Curve Evolution Implementation of the Mumford–Shah Functional for Image Segmentation, Denoising, Interpolation, and Magnification

Andy Tsai, Anthony Yezzi, Jr., *Member, IEEE*, and Alan S. Willsky, *Fellow, IEEE*

***Abstract*—In this work, we first address the problem of simultaneous image segmentation and smoothing by approaching the Mumford–Shah paradigm from a curve evolution perspective. In particular, we let a set of deformable contours define the boundaries between regions in an image where we model the data via piecewise smooth functions and employ a gradient flow to evolve these contours. Each gradient step involves solving an optimal estimation problem for the data within each region, connecting curve evolution and the Mumford–Shah functional with the theory of boundary-value stochastic processes. The resulting active contour model offers a tractable implementation of the original Mumford–Shah model (i.e., without resorting to elliptic approximations which have traditionally been favored for greater ease in implementation) to simultaneously segment and smoothly reconstruct the data within a given image in a coupled manner. Various implementations of this algorithm are introduced to increase its speed of convergence. We also outline a hierarchical implementation of this algorithm to handle important image features such as triple points and other multiple junctions.**

Next, by generalizing the data fidelity term of the original Mumford–Shah functional to incorporate a spatially varying penalty, we extend our method to problems in which data quality varies across the image and to images in which sets of pixel measurements are missing. This more general model leads us to a novel PDE-based approach for simultaneous image magnification, segmentation, and smoothing, thereby extending the traditional applications of the Mumford–Shah functional which only considers simultaneous segmentation and smoothing.

***Index Terms*—Active contours, boundary-value stochastic processes, curve evolution, denoising, image interpolation, image magnification, level sets methods, missing data problems, Mumford–Shah functional, reconstruction, segmentation, snakes.**

I. Introduction

TWO popular applications of partial differential equations in computer vision and image processing are found in the problems of segmentation and image smoothing. For segmentation, the technique of *snakes* or *active contours* has grown significantly since the seminal work of Kass, Witkin, and Terzopoulos [14] including the development of geometric models based on curve evolution theory [6], [7], [19], [47] and the progression from edge-based models [6], [7], [12], [14], [15], [19], [41], [42], [47] to region-based models [9], [27], [34], [48], [51], [52]. For image smoothing, the technique of anisotropic diffusion has become a widespread field of research ranging from techniques based upon the original formulation of Perona and Malik [29], [30] to curve and surface evolution methods based upon geometric scale spaces [13], [16], [17], [35] and to a number of recent techniques for color imagery and other forms of vector-valued data [39], [40], [44]–[46], [50].

In general, the goal of most active contour algorithms is to extract the boundaries of homogeneous regions within an image, while the goal of most anisotropic diffusion algorithms is to smooth the values of an image within homogeneous regions but not across the boundaries of such regions. We note that one of the most widely studied mathematical models in image processing and computer vision addresses both goals simultaneously, namely that of Mumford and Shah [22], [23] who presented the variational problem of minimizing a functional involving a piecewise smooth representation of an image. Their functional included a geometric term which penalized the Hausdorff measure of the set where discontinuities in the piecewise smooth estimate would be allowed. Due to the difficulties associated with implementing such a term in a numerical algorithm, one of the first practical numerical implementations of the Mumford–Shah model was developed by Richardson [32] and was not based upon the original functional but was based instead upon a elliptic approximation of the functional considered by Ambrosio and Tortorelli [3]. In this elliptic approximation of the model, the exact location of boundaries between modeled homogeneous regions was "smeared" into a set with nonzero Lebesque measure, allowing the Hausdorff term to be eliminated. The recent work by Shah [37] uses the modified boundary indicator from this relaxed model as a conformal factor in a geodesic snake model, allowing the resulting algorithm to yield exact boundary locations.

In the first part of this paper, we present a curve evolution approach to minimizing the *original* Mumford–Shah functional, thereby obtaining an algorithm for simultaneous image smoothing and segmentation.[1] In contrast to anisotropic diffusion algorithms, however, the smoothing is linear, with edge preservation based upon a global segmentation as opposed to

Manuscript received April 4, 2000; revised March 29, 2001. This work was supported by ONR Grant N00014-00-1-0089 and by AFOSR Grant F49620-00-1-0362. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Timothy J. Schulz.

A. Tsai and A. S. Willsky are with the Laboratory for Information and Decision Systems, Massachusetts Institute of Technology, Cambridge, MA 02139 USA (e-mail: atsai@mit.edu).

A. Yezzi, Jr. is with the School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, GA 30332 USA.

Publisher Item Identifier S 1057-7149(01)06033-X.

[1]A preliminary conference paper based on this work can be found in [43]. Ideas very similar to the work described here is a generalization of the work in [9]. Later in this introduction we provide a few words discussing similarities and differences with the work of Chan and Vese.

local measurements based upon the gradient. The development of this model is based upon both estimation-theoretic and geometric considerations. In particular, by viewing an active contour as the set of discontinuities considered in the original Mumford–Shah formulation, we may use the corresponding gradient flow equation to evolve the active contour. However, each gradient step involves solving an optimal estimation problem to determine piecewise smooth approximations of the image data inside and outside the active contour. We obtain these estimates by solving a linear partial differential equation (PDE) for which the solution inside the active contour is decoupled from the solution outside the active contour. This PDE, which takes the form of a Poisson equation, and the associated boundary conditions come directly from the variational problem of minimizing the Mumford–Shah functional assuming the set of discontinuities (given by our active contour) to be fixed. The same PDE and boundary conditions can also be obtained from the theory of boundary-value stochastic processes. By taking this latter approach, we obtain an algorithm that may be regarded as a curve evolution driven by a continuum of solutions to auxiliary spatial estimation problems, connecting the theories of curve evolution and optimal estimation of stochastic processes. This development may be regarded as an extension of several recent region-based approaches to curve evolution [9], [27], [48]. In particular, it naturally generalizes the recent work of Chan and Vese in [9] who consider piecewise constant generalization of the Mumford–Shah functional within a level set framework.[2]

