International Journal of Computer Vision 53(1), 31–43, 2003

Stereoscopic Segmentation

ANTHONY YEZZI
School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, GA 30332
ayezzi@ece.gatech.edu

STEFANO SOATTO
Department of Computer Science, University of California, Los Angeles, CA 90095
soatto@ucla.edu

Received August 15, 2001; Revised July 10, 2002; Accepted December 12, 2002

Abstract. We cast the problem of multiframe stereo reconstruction of a smooth surface 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 of the scene results in piecewise homogeneous image statistics. This simplifying assumption covers Lambertian scenes with 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 photometric features, where traditional reconstruction algorithms do. It enjoys significant robustness to noise under the assumptions it is designed for.

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

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 can cause mismatches.

The obvious antidote to the curse of noise is to avoid local correspondence altogether by integrating visual information over regions in each image and seeking for their global deformation across different images. This naturally leads to a segmentation problem.

The dyarchy 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.

Assumptions: We present an algorithm for reconstructing scene shape and radiance from a number of calibrated images. We make the assumption that the scene is composed by rigid objects that support

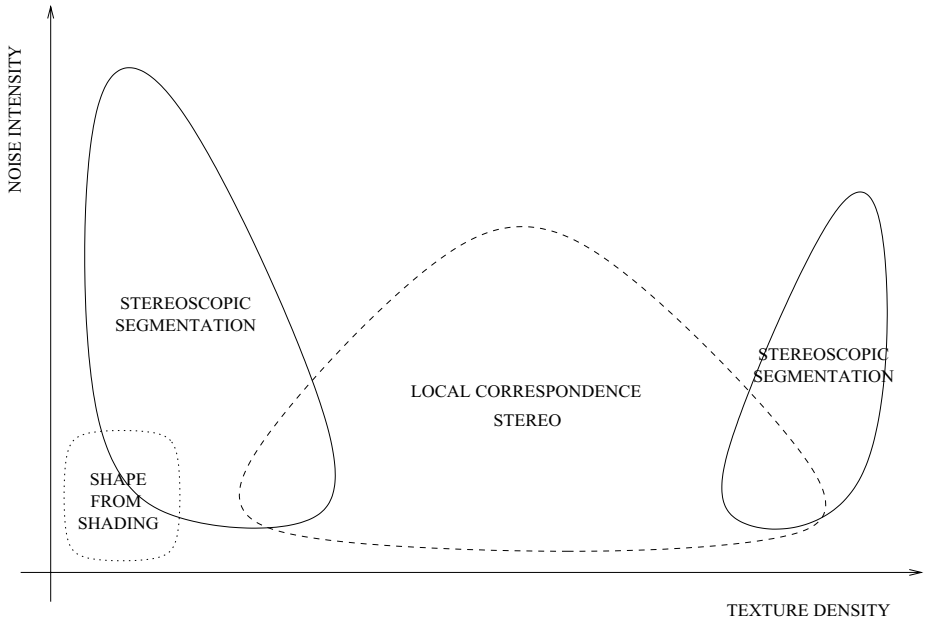

Figure 1. Domain of operation of shape from shading, stereo and stereoscopic segmentation in a noise-texture plane: the horizontal axis indicates increased density in the texture of the surfaces in the scene, the vertical axis indicates the intensity of noise in the image. Although our algorithm is designed for homogeneous image statistics, in practice it performs satisfactorily even in the presence of smooth irradiance or modest texture gradients.

radiance functions and illumination models that result in *homogeneous image statistics*.[3] This includes Lambertian objects with constant albedo (where local correspondence is ill-posed) as well as densely textured objects with isotropic statistics (where local correspondence is prone to multiple local minima). These conditions are admittedly restrictive. However, they correspond to conditions that prevent traditional stereo or shape from shading to operate. The relation between our method and traditional stereo and shape from shading is illustrated pictorially in Fig. 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.

1.1. Relation to Prior Work

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

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

In local correspondence-based stereo (see Faugeras, 1993) 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.[4] Faugeras and Keriven (1996) 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 (Faugeras and Keriven, 1996) to regions. In shape carving (Kutulakos and Seitz, 1998), 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 is smooth, to recover a different representation (the *smoothest* shape that is photometrically consistent with the data in a variational sense) as well as photometry.

The precise mathematical sense in which "smoothness" is measured is with respect to an energy functional which penalizes surface area of the modeled shape while at the same time encouraging a good fit between projections of the modeled shape (and its modeled radiance) onto the image data. 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 not be retrieved. In this sense, shape carving is *uni-directional*. Our algorithm, on the other hand, is bidirectional, in that surfaces are allowed to evolve inward or outward. This work also relates to shape from shading (Horn and Brooks, 1989) 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 (Rosenholtz and Malik, 1993) 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 (Cipolla and Blake, 1992; Laurentini, 1994), 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 (1989), and including (Caselles et al., 1993; Caselles et al., 1998; Cohen, 1991; Kass et al., 1987; Kichenassamy et al., 1996; LeVeque, 1992; Tek and Kimia, 1995; Terzopoulos and Witkin, 1988; Ronfard, 1994; Zhu et al., 1995; Zhu and Yuille, 1996). This line of work stands to complement local contour-based segmentation methods such as (Kass et al., 1987; Terzopoulos and Witkin, 1988). There are also algorithms that combine both features (Chakraborty and Duncan, 1999; Chakraborty et al., 1996).

In the methods used to perform the actual reconstruction, our work relates to the literature on level set methods (Osher and Sethian, 1988).

