathbf{x}, P, \rho_r)d\mathbf{x} + \int_{\mathcal{E}^c} \phi_{SFM}(\mathbf{x}, P)d\mathbf{x} \quad (13)$$

where $\mathcal{E}^c$ indicates the complement in Ω and ϕ_{SFM} and ϕ_{SFS} are defined in equations (11) and (12).

Indeed, it is not necessary that $\mathcal{E}$ captures all portions of the scene with constant reflectance functions, as long as it is its subset. For the minimization of (13) to be viable, $\mathcal{E}$ must be such that

$$\{\mathbf{x} \in \Omega \,|\, \nabla I_k(\pi(g_kP(\mathbf{x}))) = 0\} \subset \mathcal{E} \subset \{\mathbf{x} \in \Omega \,|\, \nabla\rho_r(P(\mathbf{x})) = 0\}. \quad (14)$$

We call $\mathcal{E}^c \doteq \Omega\backslash\mathcal{E}$ the *correspondence set*. It leads to a formulation of the correspondence problem in terms of *regions*, rather than *feature points*. Furthermore, the definition of a correspondence set is flexible, as long as it satisfies the constraints (14), and can depend upon the noise level in the image as well as upon prior information on the scene. For instance, if we choose an image as the local reference frame, so that $\mathbf{x} = \mathbf{x}_k = \pi(g_kP)$, call $\mathcal{A}$ the set of points where the gradient of I_k is not identically zero

$$\mathcal{A}_\epsilon \doteq \{\mathbf{x} \in \Omega \mid \|\nabla I_k(\mathbf{x})\| \geq \epsilon\} \quad (15)$$

for some k and ϵ, then we see that $\mathcal{E} \doteq \mathcal{A}_\epsilon^c$ leads to a feasible definition of correspondence set. In practice, we

have noticed that the functional for stereo can be applied in almost every region of the image, as long as there is a gradient of irradiance. The functional for shading is mostly used to "fill in" regions or small irradiance gradient, as we show in section 6.

4.2 Optimality conditions

Although in the previous section we have indicated a possible choice for $\mathcal{E}$, the focus of this paper is not on finding the optimal correspondence set. Rather, we will concentrate on solving the minimization of the cost functional (13), assuming that $\mathcal{E}$ has been chosen.

Of the two components of the cost, the second has been discussed extensively by Faugeras and Keriven [5]. We will therefore concentrate on the first one, and derive the optimality conditions (relative to $P(\cdot)$ and ρ_r) for

$$\int_{\mathcal{E}} \sum_{k=1}^{M} |I_k(\pi(g_k P(\mathbf{x}))) - \rho_r \langle \frac{P - L_0}{\|P - L_0\|}, N \rangle E_0 | d\mu_P(\mathbf{x}) \tag{16}$$

Notice that ρ_r and E_0 always appear as a product, which we denote by α. We will devise an iterative scheme to estimate both α and the surface, rather than solving α explicitly and plugging it into the last equation. One straightforward method is to first fix α, run a level set method solving for reconstructing the surface, and then fix the surface, and use any one-dimensional search scheme to solve for α. The necessary conditions pertaining to α are straighforward. In order to derive the necessary conditions with respect to P, we write the above equation, for simplicity in the case of a two-dimensional scene, as

$$E(t) = \int_{\mathcal{E}} \Phi(P, N) \|P_{\mathbf{x}}\| d\mathbf{x} \tag{17}$$

where $P(\mathbf{x}, t) \in \mathbb{R}^2, \mathbf{x} \in \mathbb{R}$ and $d\mu_P(\mathbf{x}) = \|P_{\mathbf{x}}\| d\mathbf{x}$. We now show that the necessary conditions for P impose that

$$\kappa\Phi - \Phi_X \cdot N - \kappa(\Phi_N \cdot N) + \kappa T^T \Phi_{NN} T - T^T \Phi_{NX} T = 0 \tag{18}$$

where the subscript indicates partial derivatives and κ is the curvature. This formula is obtained in the following subsection. There, $P(\mathbf{x}, t) \in S$ denotes a point on the evolving surface, while $\mathbf{X} \in \mathbb{R}^2$ denotes a point in the ambient (embedding) space.