We note that region-based approaches in general, enjoy a number of attractive properties including greater robustness to noise (by avoiding derivatives of the image intensity) and initial contour placement (by being less local than most edge-based approaches). In contrast to most other region based techniques however (including our own previous work [48], [49] and that of Chan-Vese [9] and Paragios-Deriche [27]), which assume highly constrained parametric models for pixel intensities within each region, our approach employs the statistical model directly implied by the Mumford–Shah functional. That is, the image is modeled as a random field within each region, a model that naturally accommodates variability across each region without the need to model such variability parametrically. In addition, while many region-based methods require a priori knowledge of the number of region types (such as [9] which assumes exactly two region types with two different mean intensities or [48] which requires separate sets of curves to deal with more than two region types), our Mumford–Shah based approach can automatically segment images with multiple region types (e.g., each with different mean intensities) without such a priori knowledge.

In the second part of this paper, we generalize the data fidelity term of the original Mumford–Shah energy functional by substituting a spatially varying penalty for the traditional constant one. This allows us to treat images in which the quality of the measurements vary depending upon location in the image. In particular, we are able to treat, as a limiting case, images containing sets of pixels without measurements. This "missing data" problem arises regularly in archived or high speed motion picture films, damaged paintings, and remote sensing and medical images with data dropouts due to speckle and sensor data gaps. By applying this missing data technique in a structured manner, we then develop a novel approach for simultaneous image magnification, segmentation, and smoothing, thereby providing a new application of the Mumford–Shah functional. This technique constitutes a more global approach to interpolating magnified data than traditional bilinear or bicubic interpolation schemes, while still maintaining sharp transitions along region boundaries. Furthermore, the curve length penalty in our Mumford–Shah based flow tends to prevent the blocky appearance of object boundaries which is a symptom of replication-based schemes. In Fig. 1, we illustrate these points by applying our magnification technique to a noisy synthetic image [Fig. 1(a)]. We show in Fig. 1(b) the blocky and noisy magnification of the original image based on zero-order interpolation. In Fig. 1(c), we show a smoother but blurry magnification of the original image based on bilinear interpolation scheme. For comparison, in Fig. 1(d), we show the magnified image based on our Mumford–Shah approach. The smooth boundaries of the magnified image are the direct result of the minimum length prior placed on the segmenting curve while the smooth sinusoidal background of the observed image is successfully captured by our PDE-based model for interpolation. It is evident that our magnification approach is better than these conventional image magnification techniques by avoiding many of the processing artifacts such as blockiness and blurring while at the same time, denoising the image.

In our work, we adopt the level set techniques of Osher and Sethian [26], [36] in the implementation of our Mumford–Shah active contour model. This numerical implementation technique, in conjunction with upwind, conservative, monotone difference schemes [19], [25], [36], allows for automatic handling of cusps, corners, and topological changes as the curves evolve.

As mentioned previously, Chan and Vese [10], [11] have recently and independently performed work very similar to ours. The basic model and formulation outlined in their work and our work are essentially the same. The differences lie in the algorithmic implementations and extensions developed in this work and their work. In [10], [11], multiphase level set techniques are implemented to capture triple points and holes in a very elegant manner. This allows one to capture up to 2^n region classes using n level set functions. In this work, we employ more standard level techniques and propose, instead, a hierarchical method to capture multiple regions and triple points (see Section III-D). An additional point of divergence comes from our extension of the basic model to images involving missing data and to simultaneous magnification, smoothing, and segmentation. We should point out, however, that our treatment of missing data shares a similar flavor with some other recent work by Chan and Shen on image inpainting [8].

This paper is organized as follows. In Section II, we present the Mumford–Shah model, its interpretation as an estimation problem, and the basic formulation of our curve evolution approach for simultaneous image segmentation and smoothing. In Section III, we present several progressive enhancements and

[2]The formulation in [9] can be viewed as the limiting form of (1) with $\alpha = \infty$.

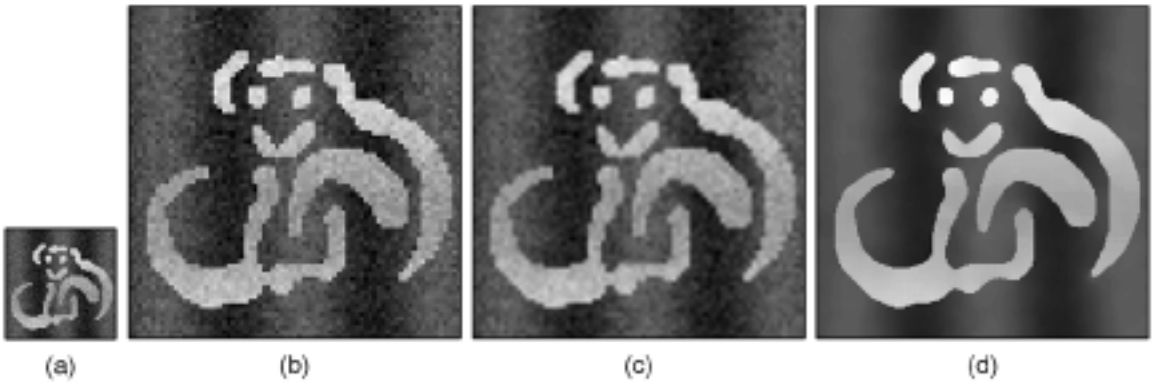

Fig. 1. Three-fold magnification of synthetic data: (a) 75×75 original image; (b) 225×225 magnified image interpolated by zero-order hold; (c) 225×225 magnified image obtained by bilinear interpolation; (d) 225×225 magnified image based on the Mumford–Shah model.

modifications to this initial algorithm to improve convergence and to handle multiple junctions. In Section IV, we extend the application of our model to one that can also handle images with missing data. The application of our model as an image magnification technique is then presented as a structured case of the missing data problem. Finally, we conclude in Section V with a summary of the paper and some further research directions that our formulation and interpretation of the Mumford–Shah functional suggests.

