Stereoscopic Segmentation

Anthony Yezzi[†] Stefano Soatto[‡*]

† School of Electrical and Computer Engineering

Georgia Institute of Technology, Atlanta – GA 30332

Tel: (404)385-1017, Fax: (404)894-4641, email ayezzi@ece.gatech.edu

‡ Department of Computer Science

University of California, Los Angeles – CA 90095

Tel: (310)825-4840, Fax: (310)794-5057, email soatto@ucla.edu

∗ Department of Electrical Engineering

Washington University, St.Louis – MO 63130

keywords: variational methods, Mumford-Shah functional, image segmentation, multi-frame stereo reconstruction, level set methods.

Abstract

We cast the problem of multiframe stereo reconstruction of a smooth shape as the global region segmentation of a collection of images of the scene. Dually, the problem of segmenting multiple calibrated images of an object becomes that of estimating the solid shape that gives rise to such images. We assume that the radiance has smooth statistics. This assumption covers Lambertian scenes with smooth or constant albedo as well as fine homogeneous textures, which are known challenges to stereo algorithms based on local correspondence. We pose the segmentation problem within a variational framework, and use fast level set methods to find the optimal solution numerically. Our algorithm does not work in the presence of strong textures, where traditional reconstruction algorithms do. It enjoys significant robustness to noise under the assumptions it is designed for.

1 Introduction

Inferring spatial properties of a scene from one or more images is a central problem in Computer Vision. When more than one image of the same scene is available, the problem is traditionally approached by first matching points or small regions across different images (local correspondence) and then combining the matches into a three-dimensional model[1]. Local correspondence, however, suffers from the presence of noise and local minima, which cause mismatches and outliers.

The obvious antidote to the curse of noise is to avoid local correspondence altogether by globally

[1]Since point-to-point matching is not possible due to the aperture problem, points are typically supported by small photometric patches that are matched using correlation methods or other cost functions based on a local deformation model. Sometime local correspondence and stereo reconstruction are combined into a single step, for instance in the variational approach to stereo championed by Faugeras and Keriven [11].

integrating visual information over regions in each image. This naturally leads to a segmentation problem. The diarchy between local and region-based methods is very clear in the literature on segmentation, where the latter are recognized as being more resistant to noise albeit more restrictive in their assumptions on the complexity of the scene[2]. The same cannot be said about stereo, where the vast majority of the algorithms proposed in the literature relies on local correspondence. Our goal in this paper is to formulate multiframe stereo as a global region segmentation problem, thus complementing existing stereo algorithms by providing tools that work when local correspondence fails.

We present an algorithm to reconstruct scene shape and radiance from a number of calibrated images. We make the *assumption* that the scene is composed by rigid objects that support radiance functions with *smooth statistics*. This includes Lambertian objects with smooth albedo (where local correspondence is ill-posed) as well as densely textured objects with isotropic or smoothly-varying statistics (where local correspondence is prone to multiple local minima). Therefore, our algorithm works under conditions that prevent traditional stereo or shape from shading to operate. However, it can provide useful results even under conditions suitable for shape from shading (constant albedo) and stereo (dense texture). The relation between these methods is illustrated pictorially in Figure 1, where the domain of operation of each scheme is shown on a plane that represents the amount of noise on the image and the texture density of the surfaces populating the scene.

[2]Local methods involve computing derivatives, and are therefore extremely sensitive to noise. Region-based methods involve computing integrals, and suffer less from noise.

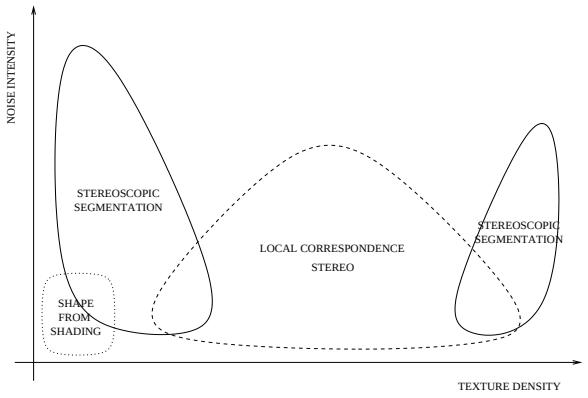

Figure 1: *Domain of operation of shape from shading, stereo and stereoscopic segmentation in a noise-texture plane: the horizontal axis indicates increased density in the texture of the surfaces in the scene, the vertical axis indicates the intensity of noise in the image.*

1.1 Relation to prior work

Since this paper touches the broad topics of segmentation and solid shape reconstruction, it relates to a vast body of work in the Computer Vision community. We emphasize the relations and differences with approaches based on different assumptions in the following table.

	ALBEDO	RADIANCE	LIGHT	OUTPUT
STEREO		$\neq$ const.		shape
CARVING		$\neq$ const.		largest phot. consist. shape, radiance
SHADING	const.		known	shape
THIS WORK		smooth statistics		smoothest phot. consist. shape, radiance

In local correspondence-based stereo (see [10] and references therein), one makes the assumption that the scene is Lambertian and the radiance is nowhere constant in order to recover a dense model of the three-dimensional structure of the scene. Faugeras and Keriven [11] pose the stereo reconstruction problem in a variational framework, where the cost function corresponds to the local matching score. In a sense, this work can be interpreted as extending the approach of [11] to regions. In shape carving [16], the same assumptions are used to recover a representation of

shape (the *largest* shape that is photometrically consistent with the data) as well as photometry. We use a different assumption, namely that radiance and shape are smooth, to recover a different representation (the *smoothest* shape that is photometrically consistent with the data in a variational sense) as well as photometry. Therefore, this work could be interpreted as performing space carving in a variational framework to minimize the effects of noise. Note, however, that once a pixel is deleted by the carving procedure, it can never be retrieved. In this sense, shape carving is *uni-directional*. Our algorithm, on the other hand, is bidirection, in that surfaces are allowed to evolve inward or outward. This work also relates to shape from shading [13] in that it can be used to recover shape from a number of images of scenes with constant albedo (although it is not bound by this assumption). However, traditional shape from shading operates on single images under the assumption of known illumination. There is also a connection to shape from texture algorithms [28] in that our algorithm can be used on scenes with dense texture, although it operates on multiple views as opposed to just one. Finally, there is a relationship between our reconstruction methods and the literature on shape from silhouettes [7], although the latter is based on local correspondence between occluding boundaries. In a sense, this work can be interpreted as a region-based method to reconstruct shape from silhouettes.

The material in this paper is tightly related to a wealth of contributions in the field of region-based segmentation, starting from Mumford and Shah's pioneering work [21], and including [2, 3, 8, 14, 15, 18, 32, 33, 27, 35]. This line of work stands to complement local contour-based segmentation methods such as [14, 33]. There are also algorithms that combine both features [4, 5].

In the methods used to perform the actual reconstruction, our work relates to the literature on level set methods of Osher and Sethian [24].