A General Gradient Flow Equation

Assuming that we want to minimize the cost functional (17) over a family of curves, $P(\mathbf{x}, t)$ where $\mathbf{x} \in [0, 1]$ parameterizes each curve and t parameterizes the family, it is natural to consider the derivative of E. In the following derivation the unit tangent T and inward normal N of P will be related by $N = JT$ where J denotes the ninety degree rotation matrix. In addition, we will always omit purely tangential terms whenever they appear inside an inner product with the curve variation P_t.

$$\begin{aligned}
E'(t) &= \int_0^1 (\Phi' \|P_{\mathbf{x}}\| + \Phi \|P_{\mathbf{x}}\|') d\mathbf{x} \\
&= \int_0^1 \left(\Phi' \|P_{\mathbf{x}}\| + \Phi \frac{\langle P_{\mathbf{x}}, P_{\mathbf{x}t} \rangle}{\|P_{\mathbf{x}}\|} \right) d\mathbf{x} \\
&= \int_0^1 ((\Phi_X \cdot P_t + \Phi_N \cdot N_t) \|P_{\mathbf{x}}\| - \langle (\Phi T)_{\mathbf{x}}, P_t \rangle) \, d\mathbf{x} \\
&= \int_0^1 ((\Phi_N \cdot N_t) \|P_{\mathbf{x}}\| - \langle \kappa\Phi N - \Phi_X, P_t \rangle \|P_{\mathbf{x}}\|) \, d\mathbf{x}
\end{aligned}$$

We now use the following expression to substitute for N_t.

$$N_t = \left(\frac{JP_{\mathbf{x}}}{\|P_{\mathbf{x}}\|} \right)_t = \frac{JP_{\mathbf{x}t}}{\|P_{\mathbf{x}}\|} - \frac{JP_{\mathbf{x}}}{\|P_{\mathbf{x}}\|^2} \frac{\langle P_{\mathbf{x}}, P_{\mathbf{x}t} \rangle}{\|P_{\mathbf{x}}\|}$$

$$\begin{aligned}
E'(t) = &\int_0^1 (\langle \Phi_N, JP_{\mathbf{x}t} \rangle - \langle (\Phi_N \cdot N)T, P_{\mathbf{x}t} \rangle - \\
&- \langle \kappa\Phi N - \Phi_X, P_t \rangle \|P_{\mathbf{x}}\|) d\mathbf{x} \\
= &\int_0^1 (-\langle \Phi_{NX} P_{\mathbf{x}} + \Phi_{NN} N_{\mathbf{x}}, JP_t \rangle + \langle (\Phi_N \cdot N) T_{\mathbf{x}}, P_t \rangle - \\
&- \langle \kappa\Phi N - \Phi_X, P_t \rangle \|P_{\mathbf{x}}\|) d\mathbf{x} \\
= &- \int_0^1 (\langle J^T \Phi_{NX} T - \kappa J^T \Phi_{NN} T - \kappa(\Phi_N \cdot N) N, P_t \rangle \|P_{\mathbf{x}}\| + \\
&+ \langle \kappa\Phi N - \Phi_X, P_t \rangle \|P_{\mathbf{x}}\|) d\mathbf{x} \\
= &- \int_0^L \langle J^T \Phi_{NX} T - \kappa J^T \Phi_{NN} T - \kappa(\Phi_N \cdot N) N + \\
&+ \kappa\Phi N - \Phi_X, P_t \rangle ds
\end{aligned}$$

In the last line, the upper limit of integration L denotes the total arclength of the curve P (at time t), and ds denotes the incremental arclength. From here it is clear to see that the gradient descent evolution is obtained by setting

$$P_t = J^T \Phi_{NX} T - \kappa J^T \Phi_{NN} T - \kappa(\Phi_N \cdot N) N + \kappa\Phi N - \Phi_X$$

Ignoring tangential components in this evolution leads to the following geometrically equivalent evolution equation.

$$P_t = (-T^T \Phi_{NX} T + \kappa T^T \Phi_{NN} T - \kappa(\Phi_N \cdot N) + \kappa\Phi - \Phi_X \cdot N) N \tag{19}$$

A More Specialized Gradient Flow Equation

In (16) we are considering a special case of (17) in which $\Phi(\mathbf{X}, N)$ is obtained by applying an L_1 norm to the error, $\epsilon(\mathbf{X}, N)$, between the measured image data $f(\mathbf{X})$ and the modeled image data, $G(\mathbf{X}) \cdot N$. The gradient flow equations will therefore depend upon:

$$\begin{aligned}
\epsilon(\mathbf{X}, N) &= f(\mathbf{X}) - G(\mathbf{X}) \cdot N \\
\epsilon_X(\mathbf{X}, N) &= f_X(\mathbf{X}) - G_X^T(\mathbf{X}) N \\
\epsilon_N(\mathbf{X}) &= -G(\mathbf{X})
\end{aligned}$$