II. Mumford-Shah Formulation as a Curve Evolution Problem

The point of reference for this paper is the Mumford–Shah functional [3]

$$E(\mathbf{f},\vec{C}) = \beta \int\int_{\Omega} (\mathbf{f}-\mathbf{g})^2\, dA + \alpha \int\int_{\Omega\setminus\vec{C}} |\nabla \mathbf{f}|^2\, dA + \gamma \oint_{\vec{C}} ds \qquad (1)$$

in which $\vec{C}$ denotes the smooth, closed segmenting curve, $\mathbf{g}$ denotes the observed data, $\mathbf{f}$ denotes the piecewise smooth approximation to $\mathbf{g}$ with discontinuities only along $\vec{C}$, and Ω denotes the image domain [22], [23]. This energy functional is also referred to as the weak membrane by Blake and Zisserman [5]. The parameters α, β, and γ are positive real scalars which control the competition between the various terms above and determine the "scale" of the segmentation and smoothing. Of course one of these parameters can be eliminated by setting it to 1 but for clarity of exposition, we will keep it as is. From an estimation-theoretic standpoint, the first term in $E(\mathbf{f},\vec{C})$, the data fidelity term, can be viewed as the measurement model for $\mathbf{f}$ with β inversely proportional to the variance of the observation noise process. The second term in $E(\mathbf{f},\vec{C})$, the smoothness term, can be viewed as the prior model for $\mathbf{f}$ given $\vec{C}$. The third term in $E(\mathbf{f},\vec{C})$ is a prior model for $\vec{C}$ which penalizes excessive arc length. With these terms, the Mumford–Shah functional elegantly captures the desired properties of segmentation and reconstruction by piecewise smooth functions. The Mumford–Shah problem is to minimize $E(\mathbf{f},\vec{C})$ over admissible $\mathbf{f}$ and $\vec{C}$. The removal of any of the three terms in (1) results in trivial solutions for $\mathbf{f}$ and $\vec{C}$, yet with all three terms, it becomes a difficult problem to solve. In this paper, we constrain the set of discontinuities in the Mumford–Shah problem to correspond to evolving sets of curves, enabling us to tackle the problem via a curve-evolution-based approach.

[3]The final term in the original Mumford–Shah functional consisted of a penalty on the Hausdorff measure of a more general set of discontinuities than we consider here. By restricting the discontinuity set to a smooth curve $\vec{C}$, we are able to replace this term by a simple arc length penalty.

A. Optimal Image Estimation and Boundary-Value Stochastic Processes

For any arbitrary closed curve $\vec{C}$ in the image domain, Ω is partitioned into R and R^c, corresponding to the image domain inside and outside the curve, respectively. Fixing such a curve, minimizing (1) corresponds to finding estimates $\hat{\mathbf{f}}_R$ and $\hat{\mathbf{f}}_{R^c}$ in regions R and R^c respectively, to minimize

$$E_{\vec{C}}(\mathbf{f}_R,\mathbf{f}_{R^c}) = \beta \int\int_R (\mathbf{f}_R-\mathbf{g})^2\, dA + \alpha \int\int_R |\nabla \mathbf{f}_R|^2\, dA + \beta \int\int_{R^c} (\mathbf{f}_{R^c}-\mathbf{g})^2\, dA + \alpha \int\int_{R^c} |\nabla \mathbf{f}_{R^c}|^2\, dA. \qquad (2)$$

The estimates $\hat{\mathbf{f}}_R$ and $\hat{\mathbf{f}}_{R^c}$ that minimize (2) satisfy (decoupled) PDEs which can be obtained using standard variational methods [22]. Alternatively each of these estimates can also be obtained from the theory of optimal estimation. This statistical interpretation suggests lines of inquiry beyond the scope of this paper (which we briefly discuss in Section V). Specifically, the estimate $\hat{\mathbf{f}}_R$ that minimizes (2) can be interpreted as the optimal estimate of a boundary-value stochastic process [1] $\mathbf{f}_R$ on the domain R whose measurement equation is

$$\mathbf{g} = \mathbf{f}_R + \mathbf{v} \qquad (3)$$

and whose prior probabilistic model is given by

$$\nabla \mathbf{f}_R = \mathbf{w} \qquad (4)$$

where $\mathbf{v}$ and $\mathbf{w}$ are independent white Gaussian random fields with covariance intensities $1/\beta$ and $1/\alpha$, respectively.