1.2 Contributions of this paper

We propose an algorithm to reconstruct solid shape and radiance from a number of calibrated views of a scene with smooth shape and radiance or homogeneous fine texture. To the best of our knowledge, work in this domain is novel. We forego local matching altogether and process (regions of) images globally, which makes our algorithms resistant to noise and local extrema in the local matching score. We work in a variational framework, which makes the enforcing of geometric priors such as smoothness simple, and use the level set methods of Osher and Sethian [24] to efficiently compute a solution.

Our algorithm does *not* work in the presence of strong textures or boundaries on the albedo; however, under those conditions traditional stereo algorithms based on local correspondence or shape carving algorithms do.

2 A variational formulation

We assume that a scene is composed of a number of smooth Lambertian surfaces supporting smooth Lambertian radiance functions (or dense textures with spatially homogeneous statistics). Under such assumptions, most of the significant irradiance discontinuities (or texture discontinuities) within any image of the scene correspond to occlusions between objects (or the background). These assumptions make the segmentation problem well-posed, although not general. In fact, "true" segmentation in this context corresponds directly to the shape and pose of the objects

in the scene[3]. Therefore, we set up a cost functional to minimize variations within each image region, where the free parameters are not the boundaries in the image themselves, but the shape of a surface in space whose occluding contours happen to project onto such boundaries.

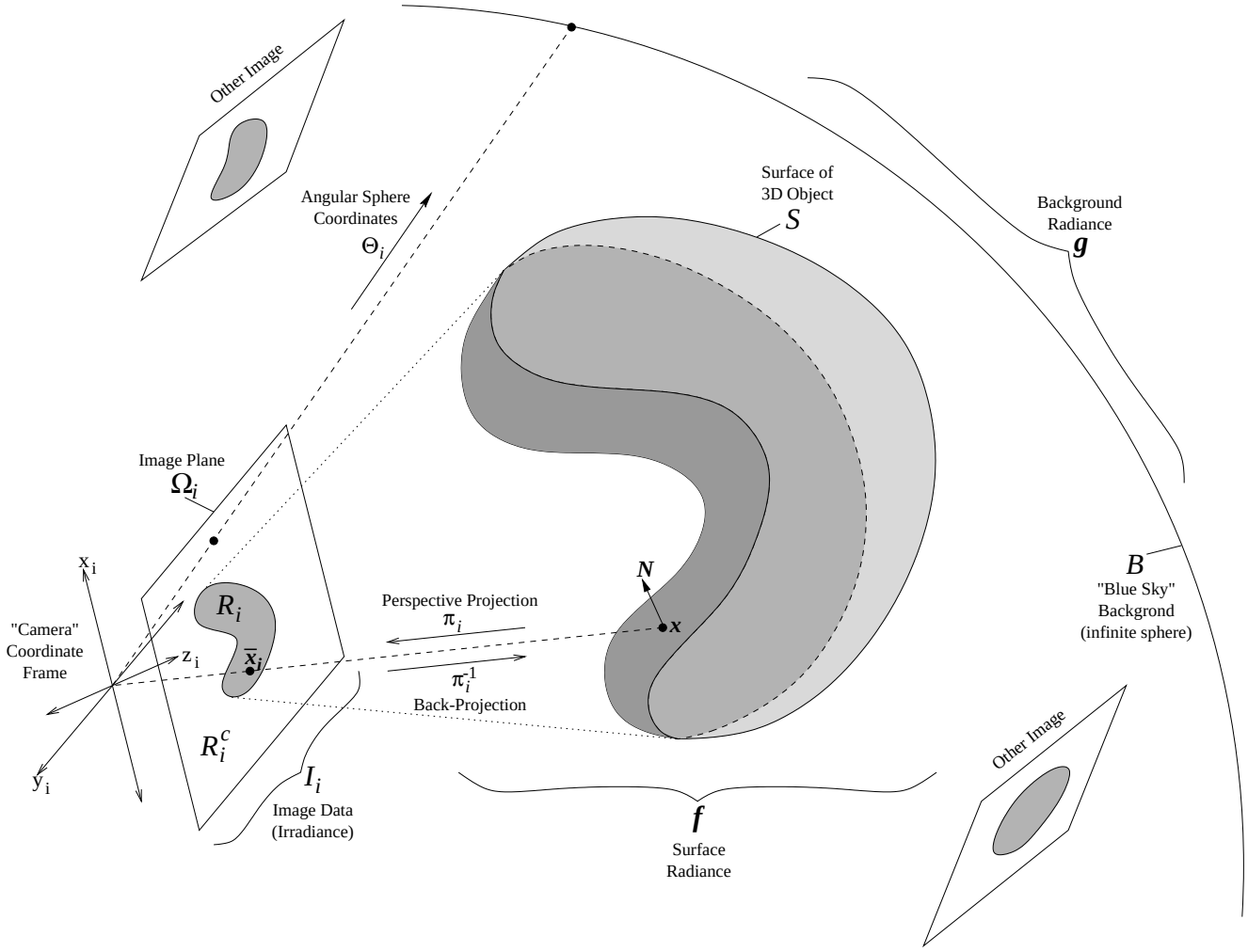

Figure 2: Illustration of notation used to describe the unknown surface and background (and their respective radiances), the camera images, and the relevant mappings between these objects.

[3]We consider the background to be yet another object that happens to occupy the entire field of view (the "blue sky" assumption).

2.1 Notation

In what follows $\mathbf{x} = (x, y, z)$ will represent a generic point of a scene in $\Re^3$ expressed in global coordinates (based upon a fixed inertial reference frame) while $\mathbf{x}_i = (x_i, y_i, z_i)$ will represent the same point expressed in "camera coordinates" relative to an image I_i (from a sequence of images $I_1, \ldots, I_n$ of the scene). To be more precise, we assume that the domain Ω_i of the image I_i belongs to a 2D plane given by $z_i = 1$ and that (x_i, y_i) constitute Cartesian coordinates within this image plane (the important point here is that the coordinate z_i corresponds to a direction which is perpendicular to the image plane). We let $\pi_i : \Re^3 \to \Omega_i; \mathbf{x} \mapsto \mathbf{x}_i = (\hat{x}_i, \hat{y}_i)$ denote an ideal perspective projection onto this image plane, where $\hat{x}_i = x_i/z_i$ and $\hat{y}_i = y_i/z_i$. The primary objects of interest will be a regular surface S in $\Re^3$ (with area element dA) supporting a radiance function $\mathbf{f} : S \to \Re$, and a background B which we treat as a sphere of infinite radius ("blue sky") with angular coordinates $\Theta = (\theta, \phi)$ that may be related in a one-to-one manner with the coordinates $\mathbf{x}_i$ of each image domain Ω_i through the mapping Θ_i (i.e. $\Theta = \Theta_i(\mathbf{x}_i)$). We assume that the background supports a different radiance function $\mathbf{g} : B \to \Re$. Given the surface S, we may partion the domain Ω_i of each image I_i into a "foreground" region $R_i = \pi_i(S) \subseteq \Omega_i$, which back-projects onto the surface S, and its complement R_i' (the "background" region), which back-projects onto the background. Although the perspective projection π_i is not one-to-one (and therefore not invertible), the operation of back-projecting a point from R_i onto the surface S (by tracing back along the ray defined by $\pi_i(\mathrm{ray}) = \mathbf{x}_i$ until the first point on S is encountered) is indeed one-to-one with π_i as it's inverse. Therefore, we will make a slight abuse of notation and denote back-projection onto the surface S by $\pi_i^{-1} : R_i \to S$. Finally, in our computations, we will make use of the relationship between the area measure dA of the surface S and the measure

$d\Omega_i = d\hat{x}_i d\hat{y}_i$ of each image domain. This arises from the form of the corresponding projection π_i and is given by $\hat{x}_i^3 d\Omega_i = -(\mathbf{x}_i \cdot N_i) dA$, where N_i denotes the outward unit normal N of S expressed in the same coordinate system as $\mathbf{x}_i$. This notation is illustrated in Figure 2.