If we use an L_1 norm, we obtain the following energy functional

$$E(t) = \int_0^1 |f(P) - G(P) \cdot N| \|P_{\mathbf{x}}\| d\mathbf{x}, \qquad (20)$$

in which case:

$$\begin{array}{lcl} \Phi & = & |\epsilon| = \pm(f - G \cdot N) \\ \Phi_X & = & \pm\epsilon_X = \pm(f_X - G_X^T N) \\ \Phi_N & = & \pm\epsilon_N = \mp G \\ \Phi_{NX} & = & \mp G_X \\ \Phi_{NN} & = & 0. \end{array}$$

Plugging these quantities into the general flow (19) yields

$$\begin{array}{rcl} P_t & = & (\pm T^T G_X T \pm \kappa G \cdot N \pm \\ & & \quad \pm \kappa(f - G \cdot N) \mp (f_X - G_X^T N) \cdot N)N \\ & = & (\pm N^T J^T G_X J N \pm \kappa f \mp (f_X - G_X^T N) \cdot N)N \\ & = & (\pm N^T (J^T G_X J + G_X^T) N \pm \kappa f \mp f_X \cdot N)N. \end{array}$$

By noting that $J^T G_X J + G_X^T = trace(G_X) I$ where I is the 2x2 identity matrix, we may simplify the flow above to obtain

$$P_t = \pm(\kappa f + trace(G_X) - f_X \cdot N)N.$$

A geometrically equivalent flow

$$P_t = \pm(\kappa f + trace(G_X))N \mp f_X \qquad (21)$$

reveals at a glance that although our cost functional (20) depends upon the unit normal N of the interface, the resulting gradient flow, surprisingly, does not (up to sign). Instead it consists of a weighted curvature term, κf, an inflationary term, $trace(G_X)$, and a gradient term, f_X, all of which depend only upon $\mathbf{X}$. Therefore, the terms needed to implement the curve evolution (21) may be pre-computed, which both simplifies the implementation of the algorithm and leads to significant savings in computation.

We discover another surprising property of this flow by substituting $G(\mathbf{X})$ given in (16) by $\alpha \frac{\mathbf{X} - L}{\|L - \mathbf{X}\|}$ where $\alpha = \rho_r E_0$ and L is the position of light source. Letting $\mathbf{X} = (x_1, x_2)$, $G(\mathbf{X}) = (g^1(x_1, x_2), g^2(x_1, x_2))$, $L = (l_1, l_2)$ and $d_L(\mathbf{X}) = \|L - \mathbf{X}\|$, we may compute

$$\begin{array}{l} trace(G_X) = g^1_{x_1} + g^2_{x_2} \\ = \alpha \left(\dfrac{1}{d_L} - \dfrac{(l_1 - x_1)^2}{d_L^3} \right) + \alpha \left(\dfrac{1}{d_L} - \dfrac{(l_2 - x_2)^2}{d_L^3} \right) \\ = \alpha \dfrac{(l_2 - x_2)^2 + (l_1 - x_1)^2}{d_L^3} = \dfrac{\alpha}{d_L} \end{array}$$

Substituting this into (21) yields

$$P_t = \pm(\kappa f + \frac{\alpha}{d_L})N \mp f_X. \qquad (22)$$

We therefore see that while our energy functional (16) depends only upon the *direction* of the light source, the gradient flow equation depends (up to sign) only upon the *distance* of the light source.

The same flow may be derived in the three dimensional setting where P_t denotes the evolution of a surface as opposed to a curve, where L, N, and f_X now denote 3D vectors, and where κ denotes the mean curvature of the evoling surface.

Remark: Note that that first term $\pm\kappa f$ in (22) may be regarded as a geometric diffusion term with a diffusion coefficient of $\pm f$. Since, f is non-negative (recall that f is just the intensity of one of the camera images at a back-projected point) this implies that the diffusion is well-posed when the coefficient is $+f$ but ill-posed when the coefficient is $-f$. In the latter case, this second order term in flow (22) gives rise to a backwards, nonlinear heat flow which is highly unstable. For this reason we refrain from implementing (22) directly, but use the following well-posed flow instead.

$$P_t = (\kappa f \pm \frac{\alpha}{d_L})N \mp f_X. \qquad (23)$$

In the two-dimensional case, flows (22) and (23) will exhibit the same steady state contour if the steady state contour is convex. This also applies in the three-dimensional case, but extends to certain non-convex surfaces as well. Namely, any steady state surface with non-negative mean curvature everywhere will be captured by both (22) and (23). This allows us to capture certain types of concavities (e.g. parabolic concavities) but not all types (e.g. elliptic concavities).