One effective approach to characterizing $\mathbf{f}_R$ is through the use of complementary processes [1]. In particular, we seek a process $\mathbf{z}$ which *complements* the observation $\mathbf{g}$ in (3) in that $\mathbf{z}$ and $\mathbf{g}$ are uncorrelated and, together, they are informationally equivalent to $\zeta = \{\mathbf{v}, \mathbf{w}\}$ (i.e., to all of the underlying random processes defining the estimation problem). Moreover, since the specification of the statistics of $\mathbf{g}$ in (3) and (4) is via a differential model and involving an internal "state" (namely $\mathbf{f}_R$), we seek an analogous model for $\mathbf{z}$. We refer the reader to [1] for the complete methodology for the direct construction of such complementary models, employing Green's identity and formal adjoints of differential operators. The application of this methodology to (3) and (4) yields the following model for the $\mathbf{z}$

$$\mathbf{z} = \lambda - \alpha \mathbf{w} \tag{5}$$

where the internal state λ satisfies

$$-\nabla\lambda = \beta \mathbf{v} \tag{6}$$

with boundary condition

$$\vec{\mathcal{N}}^T \lambda = 0 \quad \text{on } \vec{C} \tag{7}$$

where $\vec{\mathcal{N}}$ denotes the outer normal of the curve $\vec{C}$.

Eliminating $\mathbf{v}$ and $\mathbf{w}$ from (3)–(6) we can express $\mathbf{f}_R$ and λ completely in terms of $\mathbf{g}$ and $\mathbf{z}$. Then, since $\mathbf{g}$ and $\mathbf{z}$ are uncorrelated, we obtain an internal realization of the optimal estimate $\hat{\mathbf{f}}_R$

$$\begin{bmatrix} \nabla & -\frac{1}{\alpha}\mathbf{I} \\ \beta\mathbf{I} & -\nabla \end{bmatrix} \begin{bmatrix} \hat{\mathbf{f}}_R \\ \hat{\lambda} \end{bmatrix} = \begin{bmatrix} 0 \\ \beta\mathbf{g} \end{bmatrix} \quad \text{on } R$$

with the boundary condition

$$\vec{\mathcal{N}}^T \hat{\lambda} = 0 \quad \text{on } \vec{C}.$$

Eliminating $\hat{\lambda}$ and noticing that the product $\vec{\mathcal{N}} \cdot \nabla \hat{\mathbf{f}}_R$ is the derivative of $\hat{\mathbf{f}}_R$ in the direction of $\vec{\mathcal{N}}$, we obtain the following damped Poisson equation with Neumann boundary condition for $\hat{\mathbf{f}}_R$

$$\hat{\mathbf{f}}_R - \frac{\alpha}{\beta}\nabla^2 \hat{\mathbf{f}}_R = \mathbf{g} \quad \text{on } R \tag{8a}$$

$$\frac{\partial \hat{\mathbf{f}}_R}{\partial \vec{\mathcal{N}}} = 0 \quad \text{on } \vec{C}. \tag{8b}$$

In a similar fashion, $\hat{\mathbf{f}}_{R^c}$ is given as the solution to

$$\hat{\mathbf{f}}_{R^c} - \frac{\alpha}{\beta}\nabla^2 \hat{\mathbf{f}}_{R^c} = \mathbf{g} \quad \text{on } R^c \tag{9a}$$

$$\frac{\partial \hat{\mathbf{f}}_{R^c}}{\partial \vec{\mathcal{N}}} = 0 \quad \text{on } \vec{C}. \tag{9b}$$

We will refer to (8) and (9) as the estimation PDEs. The derivation of these equations using standard techniques of calculus of variations can be found in [22]. The conjugate gradient (CG) method is employed as a fast and efficient solver for these estimation PDEs.

B. Gradient Flows That Minimize the Mumford–Shah Functional

With the ability to calculate $\hat{\mathbf{f}}_R$ and $\hat{\mathbf{f}}_{R^c}$ for any given $\vec{C}$, we now wish to derive a curve evolution for $\vec{C}$ that minimizes (1). That is, as a function of $\vec{C}$, we wish to find $\vec{C}_t$ that minimizes

$$\begin{aligned} E_{\hat{\mathbf{f}}_R, \hat{\mathbf{f}}_{R^c}}(\vec{C}) = {} & \beta \int\int_R (\hat{\mathbf{f}}_R - \mathbf{g})^2 \, dA + \alpha \int\int_R |\nabla \hat{\mathbf{f}}_R|^2 \, dA \\ & + \beta \int\int_{R^c} (\hat{\mathbf{f}}_{R^c} - \mathbf{g})^2 \, dA \\ & + \alpha \int\int_{R^c} |\nabla \hat{\mathbf{f}}_{R^c}|^2 \, dA + \gamma \oint_{\vec{C}} ds. \end{aligned} \tag{10}$$

The first four terms in (10) are of the form

$$J = \int\int_D \mathcal{H} \, dA \tag{11}$$

where D denotes either the interior or the exterior of $\vec{C}$, and $\mathcal{H} : \mathbf{R}^2 \to \mathbf{R}$ is a continuous function. The gradient flow to minimize (11) is given by

$$\vec{C}_t = -\mathcal{H}\vec{\mathcal{N}}. \tag{12}$$

In addition, the gradient flow that minimizes the arc length of $\vec{C}$ is given by

$$\vec{C}_t = -\kappa \vec{\mathcal{N}} \tag{13}$$

where κ denotes the signed curvature of $\vec{C}$. Knowing gradient flows (12) and (13), the curve evolution that minimizes (10) is given by

$$\begin{aligned} \vec{C}_t = {} & \frac{\alpha}{2}(|\nabla \hat{\mathbf{f}}_{R^c}|^2 - |\nabla \hat{\mathbf{f}}_R|^2)\vec{\mathcal{N}} \\ & + \frac{\beta}{2}((\mathbf{g} - \hat{\mathbf{f}}_{R^c})^2 - (\mathbf{g} - \hat{\mathbf{f}}_R)^2)\vec{\mathcal{N}} - \gamma\kappa\vec{\mathcal{N}}. \end{aligned} \tag{14}$$

For the rest of the paper, we will refer to this gradient flow, which is also derived in [22], as the *Mumford–Shah flow*. This flow together with the optimal estimation PDEs makes explicit the coupling between the optimal estimates and the curve evolution.