2.2 Cost functional

In order to infer the shape of a surface S, we impose a cost on the discrepancy between the projection of a model surface and the actual measurements. Such a cost, $E(\mathbf{f}, \mathbf{g}, S)$, depends upon the surface S as well as upon the radiance of the surface $\mathbf{f}$ and of the background $\mathbf{g}$. We will then adjust the model surface and radiance to match the measured images. Since the unknowns (surface S and radiances $\mathbf{f}, \mathbf{g}$) live in an infinite-dimensional space, we need to impose regularization. In particular, we can leverage on our assumption that the radiance is smooth. However, this is still not enough, for the estimated surface could converge to a very irregular shape to match image noise and fine details. Therefore, we impose a geometric prior on shape (smoothness). These are the three main ingredients in our approach: a data fidelity term $E_{data}(\mathbf{f}, \mathbf{g}, S)$ that measures the discrepancy between measured images and images predicted by the model, a smoothness term for the estimated radiances $E_{rsmooth}(\mathbf{f}, \mathbf{g}, S)$ and a geometric prior $E_{gsmm}(S)$. We consider the composite cost functional to be the sum (or more generally a weighted sum) of these three terms:

$$E(\mathbf{f}, \mathbf{g}, S) = E_{data}(\mathbf{f}, \mathbf{g}, S) + E_{rsmooth}(\mathbf{f}, \mathbf{g}, S) + E_{gsmm}(S) \tag{1}$$

We conjecture that, like in the case of the Mumford-Shah functional [21], these ingredients are sufficient to define a unique solution to the minimization problem.

In particular, the geometric and smoothness terms are given by

$$E_{geom} = \int_S dA \tag{2}$$

$$E_{smooth} = \int_S \|\nabla_S \mathbf{f}\|^2 dA + \int_B \|\nabla \mathbf{g}\|^2 d\Theta \tag{3}$$

which favor surfaces S of least surface area and radiance functions $\mathbf{f}$ and $\mathbf{g}$ of least quadratic variation. (∇_S denotes the intrinsic gradient on the surface S). Finally, the data fidelity term may be measured in the sense of $\mathcal{L}^2$ by

$$E_{data} = \sum_{i=1}^{n} \left(\int_{R_i} (\mathbf{f}(\pi_i^{-1}(\mathbf{x}_i)) - I_i(\mathbf{x}_i))^2 d\Omega_i + \int_{R_i^c} (\mathbf{g}(\Theta_i(\mathbf{x}_i)) - I_i(\mathbf{x}_i))^2 d\Omega_i \right). \tag{4}$$

In order to facilitate the computation of the first variation with respect to S, we would rather express these integrals over the surface S as opposed to the partitions R_i and R_i^c. We start with the integrals over R_i and note that they are equivalent to

$$-\int_{\pi_i^{-1}(R_i)} (\mathbf{f}(\mathbf{x}) - I_i(\pi_i(\mathbf{x})))^2 \frac{\mathbf{x}_i \cdot N_i}{z_i^3} dA = \int_{\pi_i^{-1}(R_i)} \epsilon_i^2(\mathbf{x}) \sigma_i(\mathbf{x}, N)\, dA$$

where $\epsilon_i(\mathbf{x}) = \mathbf{f}(\mathbf{x}) - I_i(\pi_i(\mathbf{x}))$ and $\sigma_i(\mathbf{x}, N) = -(\mathbf{x}_i \cdot N_i)/z_i^3$. Now we move to the integrals over R_i^c and note that they are equivalent to

$$\int_{\Omega_i} \epsilon_i^2(\mathbf{x}_i)\, d\Omega_i - \int_{\pi_i^{-1}(R_i)} \epsilon_i^2(\pi_i(\mathbf{x})) \sigma_i(\mathbf{x}, N)\, dA$$

where $e_i(\hat{\mathbf{x}}_i) = \mathbf{g}(\Theta_i(\hat{\mathbf{x}}_i) - I_i(\hat{\mathbf{x}}_i))$. Combining these "restructured" integrals yields:

$$\int_{\Omega_i} e_i^2(\hat{\mathbf{x}}_i)\, d\Omega_i + \int_{\pi_i^{-1}(R_i)} (e_i^2(\mathbf{x}) - e_i^2(\pi_i(\mathbf{x})))\sigma_i(\mathbf{x}, N)\, dA$$

Note that the first integral in the above expression is independent of the surface S (and its radiance function $\mathbf{f}$) and that the second integral is taken over only a subset of S given by ${\pi_i}^{-1}(R_i)$. We may express this as an integral over all of S (and thereby avoid the use of ${\pi_i}^{-1}$ in our expression) by introducing a characteristic function $\chi_i(\mathbf{x}) \in \{0, 1\}$ into the integrand where $\chi_i(\mathbf{x}) = 1$ for $\mathbf{x} \in {\pi_i}^{-1}(R_i)$ (represented by the shaded part of the surface S in Figure 2) and $\chi_i(\mathbf{x}) = 0$ for $\mathbf{x} \notin {\pi_i}^{-1}(R_i)$ (i.e. for points that are occluded by other points on S). We therefore obtain the following equivalent expression for E_{data} given in (4):

$$E_{data} = \sum_{i=1}^{n} \left(\int_{\Omega_i} e_i^2(\hat{\mathbf{x}}_i)\, d\Omega_i + \int_S \chi_i(\mathbf{x})(e_i^2(\mathbf{x}) - e_i^2(\pi_i(\mathbf{x})))\sigma_i(\mathbf{x}, N)\, dA \right). \tag{5}$$

Finally, we express the entire cost functional by adding the geometric and smoothness priors:

$$\begin{aligned} E(\mathbf{f}, \mathbf{g}, S) &= \sum_{i=1}^{n} \left(\int_{\Omega_i} e_i^2(\hat{\mathbf{x}}_i)\, d\Omega_i + \int_S \chi_i(\mathbf{x})(e_i^2(\mathbf{x}) - e_i^2(\pi_i(\mathbf{x})))\sigma_i(\mathbf{x}, N)\, dA \right) \\ &+ \int_S \|\nabla_S \mathbf{f}\|^2 dA + \int_R \|\nabla \mathbf{g}\|^2 d\Theta + \int_S dA. \end{aligned}$$