5 Level Set Implementation

In this section we outline the level set implementation of flow (22). The level set implementation of any geometric flow begins by embedding the initial interface $P(\mathbf{x}, 0)$, whether it be a curve or a surface, as a level set of a scalar function $\psi_0(\mathbf{X})$ which is then taken to be the initial condition for a function over time $\psi(\mathbf{X}, t)$

$$\psi_0 : \mathbb{R}^n \to \mathbb{R}, \quad \psi : \mathbb{R}^n \times \mathbb{R}^+ \to \mathbb{R}, \quad \psi(\mathbf{X}, 0) = \psi_0(\mathbf{X})$$

where $n = 2$ for curves and $n = 3$ for surfaces. The choice of a particular level set is arbitrary but is typically taken to be zero. The key point is that we continue to embed the interface within the same fixed level set of ψ for all time. Thus, if we choose the zero level set we have

$$\psi_0(P(\mathbf{x}, 0)) = 0, \quad \text{and} \quad \psi(P(\mathbf{x}, t), t) = 0.$$

Differentiating with respect to t therefore yields

$$\psi_t + \nabla\psi \cdot P_t = 0 \quad \text{or} \quad \psi_t = -P_t \cdot \nabla\psi \qquad (24)$$

an evolution equation for ψ (where $\nabla\psi$ denotes ψ_X) which evolves the interface $P(\mathbf{x}, t)$ described implicitly by $\psi(\mathbf{X}, t) =$

0 for all t. Substituting (22) into (24) yields

$$\begin{aligned}\psi_t &= -[\pm(\kappa f + \frac{\alpha}{d_L})N \mp \nabla f] \cdot \nabla\psi \\ &= \pm(\kappa f + \frac{\alpha}{d_L})(-N) \cdot \nabla\psi \pm \nabla f \cdot \nabla\psi.\end{aligned}$$

Finally, by substituting $\frac{\nabla\psi}{\|\psi\|}$ for the outward normal $-N$ and $\nabla \cdot \left(\frac{\nabla\psi}{\|\psi\|}\right)$ for the curvature κ, we obtain the following level set implementation of flow (22).

$$\psi_t = \pm\left(\nabla \cdot \left(\frac{\nabla\psi}{\|\psi\|}\right) f + \frac{\alpha}{d_L}\right)\|\nabla\psi\| \pm \nabla f \cdot \nabla\psi \qquad (25)$$

The level set flow exhibits this same form in any dimension.

6 Experiments

In this section we restrict our attention to simulated three-dimensional scenes populated with Lambertian objects. Experimentation with more complex reflectance models of real scenes is part of our research agenda. In figure 2

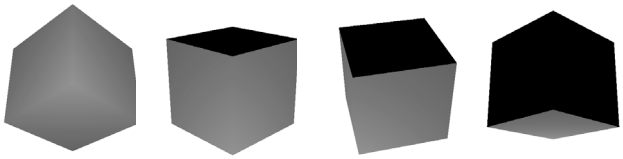

Figure 2: *Two camera views of a uniform Lambertian cube illuminated by a point light source. Some of the faces appear black since they are not reached by the light.*

we show a few views of a uniform Lambertian cube illuminated by a point light source, while in figure 3 we show the rendered reconstructed shape obtained by minimizing the second term in (13). However, if the scene contains

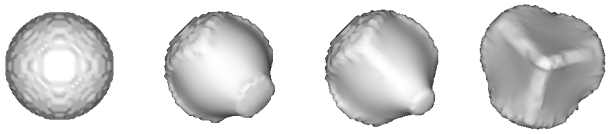

Figure 3: *Rendered shape at different stages of the reconstruction by the multi-frame shading algorithm.*

textured objects, such as the cube shown in figure 5, then the prior enforced by minimizing the second term in (13) is not valid, and therefore the reconstruction is incorrect, as we show in figure 4.