Flow (14) is implemented via the level set method [26], [36] which offers a natural and numerically reliable implementation of these solutions within a context that handles topological changes in the interface without any additional effort. In this level set framework, $\vec{C}$ is represented by the zero level set of a graph. This graph or level set function evolves so that its zero level set moves according to (14). One can then interpolate between neighboring pixels to locate the precise locations of the zero level set as the level set function evolves. Additionally, in this framework, since $\hat{f}$, $\nabla\hat{f}$, and g are only defined on the zero level set, they need to be extended to other level sets as well. Following the approach described in [26], we let these values at points on the other level sets take on their corresponding values from the closest point on the zero level set. In this manner, $\hat{f}$, $\nabla\hat{f}$, and g are extended to all other level sets. This will ensure proper evolution of the entire level set function. Finally, spatial derivatives in the level set framework associated with the curvature term in (14) are computed via central differences whereas spatial derivatives associated with the other terms in (14) (i.e.,

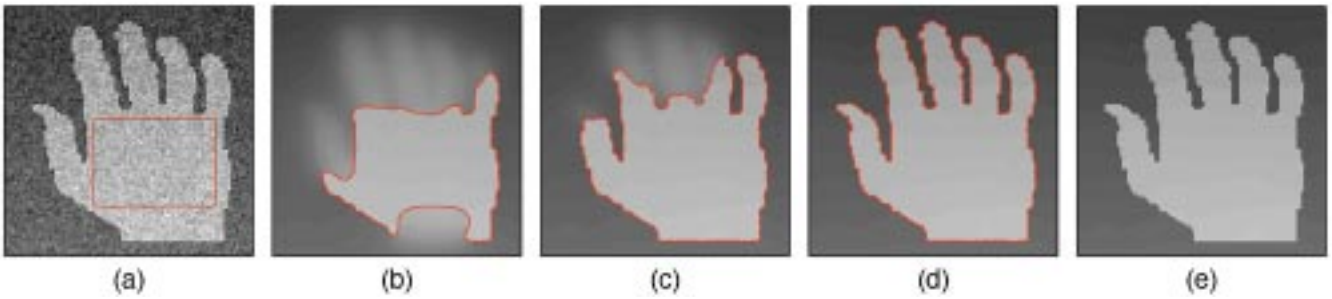

Fig. 2. Outward flow from inside.

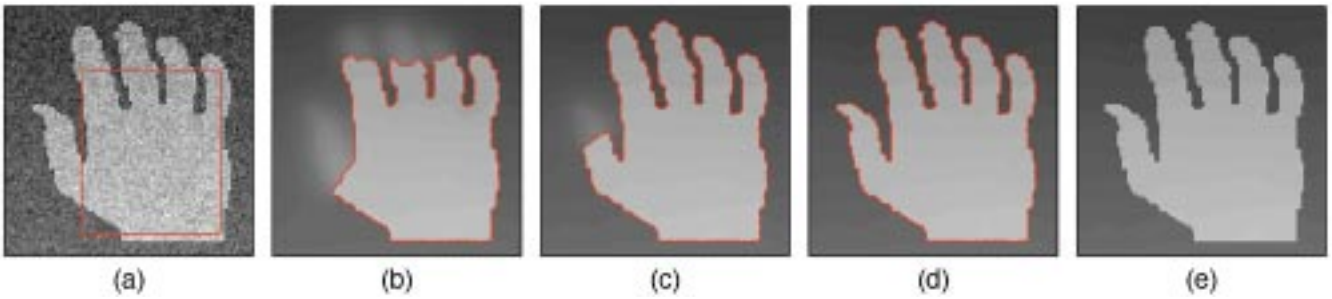

Fig. 3. Bidirectional flow.

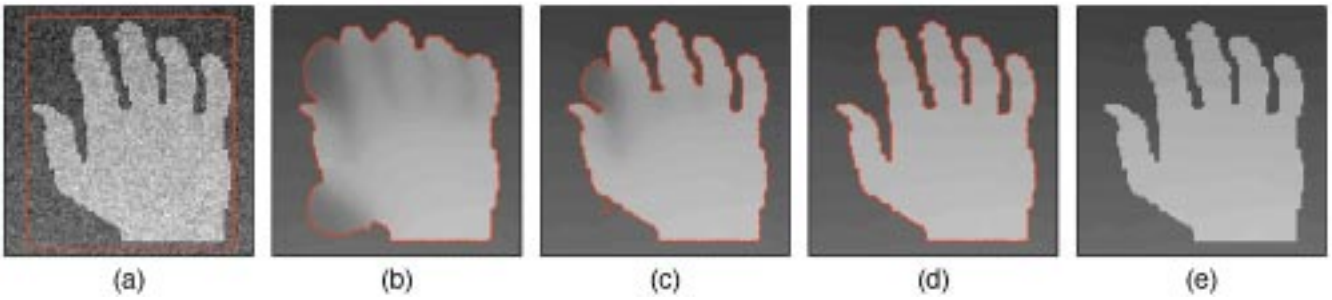

Fig. 4. Inward flow from outside.

terms involving $\hat{f}$ and g) are computed using monotone upwind difference schemes [26] in order to capture the viscosity solution (when there are discontinuities or jumps in the fitting term, for example).