2.3 Evolution equation

In order to find the surface S and the radiances $\mathbf{f}$ and $\mathbf{g}$ that minimize the functional (1) we set up an iterative procedure where we start from an initial surface S, compute optimal radiances

$\mathbf{f}$ and $\mathbf{g}$ based upon this surface, and then update S through a gradient flow based on the first variation of $E(\mathbf{f},\mathbf{g},S)$ which we denote by $\frac{d}{dS}E$ (then new radiance estimates are obtained in order to update S again). The variation of E_{geom}, which is just the surface area of S, is given by

$$-\frac{d}{dS}E_{geom} = -HN,$$

where H denotes mean curvature and N the outward unit normal. The variation of E_{smooth} is given by

$$-\frac{d}{dS}E_{smooth} = \left(K\left\langle \nabla_S \mathbf{f}, A^{-1}\nabla_S \mathbf{f}\right\rangle - \|\nabla_S \mathbf{f}\|H\right)N,$$

where K denotes the Gaussian curvature of S, where ∇_S denotes the gradient of $\mathbf{f}$ taken with respect to isothermal coordinates (the "intrinsic gradient" on S), and where A denotes the second fundamental form of S with respect to these coordinates.

The variation of E_{data} requires some attention. In fact, the data fidelity term in (5) involves an explicit model of occlusions[4] via a characteristic function. Discontinuities in the kernel cause major problems, for they can result in variations that are zero almost everywhere (e.g. for the case of constant radiance). One easy solution is to mollify the corresponding gradient flow. This can be done in a mathematically sound way by interpolating a smooth force field on the surface in space. Alternatively, the characteristic functions χ_S in the data fidelity term can be mollified, thereby making the integrands differentiable everywhere.

In order to arrive at an evolution equation, we note that the components of the data fidelity

[4]The geometric term and the smoothness term are independent of occlusions.

term, as expressed in equation (5), that depend upon S have the following form

$$E_i(S) = \int_S G_i(\mathbf{x}) \cdot N_i \, dA. \tag{6}$$

The gradient flows corresponding to such energy terms have the form

$$-\frac{d}{dS} E_i = -(\nabla_i \cdot G_i) N, \tag{7}$$

where ∇_i denotes the gradient with respect to $\mathbf{x}_i$ (recall that $\mathbf{x}_i$ is the representation of a point using the camera coordinates associated with image I_i as described in Section 2.1). In particular,

$$G_i(\mathbf{x}) = -\chi_i(\mathbf{x}) \left(\epsilon_i^2(\mathbf{x}) - \epsilon_i^2(\pi_i(\mathbf{x})) \right) \frac{\mathbf{x}_i}{z_i^3} \tag{8}$$

and the divergence of G_i, after simplification, is given by

$$-\nabla_i \cdot G_i = \frac{1}{z_i^3} \Big((\mathbf{f} - \mathbf{g}) \big[(I_i - \mathbf{f}) + (I_i - \mathbf{g}) \big] (\nabla_i \chi_i \cdot \mathbf{x}_i) + 2\chi_i (I_i - \mathbf{f}) (\nabla_i \mathbf{f} \cdot \mathbf{x}_i) \Big), \tag{9}$$

where we have omitted the arguments of $\mathbf{f}$, $\mathbf{g}$, and I_i for the sake of simplicity. A particularly nice feature of this final expression (which is shared by the standard Mumford-Shah formulation for direct image segmentation) is that it depends only upon the image values, *not upon the image gradient*, which makes it less sensitive to image noise when compared to other variational approaches to stereo (and therefore less likely to cause the resulting flow to become "trapped" in local minima). Notice that the first term in this flow involves the gradient of the characteristic function χ_i (see Appendix A) and is therefore non-zero only on the portions of S which project

(π_i) onto the *boundary* of the region R_i (illustrated by the dashed curve on the surface S in Figure 2). As such, this term may be directly associated with a curve evolution equation for the boundary of the region R_i within the domain Ω_i of the image I_i. The second term, on the other-hand, may be non-zero over the entire patch ${\pi_i}^{-1}(R_i)$ of S (illustrated by entire the shaded part of the surface S in Figure 2).

We may now write down the complete gradient flow for $E = E_{data} + E_{smooth} + E_{geom}$ as

$$\begin{aligned}\frac{dS}{dt} &= \sum_{i=1}^{n} \frac{1}{{\chi_i}^2}\Big((\mathbf{f}-\mathbf{g})\Big[(I_i-\mathbf{f})+(I_i-\mathbf{g})\Big](\nabla\chi_i\cdot\mathbf{x}_i)+2\chi_i(I_i-\mathbf{f})(\nabla\mathbf{f}\cdot\mathbf{x}_i)\Big)N \\ &+ \Big(K\left\langle\nabla_S\mathbf{f}, A^{-1}\nabla_S\mathbf{f}\right\rangle - \|\nabla_S\mathbf{f}\|H\Big)N - HN. \end{aligned} \tag{10}$$

2.4 Estimating scene radiance

Once an estimate of the surface S is available, the radiance estimates $\mathbf{f}$ and $\mathbf{g}$ must be updated. For a given surface S we may regard our energy functional $E(S,\mathbf{f},\mathbf{g})$ as a function only of $\mathbf{f}$ and $\mathbf{g}$ and minimize it accordingly. A necessary condition is that $\mathbf{f}$ and $\mathbf{g}$ satisfy the Euler-Lagrange equations for E based upon the current surface S. These optimal estimate equations are given by the following elliptic PDEs on the surface S and the background B,

$$\Delta_S\mathbf{f} = \sum_{i=1}^{n}\chi_i(\mathbf{f}-I_i)\pi_i \qquad \text{and} \qquad \Delta_B\mathbf{g} = \sum_{i=1}^{n}\hat{\chi}_i(\mathbf{g}-I_i) \tag{11}$$

where Δ_S denotes the Laplace-Beltrami operator on the surface S, where Δ_B denotes the Laplacian on the background B with respect to its spherical coordinates Θ, and where $\hat{\chi}_i(\Theta)$ denotes a characteristic function for the background B where $\hat{\chi}_i(\Theta)=1$ if ${\Theta_i}^{-1}(\Theta) \in R_i^S$ and $\hat{\chi}_i(\Theta)=0$

otherwise.

3 The piecewise constant case

The full development and implementation of the flow corresponding to the case of general piecewise smooth statistics as described in Section 2 is well beyond the scope of this paper. In Section 3.1 we specialize the derivation of Section 2 to the case of scenes with piecewise constant albedo statistics (i.e. the functions $\mathbf{f}$ and $\mathbf{g}$ defined on the surface S and background B are approximated as constants). This will result in a particularly simple and elegant flow that is easily implemented.