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 the scene is composed by a number of smooth surfaces supporting a Lambertian radiance function (or a dense texture with spatially homogeneous statistics) and illumination that guarantees that irradiance discontinuities (or texture discontinuities) correspond to occluding boundaries. These assumptions make the segmentation problem well-posed, although not general. In fact, "true" segmentation in this context corresponds in a one-to-one fashion to the shape of the objects in the scene.[5] 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.

2.1. Notation

In what follows we will indicate by $\mathbf{x} = (x, y, z)$ the coordinates of the generic point in $\mathbb{R}^3$, Ω the domain of the image I (with area element $d\Omega$) and $\pi : \mathbb{R}^3 \to \Omega; \mathbf{x} \mapsto \hat{\mathbf{x}}$ an ideal perspective projection with inverse (re-projection) $\pi^{-1} : \Omega \to \mathbb{R}^3; \hat{\mathbf{x}} = (\hat{x}, \hat{y})$ where $\hat{x} = x/z,\ \hat{y} = y/z$. Note that the projection map, restricted to the visible surfaces, is invertible.[6] S is a surface in space with coordinates (u, v) and area element dA. We describe the background ("blue sky") as a sphere with angular coordinates $\Theta = (\theta, \phi)$ supporting a radiance $\mathbf{g}$. The radiance of the surface S is $\mathbf{f}$, and we define the region $R = \pi(S) \subseteq \Omega$ which back-projects (via π^{-1}) onto the surface S and denote its complement, which back-projects (via Θ) onto the background, by R^c. The area measure $d\Omega = d\hat{x}d\hat{y}$ is related to the area measure $dA = \|S_u \times S_v\|\, du\, dv$ by $d\Omega = \frac{S \cdot N}{(S \cdot e_3)^3} dA$, where e_3 denotes the natural basis vector in the z direction and N is the inward [7] unit normal to S.

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, depends upon the surface S as well as upon the radiance of the surface $\mathbf{f}$ and of the background $\mathbf{g}$: $E = E(\mathbf{f}, \mathbf{g}, S)$. 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 $E_{smooth}(\mathbf{f}, \mathbf{g})$ and a geometric prior $E_{geom}(S)$. We consider the composite cost functional to be the sum of these three terms:

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

We conjecture that, like in the case of the Mumford-Shah functional (1989), these ingredients are sufficient to define a unique solution to the minimization problem.[8]

In particular, the geometric prior is given by $E_{geom} = \int_S dA$, while smoothness is imposed by a cost on the quadratic variation $E_{smooth} = \int_S \|\nabla \mathbf{f}\|^2 dA + \int_B \|\nabla \mathbf{g}\|^2 d\Theta$ where B denotes the background (blue sky). Finally, the data fidelity term may be measured in the sense of $\mathcal{L}^2$ by

$$\begin{aligned} E_{data} = &\int_R (\mathbf{f}(\pi^{-1}(\hat{\mathbf{x}})) - I(\hat{\mathbf{x}}))^2 d\Omega \\ &+ \int_{R^c} (\mathbf{g}(\Theta(\hat{\mathbf{x}})) - I(\hat{\mathbf{x}}))^2 d\Omega. \end{aligned} \tag{2}$$

Since an analytical expression for π^{-1} is not available, we would like to rewrite this integral so that π appears in the integrand instead of π^{-1}. Furthermore, in order to facilitate the computation of the first variation with respect to S, we want both integrals expressed as integrals over S as opposed to Ω. We begin with the first integral (over R). It is equal to

$$\begin{aligned} &\int_{\pi^{-1}(R)} (\mathbf{f}(\mathbf{x}) - I(\pi(\mathbf{x})))^2 \frac{\mathbf{x} \cdot N}{(\mathbf{x} \cdot e_3)^3} \, dA \\ &\quad = \int_{\pi^{-1}(R)} \epsilon^2(\mathbf{x}) \frac{\mathbf{x} \cdot N}{(\mathbf{x} \cdot e_3)^3} \, dA \end{aligned}$$

where $\epsilon(\mathbf{x}) = \mathbf{f}(\mathbf{x}) - I(\pi(\mathbf{x}))$. Now consider the second integral (over R^c). It equals

$$\begin{aligned} \int_{R^c} \varepsilon^2(\hat{\mathbf{x}}) \, d\Omega &= \int_\Omega \varepsilon^2(\hat{\mathbf{x}}) \, d\Omega - \int_R \varepsilon^2(\hat{\mathbf{x}}) \, d\Omega \\ &= \int_\Omega \varepsilon^2(\hat{\mathbf{x}}) \, d\Omega \\ &\quad - \int_{\pi^{-1}(R)} \varepsilon^2(\pi(\mathbf{x})) \frac{\mathbf{x} \cdot N}{(\mathbf{x} \cdot e_3)^3} \, dA \end{aligned}$$

where $\varepsilon(\hat{\mathbf{x}}) = \mathbf{g}(\Theta(\hat{\mathbf{x}})) - I(\hat{\mathbf{x}}))$. Combining these "rewritten" integrals yields:

$$\begin{aligned} &\int_\Omega \varepsilon^2(\hat{\mathbf{x}}) \, d\Omega + \int_{\pi^{-1}(R)} (\epsilon^2(\mathbf{x}) - \varepsilon^2(\pi(\mathbf{x}))) \frac{\mathbf{x} \cdot N}{(\mathbf{x} \cdot e_3)^3} \, dA \\ &\quad = \int_\Omega \varepsilon^2(\hat{\mathbf{x}}) \, d\Omega + \int_{\pi^{-1}(R)} \Upsilon(\mathbf{x}) \sigma(\mathbf{x}, N) \, dA \end{aligned}$$

where $\Upsilon(\mathbf{x}) = \epsilon^2(\mathbf{x}) - \varepsilon^2(\pi(\mathbf{x}))$ and $\sigma(\mathbf{x}, N) = \frac{\mathbf{x} \cdot N}{(\mathbf{x} \cdot e_3)^3}$. Note that the first integral in the above is independent of the surface S (and its radiance function $\mathbf{f}$) and that the second integral is taken over only a subset of the surface. We may rewrite this integral as one over the entire surface S and remove the last "remnant" of π^{-1} by introducing a characteristic function $\chi(\mathbf{x}) \in \{0, 1\}$ into the integrand where $\chi(\mathbf{x}) = 1$ for points on S that are visible and $\chi(\mathbf{x}) = 0$ otherwise:

$$E_{data} = \int_\Omega \varepsilon^2(\hat{\mathbf{x}}) \, d\Omega + \int_S \chi(\mathbf{x}) \Upsilon(\mathbf{x}) \sigma(\mathbf{x}, N) \, dA \tag{3}$$