C. Remarks on the Mumford–Shah Active Contour Model

One very attractive feature associated with our Mumford–Shah active contour model (and also present in other region-based methods) is that it automatically proceeds in the correct direction without relying upon additional inflationary terms commonly employed by many active contour algorithms. We illustrate this in Figs. 2–5 with a noisy synthetic image of a hand. An initial contour completely contained within the hand will flow outward toward the boundary (Fig. 2); an initial contour partially inside and partially outside the hand will flow in both directions toward the boundary (Fig. 3); an initial contour encircling the hand will flow inward toward the boundary (Fig. 4); and finally, an initial contour situated outside the hand will flow outward toward and wrap around the boundary (Fig. 5). In these figures, Frame (a) shows the initializing contour with the original image; Frames (b) and (c) show the estimate of curve $\vec{C}$ and the estimates of $\mathbf{f}_R$ and $\mathbf{f}_{R^c}$ associated with two intermediate steps of the algorithm; Frame (d) shows the final segmenting curve $\vec{C}$ and the final reconstruction of the image (based on the estimates $\hat{\mathbf{f}}_R$ and $\hat{\mathbf{f}}_{R^c}$); and finally, Frame (e) shows the reconstruction of the image without the overlaying curve for comparison to the original noisy image. Note that the smooth estimate of the image is continuously estimated based on the current position of the curve. In Fig. 5, in addition to the curves that outline the boundary of the hand, there exists extraneous curves around the four corners of the image which do not correspond to image edges. This is due to the fact that the algorithm has descended upon and settled on to a local minimum—a common problem faced by all algorithms which rely on gradient descent methods for minimization. However, notice that the piecewise smooth reconstruction of the image shown in Fig. 5(e) does not exhibit any ill effects from these extraneous curves; that is, the reconstruction does not show any semblance of an edge along

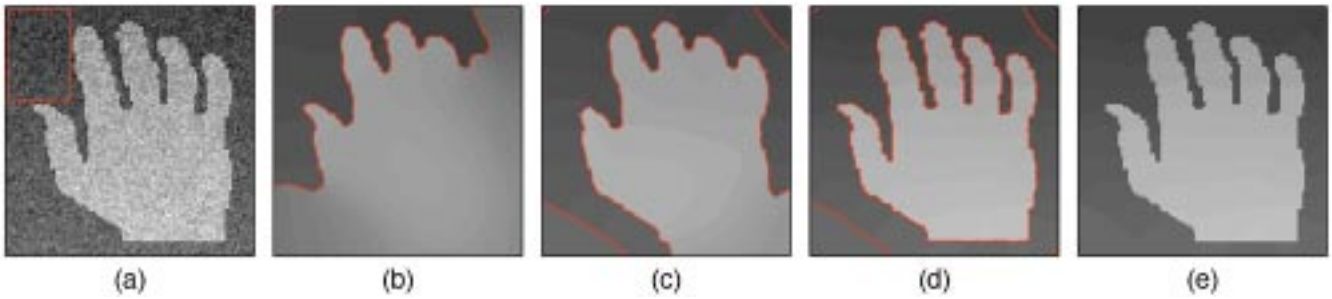

Fig. 5. Outward flow from outside.

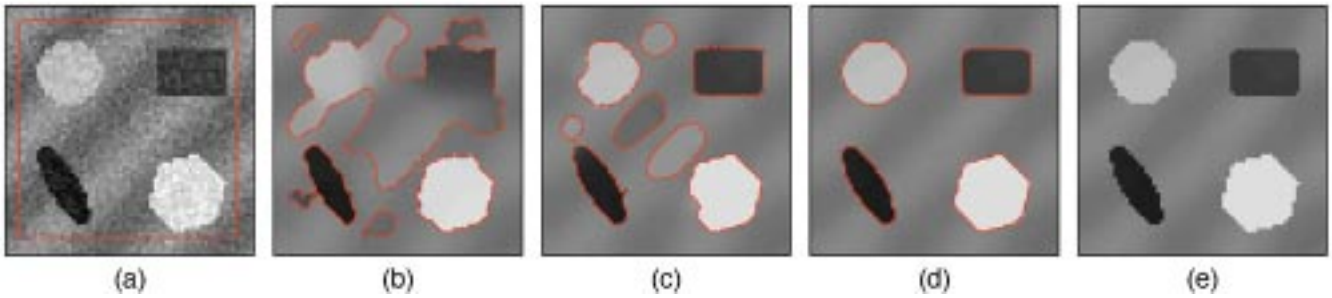

Fig. 6. Segmentation and smoothing of an image with four distinct foreground regions.

these extraneous curves. Thus even if the curve is trapped at a local minimum, the reconstruction of the image is still accurate.

The class of imagery that our algorithm can handle is not restricted just to images with only two distinct means but is equally applicable to images with multiple nonoverlapping regions each with different means. Moreover, we do not need to know in advance the number of such regions or distinct means are present. As shown in Fig. 6, segmentation and smoothing are performed on a noisy synthetic image with four foreground regions of different means situated on a spatially varying background region. Multiple disjoint regions are captured by a single contour demonstrating the topological transitions allowed by the model's level set implementation. However, it is important to point out that this single curve cannot detect a region if it is inside another region. In Section III-D, we demonstrate how to handle this type of problem.