Despite the apparent restrictiveness of the assumptions (i.e. despite the simplicity of the class of scenes captured by this model) we show that the resulting implementation – which uses level set methods – is extremely robust, to the point of easily tolerating significant violations of such assumptions. We do so in Section 3.2 by means of experiments on real image sequences that patently depart from the model

3.1 Optimal Estimates and Gradient Flow

We obtain a special piecewise constant[5] energy functional as a limiting case of the more general energy functional (1) by giving the smoothness term E_{smooth} infinite weight. In this case, the only critical points are constant radiance functions. We may obtain an equivalent formulation, by

[5] We say "piecewise constant" even though each radiance function is treated as a single constant since the segmentations obtained by projecting these objects with constant radiances onto the camera images yield *piecewise constant* approximations to the image data.

dropping the E_{smooth} term

$$\begin{aligned} E_{constant} &= E_{data} + E_{smooth} \qquad (12) \\ &= \sum_{i=1}^{n} \left(\int_{R_i} (\mathbf{f}(\pi_i^{-1}(\hat{\mathbf{x}}_i)) - I_i(\hat{\mathbf{x}}_i))^2 d\Omega_i + \int_{R_i^c} (\mathbf{g}(\Theta_i(\hat{\mathbf{x}}_i)) - I_i(\hat{\mathbf{x}}_i))^2 d\Omega_i \right) + \int_S dA, \end{aligned}$$

and by restricting our class of admissible radiance functions $\mathbf{f}$ and $\mathbf{g}$ to be only constants. This is analogous to the segmentation work of Chan and Vese in [6] who consider the piecewise-constant version of the Mumford-Shah functional for the segmentation of images with bimodal statistics. In this simpler formulation, one no longer needs to solve a PDE on the surface S nor on the background B, to obtain optimal estimates for $\mathbf{f}$ and $\mathbf{g}$ (given the current location of the surface S). In this case, $E_{constant}$ is minimized by setting the constants $\mathbf{f}$ and $\mathbf{g}$ to be the overall sample mean of I_i over the regions R_i (for each $1 \leq i \leq n$) and the overall sample mean of I_i over the complementary regions R_i^c respectively. The gradient flow associated with the E_{data} term simplifies. Recall that, in the general case, the E_{data} gradient flow depends upon two terms (given by (9)), one of which only acts upon the points of S which project to the boundaries of the regions R_i, giving rise to curve evolutions for these segmentation boundaries, while the second term acts upon each entire patch of S associated with each region R_i. In the piecewise constant case, this second term drops out (since it depends upon the gradient of $\mathbf{f}$), and therefore only the boundary evolution term remains. As such, the gradient flow for $E_{constant}$ is given by

$$\frac{dS}{dt} = \Big(\sum_{i=1}^{n} \frac{1}{\hat{z}_i^3} (\mathbf{f} - \mathbf{g}) \Big[(I_i - \mathbf{f}) + (I_i - \mathbf{g}) \Big] (\nabla \chi_i \cdot \mathbf{x}_i) N \Big) - HN. \qquad (13)$$

3.2 Experiments

A numerical implementation of the evolution equation above has been carried out within the level set framework of Osher and Sethian [24]. A number of sequences has been captured and the relative position and orientation of each camera has been computed using standard camera calibration methods. Here we show the results on two representative experiments. Although the equation above assumes that the scene is populated by objects with constant albedo, the reader will recognize that the scenes we have tested our algorithms on represent significant departures from such assumptions. They include fine textures, specular highlights and even substantial calibration errors.

In Figure 3 we show 4 of 22 calibrated views of a scene that contains three objects: two shakers and the background. This scene would represent an insurmountable challenge to traditional correspondence-based stereo algorithms: the shakers exhibit very little texture (making local correspondence ill-posed), while the background exhibits very dense texture (making local correspondence prone to local minima). In addition, the shakers have a dark but shiny surface, that reflects highlights that move relative to the camera since the scene is rotated while the light is kept stationary. In Figure 4 we show the surface evolving from a large ellipse that neither contains nor is contained in the shape of the scene, to a final solid model. Notice that some parts of the initial surface evolve outward, while other parts evolve inward in order to converge to the final shape. This bi-directionality is a feature of our algorithm, which is not shared - for instance - by shape carving methodologies. There, once a pixel has been deleted, it cannot be retrieved. In Figure 5 we show the final result from various vantage points. In Figure 6 we show the final segmentation in some of the original views (top). We also show the segmented foreground superimposed to the

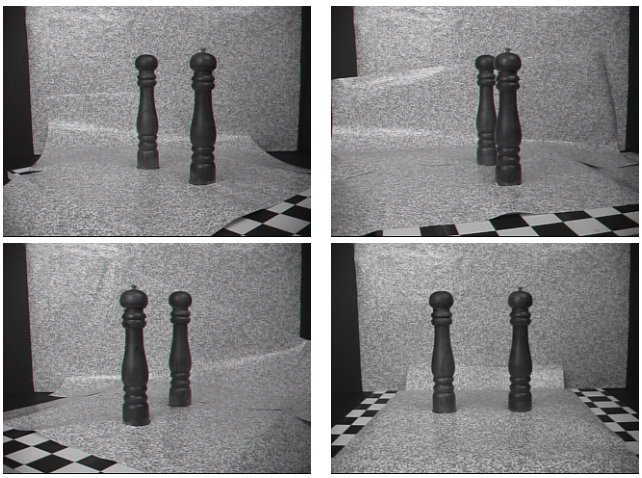

Figure 3: *Original "salt and pepper" sequence (4 of 22 views).*

original images. Two of the 22 views were poorly calibrated, as it can be seen from the large reprojection error. However, this does not significantly impact the final reconstruction, for there is an averaging effect by integrating data from all views.

In Figure 7 we show an image from a sequence of views of a watering can, together with the initial surface. The estimated shape is shown in Figure 8

4 Discussion

We have presented a fast and efficient algorithm to reconstruct solid shape and radiance from a number of calibrated views of a scene with smooth shape and constant albedo statistics. To the best of our knowledge the algorithm is new. It enjoys significant resistance to noise since it does not rely on local correspondence, and its domain of operation nicely complements existing stereo reconstruction algorithms. The experiments show that the algorithm can tolerate errors in camera calibration as well as significant violations of the (admittedly restrictive) assumptions that the algorithm is based on. Extending our implementation to handle piecewise smooth radiances and