On the other hand, in figure 5 we show a number of views of a cube with smoothly textured faces. Despite the absence of distinct features, our multi-frame stereo algorithm - which is similar to that of Faugeras and Keriven

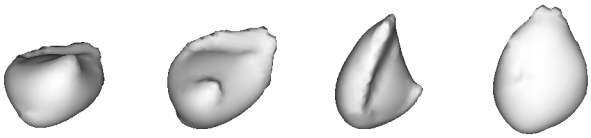

Figure 4: *Rendered shape as reconstructed by the multi-frame shading algorithm when the prior on shading is not satisfied (scene in figure 5). Of course, the reconstruction is incorrect.*

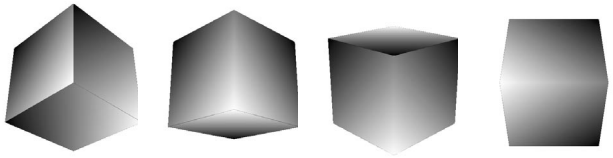

Figure 5: *Camera views of a cube with smoothly textured surfaces.*

[5] - is capable of capturing the shape of the cube (figure 6). However, if one of the faces of the cube is painted

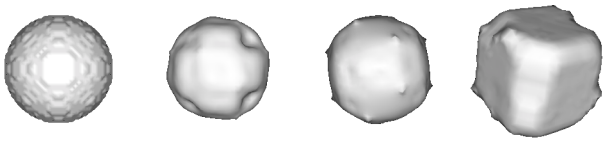

Figure 6: *Rendered reconstruction of the scene in figure 5 at various stages of processing.*

white, as in figure 7, no matter what the matching criterion, whether total variation, normalized cross-correlation, or the two-norm, a stereo algorithm cannot converge to the correct shape without additional prior information (figure 8). If, however, one enforces that - in the regions of the images in figure 7 where the gradient of irradiance is below a threshold - the reflectance of the scene is constant, so that the shading term in (13) is non-zero, then the algorithm can recover the correct shape, as shown in figure 9.

7 Conclusions

We have presented a framework to merge different shape cues by solving an optimization problem in the (infinite-dimensional) space of surfaces. The data are intensity images collected by a number of cameras, and the correspondence problem is re-formulated in terms of regions where prior assumptions on the reflectance properties of the surfaces can be enforced. For a simplified model, we prove the necessary conditions for optimality, and generate an iterative optimization algorithm using the tools of

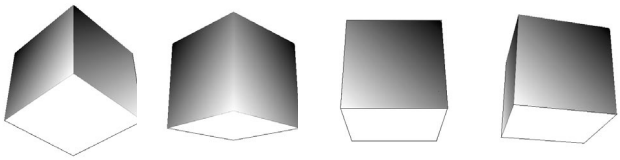

Figure 7: *Views of a cube with a white face.*

Figure 8: *Stereo-based reconstruction of the scene in figure 7: convergence is achieved everywhere but on the white face of the cube.*

calculus of variation. We implement the algorithm using ultra-narrowband level set methods, and perform some preliminary experiments on synthetic scenes. Although in the experiments we use a single light source illuminating a Lambertian surface, the algorithm is general and can – in principle – handle multiple sources (including mutual illumination between objects), more complex reflectance functions (including specular reflections) and topological changes in the estimated surface.

Our short-term agenda includes experimenting with real images and more general reflection models. We are also interested in including accommodation cues into our models, and addressing the characterization and computation of the correspondence set.

Although we use a deterministic formulation of the problem, a statistical interpretation is also possible, albeit not explored in this paper.

References

[1] A. Blake, A. Zisserman, and G. Knowles. Surface descriptions from stereo and shading. *Image and vision computing*, 3(4):183–191, 1985.

[2] A. Bruckstein. On shape from shading. *Computer Vision, Graphics and Image Processing*, 44(2):139–154, 1988.

[3] J. Cryer, P. Tsai, and M. Shah. Integration of shape from shading and stereo. In *Proc. IEEE Conf. on Comp. Vision and Pattern Recogn.*, pages 720–721, 1993.

[4] O. Faugeras. *Three dimensional vision, a geometric viewpoint*. MIT Press, 1993.

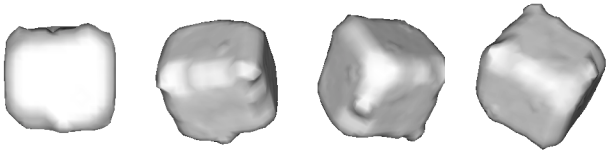

Figure 9: *Stereoscopic shading reconstruction of the scene in figure 7: convergence is achieved everywhere by enforcing uniform reflectance on the complement of the correspondence set.*

[5] O. Faugeras and R. Keriven. Variational principles, surface evolution pdes, level set methods and the stereo problem. *INRIA Technical report*, 3021:1–37, 1996.