The global cost functional is obtained by adding the smoothness prior and geometric prior as follows

$$\begin{aligned} E(\mathbf{f}, \mathbf{g}, S) = &\int_\Omega \varepsilon^2(\hat{\mathbf{x}}) \, d\Omega + \int_S \chi(\mathbf{x}) \Upsilon(\mathbf{x}) \sigma(\mathbf{x}, N) \, dA \\ &+ \int_S dA + \int_S \|\nabla \mathbf{f}\|^2 dA + \int_B \|\nabla \mathbf{g}\|^2 d\Theta \end{aligned} \tag{4}$$

2.3. Evolution Equation

In order to find the surface S and the radiances $\mathbf{f}, \mathbf{g}$ that minimize the functional (1) we set up an iterative procedure where we start from an initial radiance $\mathbf{f}, \mathbf{g}$ and update S by a gradient flow equation based on the first variation with respect to S. While the variation of E_{smooth} does not, to a first approximation, [9] depend on

S, the variation of E_{geom} is given by κN. The variation of E_{data} requires some attention. In fact, the data fidelity term in (3) involves an explicit model of occlusions[10] 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 integrand. This can be done in a mathematically sound way by interpolating a smooth force field on the surface in space. Or, alternatively, the integrand in data fidelity term can be mollified thereby removing the need to extend cost functional onto the surface. We discuss this issue in Appendix A.

In order to arrive to an evolution equation, let us first consider the data fidelity term in Eq. (3), which we write in the following concise form

$$\int_{\mathcal{E}} G(X) \cdot N \, dA. \tag{5}$$

It can be shown that zeroing the first variation leads to the following evolution equation

$$\frac{dS}{dt} = -(\nabla \cdot G)N \tag{6}$$

In (3) we are considering a particular case of (5) for which

$$G(X) = \chi(X)\frac{\Upsilon(X)}{(X \cdot e_3)^3}X \tag{7}$$

We now compute the divergence of G:

$$\nabla \cdot G = \frac{1}{z^3}(\chi \nabla \Upsilon + \Upsilon \nabla \chi) \cdot X \tag{8}$$

and substitute $-\Upsilon = (\mathbf{f} - \mathbf{g})[(I - \mathbf{f}) + (I - \mathbf{g})]$ and $-\nabla\Upsilon = 2[(\mathbf{f} - \mathbf{g})\nabla I + (I - \mathbf{f})\nabla\mathbf{f} - (I - \mathbf{g})\nabla\mathbf{g}]$ which leads, after simplifications, to the following gradient flow:

$$\begin{aligned}\frac{dS}{dt} = \frac{1}{z^3}((\mathbf{f} - \mathbf{g})[(I - \mathbf{f}) + (I - \mathbf{g})](\nabla\chi \cdot S) \\ + 2\chi(I - \mathbf{f})(\nabla\mathbf{f} \cdot S))N + \kappa N.\end{aligned} \tag{9}$$

Notice that this flow depends only upon the image values, *not the image gradient*, which makes it more robust to image noise when compared to other variational approaches to stereo (i.e. less prone to become "trapped" in local minima).

2.4. Estimating Scene Radiance

Once an estimate of the surface S is available, the radiances $\mathbf{f}$, $\mathbf{g}$ must be updated. This is immediate for the case of piecewise constant images (images with piecewise constant irradiance statistics), since $\nabla\mathbf{f} = 0$ and $\nabla\mathbf{g} = 0$ and therefore the best radiances are given, from Eq. (2), by the sample average of I in the regions R and R^c. This is what we do in the experimental section that follows. In this respect, our segmentation scheme reflects the piecewise constant model used by Chan and Vese (1999). However, rather than utilizing individual active contours to segment each image directly as in Chan and Vese (1999), we couple the segmentations of each image through the evolution of a single 3D surface rather than separate 2D active contours.

If the radiances are smooth, one can set up an optimization problem for $\mathbf{f}$ on the manifold S and $\mathbf{g}$ on the manifold B; this is solved by a Poisson equation on the manifold, where the Laplace-Beltrami operator takes the place of the standard Laplacian. Since the images we tested were well approximated by piecewise constant statistics, we do not pursue this approach further at this stage. Implementation details of the full piecewise smooth case may be found in Jin et al. (2003).