Our model can also be generalized, in a very straight forward manner, to handle vector-valued images[4] (e.g., color images or images obtained from scale and orientation decompositions commonly used for texture analysis). Consider the following vector version of the Mumford–Shah functional:

$$E(\mathbf{f}_1, \mathbf{f}_2, \ldots, \mathbf{f}_k, \vec{C}) = \beta \int\int_{\Omega} \sum_{i=1}^{k} (\mathbf{f}_i - \mathbf{g}_i)^2 \, dA + \alpha \int\int_{\Omega \setminus \vec{C}} \sum_{i=1}^{k} |\nabla \mathbf{f}_i|^2 \, dA + \gamma \oint_{\vec{C}} ds$$

[4]Chan and Vese, who have considered the piecewise constant version of the Mumford–Shah functional [9], have also extended their framework to vector-valued data in "Active Contours Without Edges for Vector-Valued Images" (see http://www.math.ucla.edu/applied/cam).

where $\mathbf{g}_i$ and $\mathbf{f}_i$ denote the ith component of the k-dimensional vector-valued observed data and its smooth estimate, respectively. The curve evolution that minimizes this energy functional is given by

$$\vec{C}_t = \frac{\alpha}{2} \sum_{i=1}^{k} (|\nabla \hat{\mathbf{f}}_{iR^c}|^2 - |\nabla \hat{\mathbf{f}}_{iR}|^2) \vec{\mathcal{N}} + \frac{\beta}{2} \sum_{i=1}^{k} ((\mathbf{g}_i - \hat{\mathbf{f}}_{iR^c})^2 - (\mathbf{g}_i - \hat{\mathbf{f}}_{iR})^2) \vec{\mathcal{N}} - \gamma \kappa \vec{\mathcal{N}}. \quad (15)$$

The $\hat{\mathbf{f}}_{iR}$ and $\hat{\mathbf{f}}_{iR^c}$ for $i = 1, \ldots, k$ in (15) is given by the solutions to the following:

$$\hat{\mathbf{f}}_{iR} - \frac{\alpha}{\beta} \nabla^2 \hat{\mathbf{f}}_{iR} = \mathbf{g}_i \quad \text{on } R$$

$$\frac{\partial \hat{\mathbf{f}}_{iR}}{\partial \vec{\mathcal{N}}} = 0 \quad \text{on } \vec{C}$$

and

$$\hat{\mathbf{f}}_{iR^c} - \frac{\alpha}{\beta} \nabla^2 \hat{\mathbf{f}}_{iR^c} = \mathbf{g}_i \quad \text{on } R^c$$

$$\frac{\partial \hat{\mathbf{f}}_{iR^c}}{\partial \vec{\mathcal{N}}} = 0 \quad \text{on } \vec{C}.$$

For demonstration, in Fig. 7, we show the segmentation and smoothing of a noisy color image of six different types of gemstones.

III. Implementation

There are two ways in which we can improve the Mumford–Shah active contour model presented so far. One, we can speed up the convergence of the algorithm by reducing the

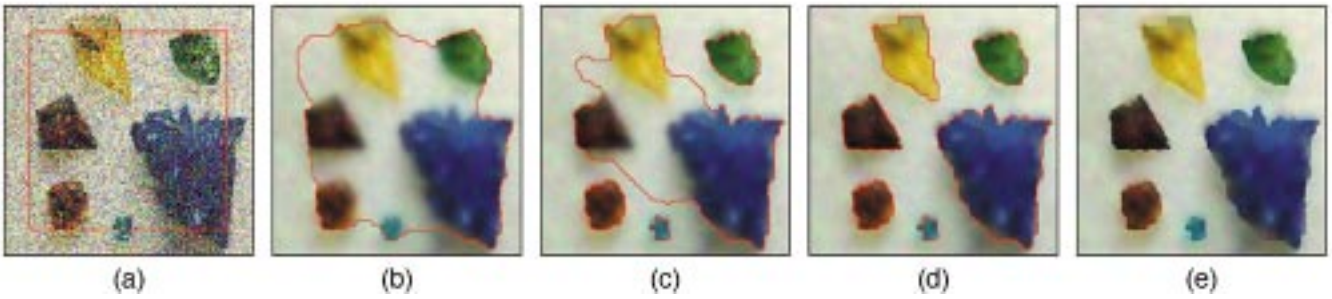

Fig. 7. Segmentation and smoothing of a color image with six distinct foreground regions.

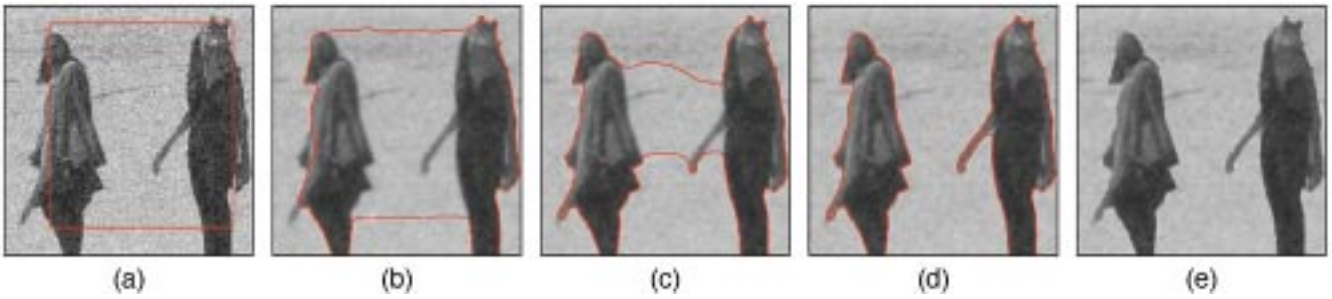

Fig. 8. Noisy image of Star Wars characters Qui-Gon Jinn and Jar Jar Binks in the Tatooine Desert.

required number of curve evolution steps and/or reducing the computational cost of solving the estimation PDEs. Two, we can build on the level set implementation of our algorithm to enable it to handle important image features such as triple points and other multiple junctions (without having to resort to more sophisticated level set techniques [21]). In this section, we present several progressive modifications to the implementation of the Mumford–Shah active contour model in order to make these improvements.