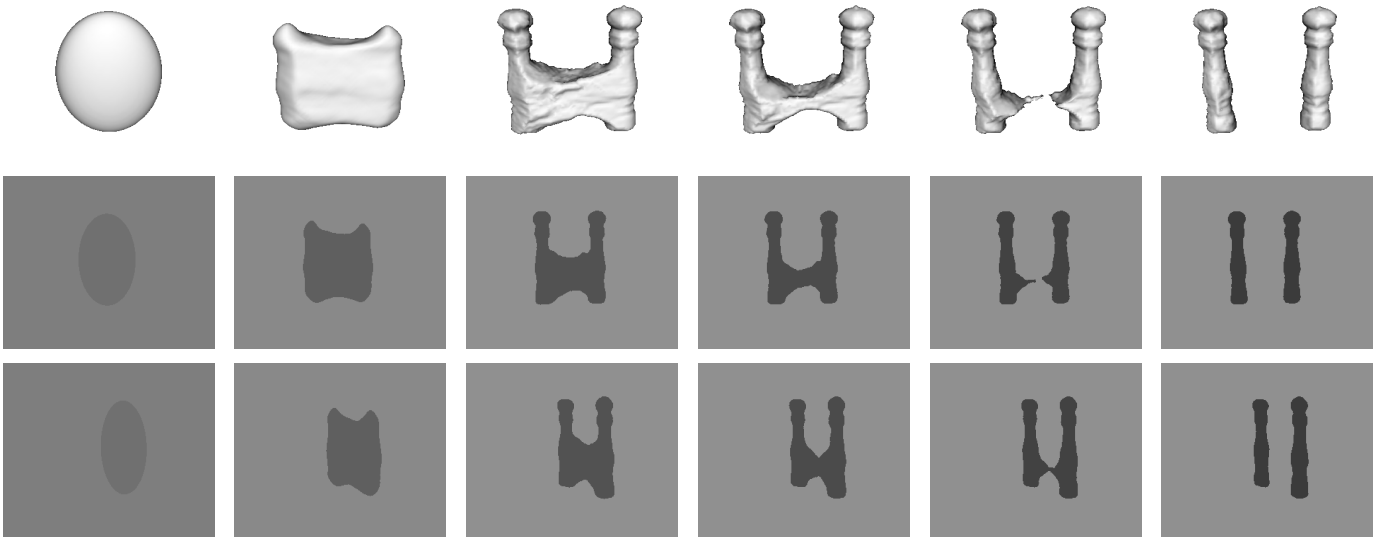

Figure 4: (Top) Rendered surface during evolution (6 of 800 steps). Notice that the initial surface is neither contains nor is contained by the final surface. (Bottom) segmented image during the evolution from two different viewpoints.

textures is by no means straightforward and is part of our future research agenda.

Acknowledgements

This research is supported in part by NSF grant IIS-9876145, ARO grant DAAD19-99-1-0139 and Intel grant 8029. The authors wish to thank Hailin Jin for his assistance on the experiments. We thank Wilfrid Gangbo and Luigi Ambrosio for very helpful discussions and insights, Allen Tannenbaum and Guillermo Sapiro for comments and suggestions. We also wish to thank Luminita Vese, Tony Chan and Stan Osher for comments.

References

[1] M. Bertalmio, L. Cheng, S. Osher, and G. Sapiro, "Variational Problems and Partial Differential Equations on Implicit Surfaces: The Framework and Examples in Image Processing

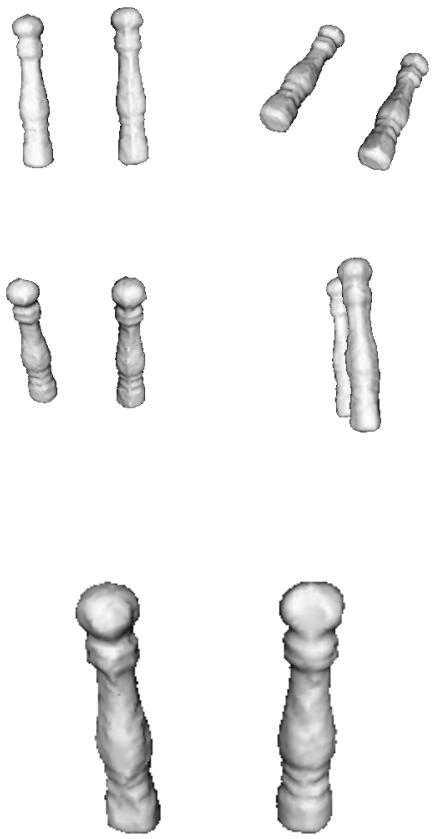

Figure 5: *Final estimated surface shown from several viewpoints. Notice that the bottoms of the salt and pepper shakers are flat, even though no data was available. This is due to the geometric prior, which in the absence of data results in a minimal surface being computed.*

and Pattern Formation". *CAM Technical Report 00-23*, UCLA, June 2000.

[2] V. Caselles, F. Catte, T. Coll, and F. Dibos, "A geometric model for active contours in image processing," *Numerische Mathematik*, vol. 66, pp. 1–31, 1993.

[3] V. Caselles, R. Kimmel, G. Sapiro, "Geodesic snakes," *Int. J. Computer Vision*, 1998.

[4] A. Chakraborty and J. Duncan, "Game-Theoretic Integration for Image Segmentation," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 21, no. 1, pp. 12–30, Jan. 1999.

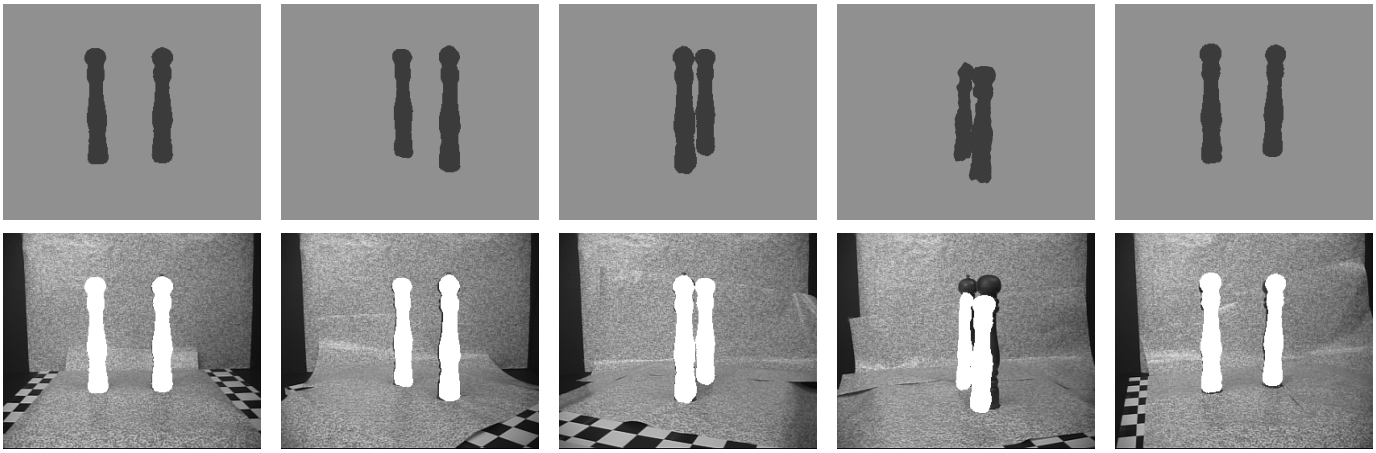

Figure 6: *(Top) Image segmentation for the salt and pepper sequence. (Bottom) Segmented foreground superimposed to the original sequence. The calibration in two of the 22 images was inaccurate (one of which is shown above; the fourth image from the left). However, the effect is mitigated by the global integration, and the overall shape is only marginally affected by the calibration errors.*