3. Experiments

In Fig. 2 we show 4 of 22 calibrated views of a scene meant to illustrate the domain of applicability of our algorithm. The scene contains three objects: two shakers and the background. 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 Fig. 3 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 the parts of the initial surface evolve outwards, while parts evolve inwards in order to converge to the final shape. This bi-directionality is a feature of our algorithm. In Fig. 4 we show the final result from various vantage points. In Fig. 5 we show the final segmentation in some of the original views (Top). We also show the segmented foreground superimposed to the original images. Two

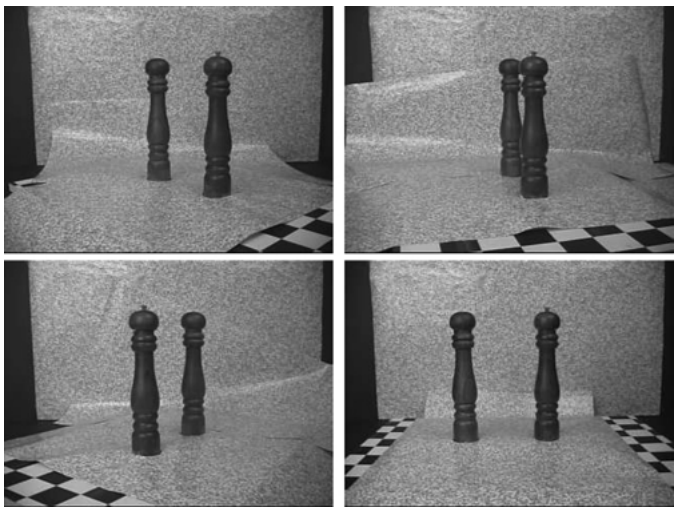

Figure 2. Original "salt and pepper" sequence (4 of 22 views).

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 Fig. 6 we show the results of first segmenting each of the camera images separately and independently and then intersecting the projected cones of these segmentations in space to "carve out" an estimate of the shape. This procedure is suboptimal for a number of reasons. First, the smoothness constraints are placed on the boundaries of the segmenting curves within each image, rather than on the reconstructed shape itself. As clearly seen in the figure, smooth projections of a shape do not guarantee a smooth reconstruction of the shape itself when such carving procedures are employed (unless an infinite number of projections are available). Second, this approach is extremely sensitive to calibration errors. Many of the images in this experiment contained small calibration errors which cumulatively cause a noticeable portion of the true shape to be "eaten away" near its boundary under this

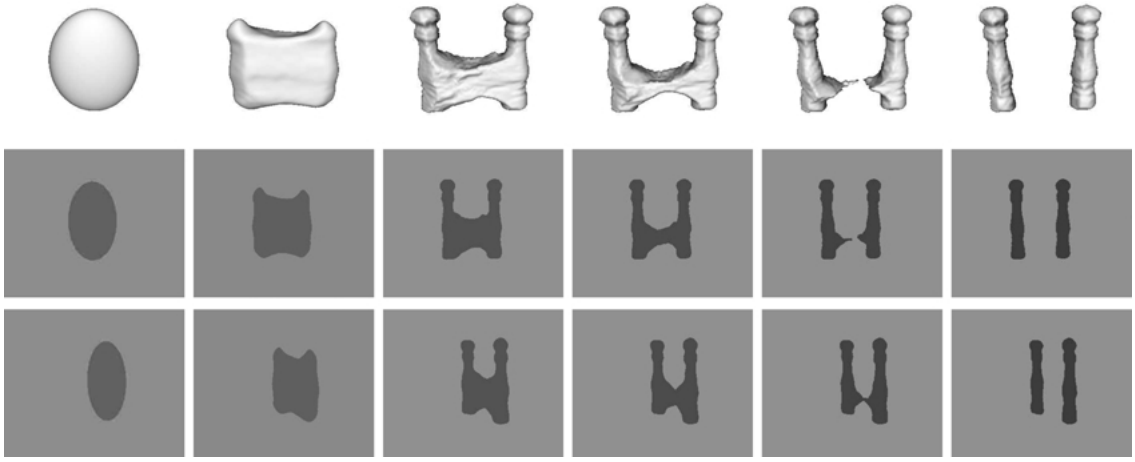

Figure 3. (Top) Rendered surface during evolution (6 of 800 steps). Notice that the initial surface is neither contained nor contains the actual scene. (Bottom) segmented image during the evolution from two different viewpoints.

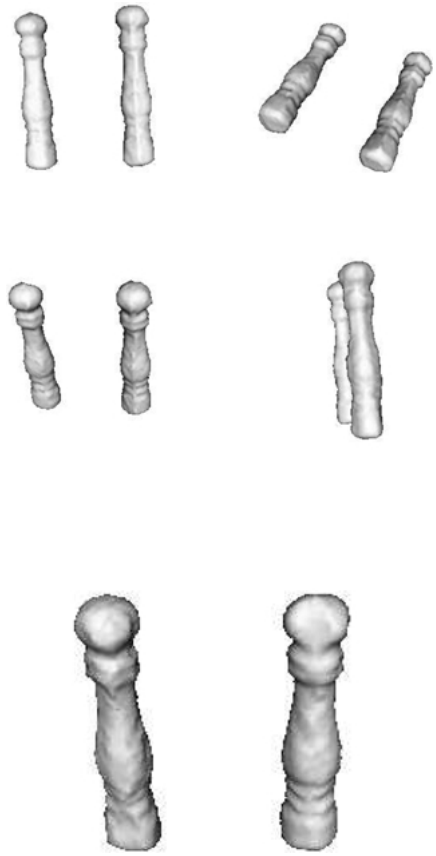

Figure 4. Final estimated surface, seen from several viewpoints. Notice that the bottoms of the salt and pepper shakers are flat, even though no data was available. This is due to the geometric prior, which in the absence of data results in a minimal surface being computed.

carving procedure. Thus, even if the portions of the reconstruction that appear "reasonable", the fine details captured by our more integrated and coupled model are completely missed. More severe effects however are seen from two significantly miscalibrated images which cause very large "chunks" of the true shape to be removed from the reconstruction. Clearly, coupling the segmentation of each image by directly estimating the object we are trying to reconstruct yields better results.