[6] D. Forsyth and A. Zisserman. Reflections on shading. *IEEE Trans. Pattern Anal. Mach. Intell.*, 13(7):671–679, 1991.

[7] P. Fua and Y. Leclerc. Object-centered surface reconstruction: combining multi-image stereo and shading. *Int. J. of Computer Vision*, 16:35–56, 1995.

[8] B. Horn and M. Brooks. The variational approach to shape from shading. *Computer Vision, Graphics and Image Processing*, 33(2):174–208, 1985.

[9] B. Horn and M. Brooks (eds.). *Shape from Shading*. MIT Press, 1989.

[10] B. Horn, R. Szeliski, and A. Yuille. Impossible shaded images. *IEEE Trans. Pattern Anal. Mach. Intell.*, 15(2):166–170, 1993.

[11] K. Ikeuchi. Reconstructing a depth map from intensity maps. In *Proc. Intl. Conf. on Pattern Recognition*, pages 736–738, 1984.

[12] A. Jones and C. Taylor. Robust shape from shading. *Image and Vision Computing*, 12(7):411–421, 1994.

[13] R. Kimmel and A. Bruckstein. Tracking level sets by level sets: a method for solving the shape from shading problem. *Computer Vision, Graphics and Image Understanding*, (62)1:47–58, 1995.

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

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

[16] P.L. Lions, E. Roucy and A. Tourin. Shape from shading, viscosity solutions and edges. *Numerische Mathematik*, 64:323–353, 1993.

[17] M. Langer and S. Zucker. Shape from shading on a cloudy day. *J. Opt. Soc. Am. A*, 11(2):467–478, 1994.

[18] Y. Leclerc and A. Bobick. The direct computation of height from shading. In *Proc. IEEE Conf. on Comp. Vision and Pattern Recogn*, pages 552–558, 1991.

[19] R. J. LeVeque, *Numerical Methods for Conservation Laws* Birkhäuser, Boston, 1992.

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

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

[22] J. Oliensis and P. Dupuis. A global algorithm for shape from shading. In *Proc. of the ICCV*, pp. 692–701, 1993.

[23] S. Osher, "Riemann solvers, the entropy condition, and difference approximations," *SIAM J. Numer. Anal.*, vol. 21, pp. 217–235, 1984.

[24] S. Osher and J. Sethian, "Fronts propagation with curvature dependent speed: Algorithms based on Hamilton-Jacobi formulations," *Journal of Computational Physics*, vol. 79, pp. 12–49, 1988.

[25] S. Peleg and G. Ron. Nonlinear multiresolution: a shape from shading example. *IEEE Trans. Pattern Anal. Mach. Intell.*, 1(12):1206–1210, 1990.

[26] A. Pentland. Local shading analysis. *IEEE Trans. Pattern Anal. Mach. Intell.*, 6(2):170–187, 1984.

[27] T. Pong, R. Haralick, and L. Shapiro. Shape from shading using the facet model. In *Proc. IEEE Conf. on Comp. Vision and Pattern Recogn*, pages 683–695, 1989.

[28] V. Ramachandran. Perceiving shape from shading. *Scientific American*, 259(2):76–83, 1988.

[29] J. Sethian, *Level Set Methods: Evolving Interfaces in Geometry, Fluid Mechanics, Computer Vision, and Material Science*, Cambridge University Press, 1996.

[30] J. Shah, H. Pien, and J. Gauch. Recovery of surfaces with discontinuities by fusing shading and range data within a variational framework. *IEEE Trans. on Image Processing* 5(8):1243–1251, 1996.

[31] T. Simchony, R. Chellappa, and M. Shao. Direct analytical methods for solving Poisson equations in computer vision problems. *IEEE Trans. Pattern Anal. Mach. Intell.*, 12(5):435–446, 1990.

[32] G. Ward. Measuring and modeling anisotropic reflection. In *SIGGRAPH*, pages 265–272, 1992.

[33] D. Weinshall. Local shape approximation from shading. In *Proc. IEEE Conf. on Comp. Vision and Pattern Recogn.*, pages 716–718, 1992.

[34] A. Yuille and D. Snow. Shape and albedo from multiple images using integrability. In *Proc. of the Intl. Conf. on Comp. Vision and Pattern Recognition* pages 158–164, 1997.

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

[36] Q. Zheng and R. Chellappa. Estimation of illuminant direction, albedo and shape from shading. *IEEE Trans. Pattern Anal. Mach. Intell.*, 13(7):680–702, 1991.