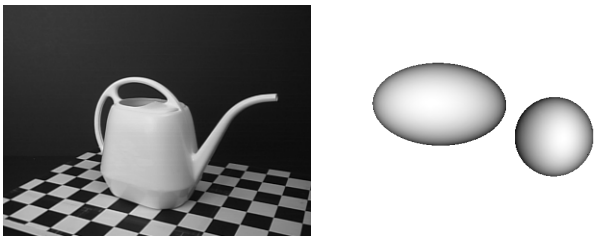

Figure 7: *The "watering can" sequence and the initial surface. Notice that the initial surface is not simply connected and neither contains nor is contained by the final surface. In order to capture a hole it is necessary that it is intersected by the initial surface. One way to guarantee this is to start with a number of small surfaces.*

[5] A. Chakraborty, L. Staib, and J. Duncan, "Deformable Boundary Finding in Medical Images by Integrating Gradient and Region Information," *IEEE Trans. Medical Imaging*, vol. 15, no. 6, pp. 859–870, Dec. 1996.

[6] T. Chan and L. Vese, "An active contours model without edges," *Int. Conf. Scale-Space Theories in Computer Vision*, pp. 141-151, 1999.

[7] R. Cipolla and A. Blake. Surface shape from the deformation of apparent contours. *Int. J.*

of Computer Vision, 9 (2), 1992.

[8] L. Cohen, "On active contour models and balloons," *CVGIP: Image Understanding*, vol. 53, pp. 211–218, 1991.

[9] M. Crandall, H. Ishii, and P. Lions, "Users guide to viscosity solutions of second order partial differential equations," *Bulletin of Amer. Math. Soc.*, vol. 27, pp. 1–67, 1992.

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

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

[12] W. Fleming and H. Soner, *Controlled Markov processes and viscosity solutions*. Springer-Verlag, New York, 1993.

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

[14] M. Kass, A. Witkin, and D. Terzopoulos, "Snakes: active contour models," *Int. Journal of Computer Vision*, vol. 1, pp. 321–331, 1987.

[15] S. Kichenassamy, A. Kumar, P. Olver, A. Tannenbaum, and A. Yezzi, "Conformal Curvature Flows: From Phase Transitions to Active Vision," *Arch. Rational Mech. Anal.*, vol. 134, pp. 275–301, 1996.

[16] K. Kutulakos and S. Seitz. A theory of shape by space carving. In *Proc. of the Intl. Conf. on Comp. Vision*, 1998.

[17] Y. Leclerc, "Constructing stable descriptions for image partitioning," *Int. J. Computer Vision*, vol. 3, pp. 73–102, 1989.

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

[19] P. L. Lions, *Generalized Solutions of Hamilton-Jacobi Equations*, Pitman Publishing, Boston, 1982.

[20] R. Malladi, J. Sethian, and B. Vemuri, "Shape modeling with front propagation: a level set approach," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 17, pp. 158–175, 1995.

[21] D. Mumford and J. Shah. Optimal approximations by piecewise smooth functions and associated variational problems. *Comm. on Pure and Applied Mathematics*, 42:577–685, 1989.

[22] D. Mumford and J. Shah, "Boundary detection by minimizing functionals," *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, San Francisco, 1985.

[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 propagating with curvature-dependent speed: algorithms based on Hamilton-Jacobi equations. *J. of Comp. Physics*, 79:12–49, 1988.

[25] N. Paragios and R. Deriche, "Geodesic Active Regions for Supervised Texture Segmentation," *Proceedings of ICCV*, Sept. 1999, Corfu, Greece.

[26] N. Paragios and R. Deriche, "Coupled Geodesic Active Regions for Image Segmentation: a level set approach," *Proceedings of ECCV*, June 2000, Dublin, Ireland.

[27] R. Ronfard, "Region-Based Strategies for Active Contour Models," *Int. J. Computer Vision*, vol. 13, no. 2, pp. 229–251, 1994.

[28] R. Rosenholtz and J. Malik. A differential method for computing local shape-from-texture for planar and curved surfaces. UCB-CSD 93-775, Computer Science Division, University of California at Berkeley, 1993.

[29] C. Samson, L. Blanc-Feraud, G. Aubert, and J. Zerubia. "A Level Set Method for Image Classification," *Int. Conf. Scale-Space Theories in Computer Vision*, pp. 306-317, 1999.

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

[31] K. Siddiqi, Y. Lauziere, A. Tannenbaum, and S. Zucker, "Area and length minimizing flows for segmentation," *IEEE Trans. Image Processing*, vol. 7, pp. 433–444, 1998.

[32] H. Tek and B. Kimia, "Image segmentation by reaction diffusion bubbles," *Proc. Int. Conf. Computer Vision*, pp. 156–162, 1995.

[33] D. Terzopoulos and A. Witkin, "Constraints on deformable models: recovering shape and non-rigid motion," *Artificial Intelligence*, vol. 36, pp. 91–123, 1988.

[34] A. Yezzi, A. Tsai, and A. Willsky, "A Statistical Approach to Image Segmentation for Bimodal and Trimodal Imagery," *Proceedings of ICCV*, September, 1999.

[35] S. Zhu and A. Yuille, "Region Competition: Unifying snakes, Region Growing, and Bayes/MDL for Multiband Image Segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 18, no. 9, pp. 884–900, Sep. 1996.

A Computing the gradient of χ

Now we will address the term $\nabla\chi \cdot S$, where $\nabla\chi$ must be defined in the distributional sense because the characteristic function χ is discontinuous. In the case of a convex surface, $\chi(S) = 1$ when $S \cdot N < 0$, and $\chi(S) = 0$ when $S \cdot N > 0$, or

$$\chi(S) = \mathcal{H}(-S \cdot N(S))$$

where $\mathcal{H}$ denotes the standard Heaviside (unit step) function. This expression, however, is only defined for points on the surface[6]. We may obtain an expression valid everywhere if we extend the vector field of unit outward normals N from the surface S to the surrounding embedding space. Let us call this extension $\bar{N}$. We will soon see that the particular choice of extension will not affect the final gradient flow. We now write

$$\chi(X) = \mathcal{H}(-X \cdot \bar{N}(X)) \quad \text{where} \quad \bar{N}(S) = N(S),$$

and express the gradient as

$$\nabla\chi(X) = -\left(\bar{N} + \bar{N}_X^T X\right) \delta(X \cdot \bar{N})$$

[6]This is only one possible expression for $\chi(S)$. There are many other valid ways to express this function that will lead to different expressions for its gradient. However, in each case the expression will ultimately involve a delta function. The primary purpose of this appendix is to establish that it is possible to construct expressions for $\nabla\chi(S)$ which are intrinsic to S (i.e. do not depend upon how χ is extended away from the surface S).