In order to quantify the improvement of our technique over volume carving techniques, we have computed the reprojection error between the two models. In each case we have endowed the reconstructed shape and the background with two different radiances equal to the respective overall means of the segmented foreground and background image intensities. The reprojection error is computed as the root mean square brightness error, normalized with respect to the image size and to the average brightness of the 22 original images. The final reprojection error using the carving procedure was $9.849 \times 10^{-2}\%$ per pixel, whereas the reconstruction using our algorithm yielded an error of $8.809 \times 10^{-2}\%$) obtained using our variational approach. In our algorithm, most of the error is concentrated on the two miscalibrated images, whereas most views contain significant reprojection errors in the carving results. More than the reprojection error, however, the improvement of our algorithm over carving techniques is visible in the reconstruction (compare Fig. 4 with Fig. 6).

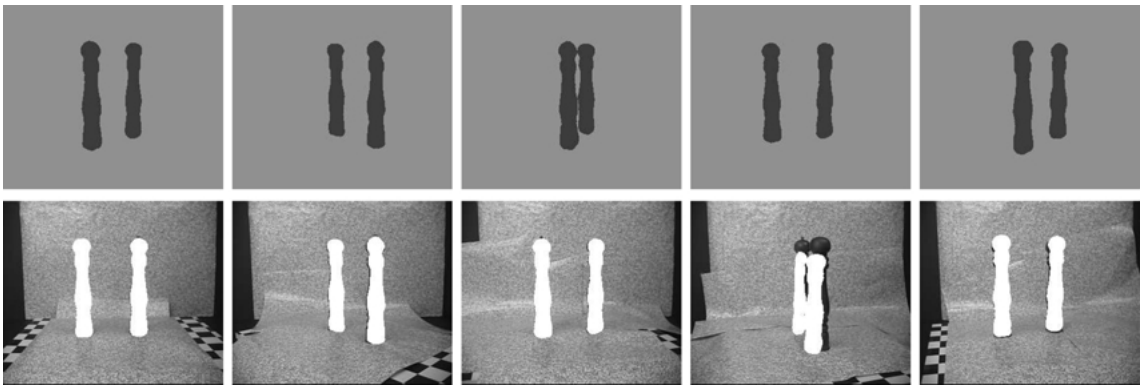

Figure 5. (Top) image segmentation for the salt and pepper sequence. (Bottom) Segmented foreground superimposed to the original sequence. The calibration in two of the 22 images was dramatically wrong. However, the effect is mitigated by the global integration, and the overall shape is only marginally affected by the calibration errors.

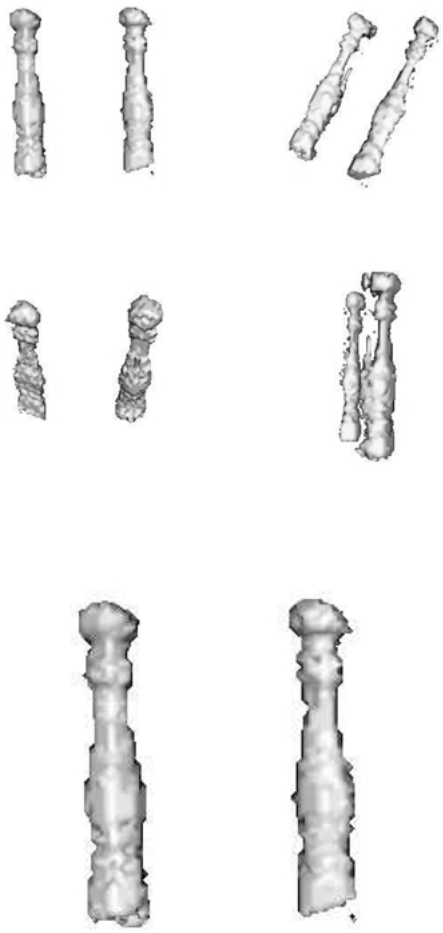

Figure 6. Reconstruction obtained by carving techniques. Individual images are first segmented, and then visibility cones are carved to yield the reconstruction seen above. Even though individual images are segmented successfully, calibration errors cause uni-directional carving techniques to miss important details and to cut holes into the reconstruction.

In Fig. 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 Fig. 8. The results shown were obtained using a C++ implementation running on a 700 MHz laptop. For 22 640 $\times$ 480 images and a cubic grid of 128 $\times$ 128 $\times$ 128 the algorithm takes about 20 minutes to converge (tested by threshold on the iteration residual).

4. Discussion

We have presented an algorithm to reconstruct solid shape and radiance from a number of calibrated views of a scene with smooth shape and homogeneous radiance statistics. It enjoys significant resistance to noise since it does not rely on local correspondence, and its domain of operation complements existing stereo reconstruction algorithms. In particular, our assumptions demand that the scene radiance results in piecewise homogeneous image statistics (e.g. the response of a filter bank). While restrictive, the algorithm is robust enough to operate despite gross violations of the assumptions. For instance, we have shown experiments where the image irradiance had some specularity and regions were well separated, but not strictly constant. Our experiments show that the algorithm can tolerate errors in camera calibration. The computational complexity depends on the size of the grid where the PDE evolves, and not on the complexity of the scene. Therefore, although the algorithm is far from operating in real time, it can infer scenes with complex geometry and topology effortlessly.

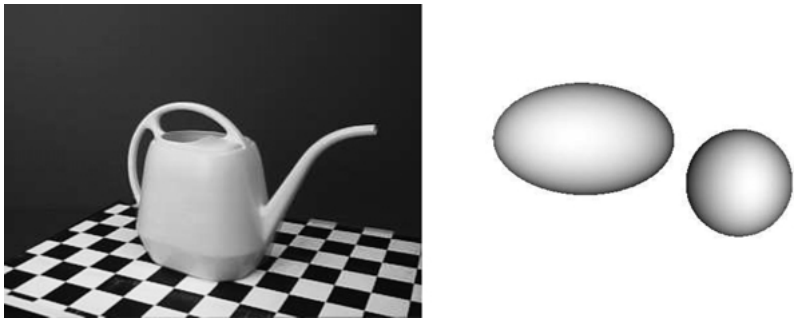

Figure 7. The "watering can" sequence and the initial surface. Notice that the initial surface is not simply connected and does not include and is not included by the shape. In order to capture a hole it is necessary that it intersects the initial surface. One way to guarantee this is to start with a number of small surfaces.

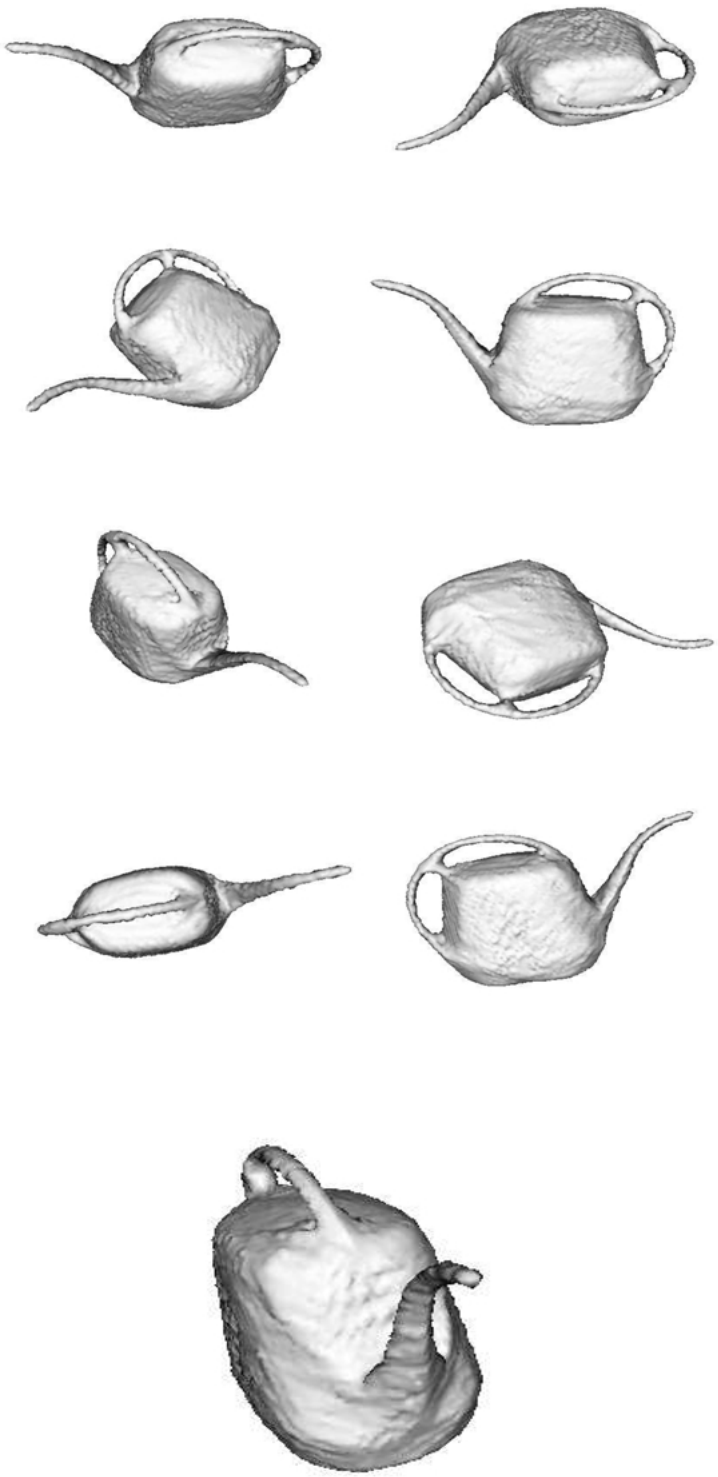

Figure 8. Final estimated shape for the watering can. The two initial surfaces have merged. and the topology and geometry of the watering can has been correctly captured. The evolution is shown in Fig. 9.

Figure 9. (Top) Rendered surface during evolution for the watering can.

Appendix 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 S. We may obtain an expression valid everywhere if we extend the vector field of unit inward normals N from the surface S to the surrounding embedding space. Let us call this extension $\tilde{N}$. We will soon see that the particular choice of extension will not affect the final gradient flow. We now write

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

and express the gradient as

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

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

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

To perform the computation of $\tilde{N}_X^T S$, 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. $S_u \times S_v = N$, $S_v \times N = S_u$, $N \times S_u = S_v$ ($E = 1$, $F = 0$, $G = 1$)
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 $\tilde{N}_X$ by noting that $\tilde{N}_U = \tilde{N}_X X_U$ and so

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

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

$$\begin{aligned}\tilde{N}_X &= [N_u \quad N_v \quad \tilde{N}_w] \\ &\quad \times \frac{[S_v \times N \quad N \times S_u \quad S_u \times S_v]^T}{(S_u \times S_v) \cdot N} \\ \tilde{N}_X^T S_0 &= \frac{[S_u \quad S_v \quad N]}{N \cdot N} \begin{bmatrix} N_u \cdot S_0 \\ N_v \cdot S_0 \\ \tilde{N}_w \cdot S_0 \end{bmatrix} \\ &= (N_u \cdot S_0)S_u + (N_v \cdot S_0)S_v + (\tilde{N}_w \cdot S_0)N.\end{aligned}$$

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

$$\begin{aligned}\left(\tilde{N}_X^T S_0\right) \cdot S_0 &= \|S_0\|(N_u \cdot S_0) = \|S_0\|^2(N_u \cdot S_u) \\ &= -\|S_0\|^2(N \cdot S_{uu}) = -\kappa_u\|S_0\|^2\end{aligned}$$

yielding

$$\nabla\chi \cdot S = -\kappa_u\|S\|^2\delta(S \cdot N) \tag{10}$$

where κ_u denotes the normal curvature of S in the u-direction (the direction S when $S \cdot N = 0$).