which, at a point S on the surface, has the following form

$$\nabla\chi = -\left(N + \bar{N}_X^T S\right)\delta(S \cdot N).$$

To perform the computation of $\bar{N}_X^T S$, we note that it is possible to express the spatial coordinates $X = (x, y, z)$ in a neighborhood of any point S_0 on the surface S where $S_0 \cdot N_0 = 0$ (the only points for which the above expression is nonzero) in terms of local curvilinear coordinates $U = (u, v, w)$ where (u, v) are coordinates for S which are related to each other and to w as follows:

1. $X(U) = (x(u,v,w), y(u,v,w), z(u,v,w))$
2. $X(u,v,0) = S(u,v)$ and therefore $X_u(u,v,0) = S_u(u,v),\ X_v(u,v,0) = S_v(u,v)$
3. $X_w(u,v,0) = N(u,v)$
4. $S_u \times S_v = N, \quad S_u \times N = S_u, \quad N \times S_v = S_v \qquad (E{=}1, F{=}0, G{=}1)$
5. $S(0,0) = S_0$
6. $S_u(0,0) = S_0/\|S_0\|$ (possible since $S_0 \cdot N_0 = 0$).

We may now obtain an expression for $\bar{N}_X$ by noting that $\bar{N}_U = \bar{N}_X X_U$ and so

$$\begin{aligned}
\bar{N}_X &= \bar{N}_U {X_U}^{-1} \\
&= \begin{bmatrix} \bar{N}_u & \bar{N}_v & \bar{N}_w \end{bmatrix} \begin{bmatrix} X_u & X_v & X_w \end{bmatrix}^{-1} \\
&= \begin{bmatrix} \bar{N}_u & \bar{N}_v & \bar{N}_w \end{bmatrix} \frac{\begin{bmatrix} X_v \times X_w & X_w \times X_u & X_u \times X_v \end{bmatrix}^T}{(X_u \times X_v) \cdot X_w}.
\end{aligned}$$

When evaluated at the surface point S_0, we may use the relations above to obtain

$$\begin{aligned}
\bar{N}_\chi &= \begin{bmatrix} N_u & N_v & \bar{N}_w \end{bmatrix} \frac{\begin{bmatrix} S_u \times N & N \times S_v & S_u \times S_v \end{bmatrix}^T}{(S_u \times S_v) \cdot N} \\
\bar{N}_\chi^T S_0 &= \frac{\begin{bmatrix} S_u & S_v & N \end{bmatrix}}{N \cdot N} \begin{bmatrix} N_u \cdot S_0 \\ N_v \cdot S_0 \\ \bar{N}_w \cdot S_0 \end{bmatrix} \\
&= (N_u \cdot S_0) S_u + (N_v \cdot S_0) S_v + (\bar{N}_w \cdot S_0) N.
\end{aligned}$$

We now use the fact that S_u, S_v, N are orthonormal and that $S_0 = S_u \|S_0\|$ to obtain

$$\begin{aligned}
(\bar{N}_\chi^T S_0) \cdot S_0 &= \|S_0\| (N_u \cdot S_0) \\
= \|S_0\|^2 (N_u \cdot S_u) &= -\|S_0\|^2 (N \cdot S_{uu}) = -\kappa_u \|S_0\|^2
\end{aligned}$$

yielding

$$\nabla \chi \cdot S = \kappa_u \|S\|^2 \delta(S \cdot N)$$

where κ_u denotes the normal curvature of S in the u-direction (the direction S when $S \cdot N = 0$). The key point to note here is that one may derive an intrinsic expression for $\nabla_S \chi \cdot \mathbf{x}_i$ which ultimately does not depend upon the extension of any function (or vector field in this case) from the surface into its local neighborhood. As such, we may legitimately model $\nabla_S \chi_S \cdot \mathbf{x}_i$ in (10) and (13) as a delta function without being concerned that ∇_S is an operator on the ambient space of the surface S as opposed to the surface itself.

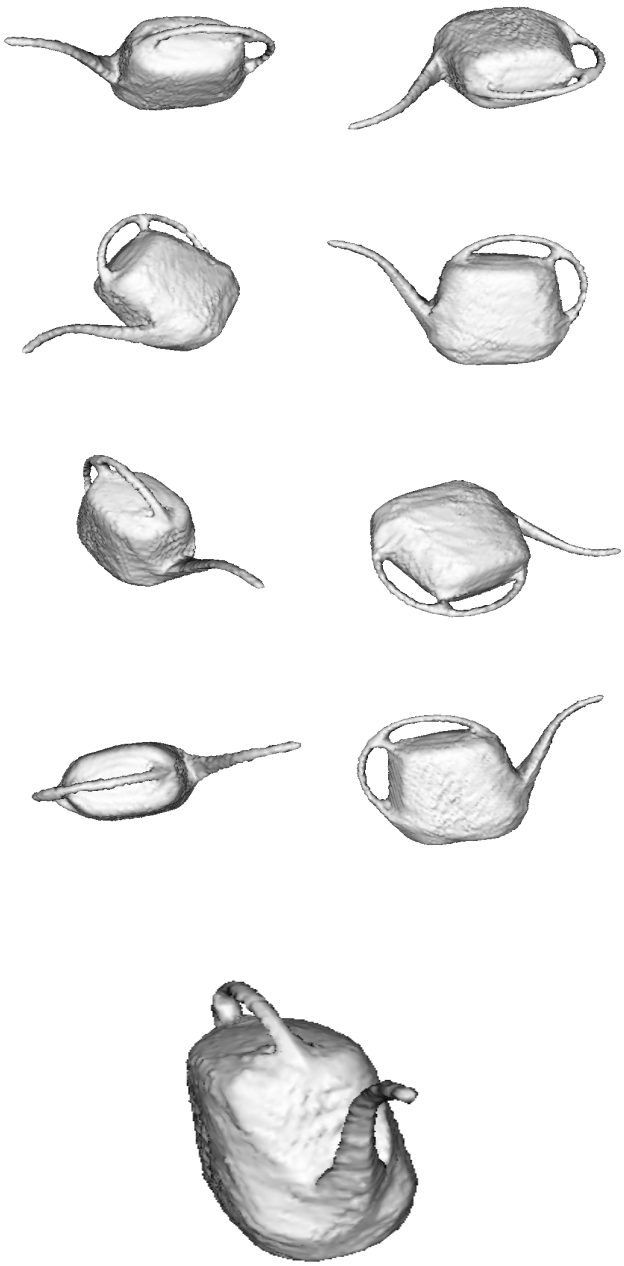

Figure 8: Final estimated shape for the watering can. The two initial surfaces, as seen in Figure 7, have merged. Although no ground truth is available for these sequences, it is evident that the topology and geometry of the watering can has been correctly captured.

Figure 9: *(top) Rendered surface during evolution for the watering can.*