If we approximate the surface radiance by a constant $\mathbf{f}$ then $\nabla\mathbf{f} = 0$, causing the second term in (9) to vanish.

$$\begin{aligned}\frac{dS}{dt} &= \frac{1}{z^3}(\mathbf{f}-\mathbf{g})[(I-\mathbf{f})+(I-\mathbf{g})](\nabla\chi\cdot S)N \\ &= -\frac{\kappa_u\|S\|^2}{z^3}(\mathbf{f}-\mathbf{g})[(I-\mathbf{f})+(I-\mathbf{g})]\delta(S\cdot N)N\end{aligned}$$

Appendix B: Projecting ϵ-Neighborhoods of the Tangent Plane

Given a point $X_0 \in S$ and a normal vector N at that point (does not have to be unit normal and can be either inward or outward) we may express the tangent plane to X_0 through the following equation,

$$N \cdot X = N \cdot X_0 \tag{11}$$

where $X = (x, y, z) \in \mathbb{R}^3$ and $(x_0, y_0, z_0) = X_0$. We assume that a point X projects to a point $(u, v) = \pi(X)$ in a given image as follows (where L_u and L_v denote separate focal lengths for the u and v axes of the image).

$$u = L_u\frac{x}{z} \quad \text{and} \quad v = L_v\frac{y}{z} \tag{12}$$

The tangent plane may be parameterized via the image coordinates (u, v) by combining (11) and (12).

$$N\left(\frac{zu}{L_u}, \frac{zv}{L_v}, z\right) = N\cdot X_0 \longrightarrow z = \frac{N\cdot X_0}{N\left(\frac{u}{L_u}, \frac{v}{L_v}, 1\right)}$$

$$\longrightarrow X(u,v) = \frac{N\cdot X_0}{N\left(\frac{u}{L_u}, \frac{v}{L_v}, 1\right)}\left(\frac{u}{L_u}, \frac{v}{L_v}, 1\right)$$

If we denote $\pi(X_0)$ by (u_0, v_0), i.e. $X(u_0, v_0) = X_0$, then we may write

$$\left(\frac{u_0}{L_u}, \frac{v_0}{L_v}, 1\right) = \frac{X_0}{z_0} \longrightarrow \left(\frac{u}{L_u}, \frac{v}{L_v}, 1\right) = \frac{X_0}{z_0} + \Delta$$

where $\Delta = (\frac{u-u_0}{L_u}, \frac{v-v_0}{L_v}, 0)$ which, when plugged into (13), gives the following expression for the tangent plane in terms of Δ.

$$X(u,v) = \frac{N\cdot X_0}{N(X_0 + z_0\Delta)}(X_0 + z_0\Delta) \tag{13}$$

Now, based upon this local approximation of S by its tangent plane through X_0, we may test whether an image pixel (u, v) in a neighborhood of $\pi(X_0) = (u_0, v_0)$ back-projects onto S within an ϵ-neighborhood of X_0 (which will be used to approximate a discrete grid voxel) by checking if $\|X(u, v) - X_0\| < \epsilon$. To do so, we will make use of the following expression.

$$\begin{aligned}\|X(u,v)-X_0\|^2 &= X(u,v)\cdot X(u,v) - 2X_0\cdot X(u,v) + X_0\cdot X_0 \\ &= \frac{(N\cdot X_0)^2}{(N\cdot X_0 + z_0 N\cdot\Delta)^2}\left(X_0\cdot X_0 + 2z_0X_0\cdot\Delta + z_0^2\Delta\cdot\Delta\right) \\ &\quad - \frac{2N\cdot X_0}{N\cdot X_0 + z_0N\cdot\Delta}(X_0\cdot X_0 + z_0X_0\cdot\Delta) + X_0\cdot X_0 \\ &= z_0^2\frac{(N\cdot X_0)^2(\Delta\cdot\Delta) - 2(N\cdot X_0)(N\cdot\Delta)(X_0\cdot\Delta) + (N\cdot\Delta)^2(X_0\cdot X_0)}{(N\cdot X_0 + z_0N\cdot\Delta)^2}.\end{aligned}$$

Acknowledgments

This research is supported by NSF IIS-0208197 and CCR-0133736, ONR N00014-02-1-0720, ARO DAAD19-99-1-0139 and Intel 8029. The authors wish to thank Hailin Jin for his assistance with the experiments.

Notes

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 (1996).
2. Local methods involve computing derivatives, and are therefore sensitive to noise. Region-based methods involve computing integrals, and suffer noise less.
3. The term statistics is used here as an arbitrary function of an image realization. For instance, the image itself, its gradient, or the output of a bank of filters are image statistics.

4. An anonymous reviewer pointed out that (Szeliski and Scharstein, 2002), which appeared after this manuscript was submitted for review, is also representative of this line of work.
5. We consider the background to be yet another object that happens to occupy the entire field of view (the "blue sky" assumption).
6. Naive visibility computation is straightforward. For an efficient computation of visibility in the level set framework, the reader can refer to Tsai (2002).
7. While there is no explicit assumption of an orientable surface in our mathematical model, we have assigned an orientation to N so that the corresponding level set implementations of the resulting flows may be written down immediately. Level set techniques, which we have used in our experiments, require an orientable surface even though our cost functional does not.
8. An anonymous reviewer has pointed out that it is dangerous to conjecture uniqueness for functionals of this type.
9. Technically, the variation E_{smooth} does depend upon the surface S since it affects the intrinsic gradient operator ∇_S. We compute this variation and discuss its behavior in detail in a technical report (Jin et al., 2003). Our experiments have shown, however, that it is not critical in the final reconstruction to model the explicit dependence of E_{smooth} during the gradient descent procedure since smoothness in the surface itself is more directly modeled through the term E_{geom}. The more important variation of E_{smooth} is with respect to the radiance function $\mathbf{f}$ since only this term affects its smoothness.
10. The geometric term and the smoothness constraints are independent of occlusions.

References

Blake, A. and Yuille, A. 1992. *Active Vision*. MIT Press: Cambridge, MA.

Caselles, V., Catte, F., Coll, T., and Dibos, F. 1993. A geometric model for active contours in image processing, *Numerische Mathematik*, 66:1–31.

Caselles, V., Kimmel, R., and Sapiro, G. 1997. Geodesic snakes, *Int. J. Computer Vision*, 22(1):61–79.

Chakraborty, A. and Duncan, J. 1999. Game-theoretic integration for image segmentation, *IEEE Trans. Pattern Anal. Machine Intell.*, 21(1):12–30.

Chakraborty, A., Staib, L., and Duncan, J. 1996. Deformable boundary finding in medical images by integrating gradient and region information, *IEEE Trans. Medical Imaging*, 15(6):859–870.

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

Cipolla, R. and Blake, A. 1992. Surface shape from the deformation of apparent contours. *Int. J. of Computer Vision*, 9(2).

Cohen, L. 1991. On active contour models and balloons, *CVGIP: Image Understanding*, 53:211–218.

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

Faugeras, O. 1993. *Three Dimensional Vision, a Geometric Viewpoint.* MIT Press.

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

Fleming, W. and Soner, H. 1993. *Controlled Markov Processes and Viscosity Solutions*. Springer-Verlag: New York.

Horn, B. and Brooks, M. (Eds.). 1989. *Shape from Shading*. MIT Press.

Jin, H., Yezzi, A., Tsai, R., Cheng, L., and Soatto, S. 2003. Estimation of 3D surface shape and smooth radiance from 2D images: A level set approach, *J. of Comp. Physics* (in press).

Kass, M., Witkin, A., and Terzopoulos, D. 1987. Snakes: Active contour models, *Int. Journal of Computer Vision*, 1:321–331.

Kichenassamy, S., Kumar, A., Olver, P., Tannenbaum, A., and Yezzi, A. 1996. Conformal curvature flows: From phase transitions to active vision, *Arch. Rational Mech. Anal.*, 134:275–301.

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

Laurentini, A. 1994. The visual hull concept for silhouette-based image understanding. *PAMI* (16)2:150–162.

Leclerc, Y. 1989. Constructing stable descriptions for image partitioning, *Int. J. Computer Vision*, 3:73–102.

LeVeque, R.J. 1992. *Numerical Methods for Conservation Laws*. Birkhäuser: Boston.

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

Malladi, R., Sethian, J., and Vemuri, B. 1995. Shape modeling with front propagation: A level set approach, *IEEE Trans. Pattern Anal. Machine Intell.*, 17:158–175.

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

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

Osher, S. 1984. Riemann solvers, the entropy condition, and difference approximations, *SIAM J. Numer. Anal.*, 21:217–235.

Osher, S. and Sethian, J. 1988. Fronts propagating with curvature-dependent speed: Algorithms based on Hamilton-Jacobi equations. *J. of Comp. Physics*, 79:12–49.

Paragios, N. and Deriche, R. 1999. Geodesic active regions for supervised texture segmentation, In *Proceedings of ICCV*, Corfu, Greece.

Paragios, N. and Deriche, R. 2000. Coupled geodesic active regions for image segmentation: A level set approach, In *Proceedings of ECCV*, Dublin, Ireland.

Ronfard, R. 1994. Region-based strategies for active contour models, *Int. J. Computer Vision*, 13(2):229–251.

Rosenholtz, R. and Malik, J. 1993. 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.

Samson, C., Blanc-Feraud, L., Aubert, G., and Zerubia, J. 1999. A level set method for image classification, In *Int. Conf. Scale-Space Theories in Computer Vision*, pp. 306–317.

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

Siddiqi, K., Lauziere, Y., Tannenbaum, A., and Zucker, S. 1998. Area and length minimizing flows for segmentation, *IEEE Trans. Image Processing*, 7:433–444.

Szeliski, R. and Scharstein, D. 2002. Symmetric subpixel stereo matching. In *Proc. of the Seventh European Conference on Computer Vision*, Copenhagen, Denmark.

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

Tsai, R., Burchard, P., Cheng, L., Osher, S., and Sapiro, G. 2002. Dynamic visibility in a level set based implicit frame work, CAM Technical Report 02-06.

Terzopoulos, D. and Witkin, A. 1988. Constraints on deformable models: Recovering shape and non-rigid motion, *Artificial Intelligence*, 36:91–123.

Yezzi, A., Tsai, A., and Willsky, A. 1999. A statistical approach to image segmentation for bimodal and trimodal imagery, In *Proceedings of ICCV*.

Zhu, S., Lee, T., and Yuille, A. 1995. Region competition: Unifying snakes, region growing, and bayes/MDL for multiband image segmentation, In *Proc. of ICCV*, pp. 416–423.

Zhu, S. and Yuille, A. 1996. Region competition: Unifying snakes, region growing, and bayes/MDL for multiband image segmentation, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18(9):884–900.