A. Approximate Gradient Descent

We propose an approximate gradient descent approach to calculate $\hat{\mathbf{f}}_R$, $\hat{\mathbf{f}}_{R^c}$, and $\vec{C}$ that minimize the Mumford–Shah functional shown in (1). This approach consists of alternating between the following two steps:

- fix $\hat{\mathbf{f}}_R$ and $\hat{\mathbf{f}}_{R^c}$, and take several gradient descent curve evolution steps to move the curve $\vec{C}$;
- fix $\vec{C}$, and perform just a few iterations of the CG method for the estimation PDEs—without taking it to convergence—to obtain a *rough* estimate of $\mathbf{f}_R$ and $\mathbf{f}_{R^c}$.

We have found that it is not necessary to get an accurate estimate of $\mathbf{f}_R$ and $\mathbf{f}_{R^c}$ at each evolution. All that is required is a rough estimate of these values to direct the curve to move in the general descent direction. The idea is to make the algorithm faster by reducing the number of times $\hat{\mathbf{f}}_R$ and $\hat{\mathbf{f}}_{R^c}$ are estimated and also the amount of time required to calculate each of them. The CG procedure is then carried to convergence in the last iteration to obtain an accurate final estimate of $\mathbf{f}_R$ and $\mathbf{f}_{R^c}$.

Fig. 8 illustrates the performance of this approach applied to a noisy image of two Star Wars characters with different mean intensities standing in a spatially varying background. We obtained the results shown in this figure by alternating between 20 curve evolution steps and ten CG iterations. This reduced number of CG iterations compares quite favorably to the 100 CG iterations we use in the end to obtain an accurate estimate of $\mathbf{f}_R$ and $\mathbf{f}_{R_c}$. The segmentation clearly delineated the two Star Wars characters. The reconstruction of the image accurately captured the spatially varying background and preserved the structures within each Star Wars character. Obviously, this is not possible with lower dimensional models (such as ones based on mean intensities [9]).

As one can appreciate, a large number of curve evolution steps is required to take the curve from the initial curve in Fig. 8(a) to the final segmenting curve in Fig. 8(d). This translates to calculating $\hat{\mathbf{f}}_R$ and $\hat{\mathbf{f}}_{R^c}$ many times, a significant load even if we only perform a few CG iterations at each step. In the next two subsections, we address this computational issue.

B. Two-Step Approach

One way to reduce the number of curve evolution steps is to obtain a good initial estimate of the curve $\vec{C}$ so that the travel distance of the initializing curve to the correct image boundary is reduced. One approach to doing so, that works if there are only two distinct means in the image, is to employ the method of Chan and Vese [9] referred to earlier. Chan and Vese restrict the two regions, R and R^c, to have constant values. For the class of bimodal images, this restriction is equivalent to taking $\alpha = \infty$ in our Mumford–Shah active contour model. This reduces flow (14) to

$$\vec{C}_t = \frac{\beta}{2}(u - v)(\mathbf{g} - u + \mathbf{g} - v)\vec{\mathcal{N}} - \gamma\kappa\vec{\mathcal{N}} \qquad (16)$$

where u and v are the average intensities of R and R^c, respectively. This is precisely the flow presented in [9]. Evolving the curve according to this flow is fast since each evolution only requires updating the mean values inside and outside the curve. We evolve any starting curve according to this flow in order to obtain a good initial estimate of $\vec{C}$. Once we have this estimate,

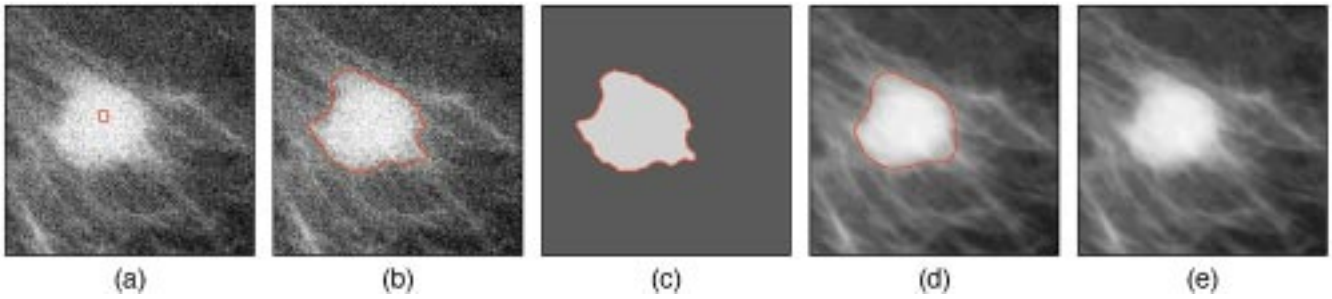

Fig. 9. Mammogram showing cyst in the breast tissue.

we relax α to a finite value, and employ the approximate gradient descent method introduced earlier to minimize the general form of the Mumford–Shah functional. Since the initializing curve is presumably close to the correct boundary, the number of evolution steps required for convergence is greatly reduced. However, due to the use of flow (16) in calculating the initial estimate of $\vec{C}$, this two-step implementation approach of our Mumford–Shah active contour model can only handle images with two distinct means. In Section III-D, we discuss ways to circumvent this restriction.

Fig. 9 illustrates this two-step implementation of our model. In Fig. 9(a), a noisy mammogram showing a cyst in the breast tissue is displayed, along with the starting curve. The next frame shows the estimate of $\vec{C}$ obtained by assuming piecewise constant regions; that is, obtained by employing flow (16). This curve is superimposed on top of the original image. Fig. 9(c) shows the piecewise constant approximation of the image based on this segmenting curve. In Fig. 9(d) we show the results of applying the approximate gradient descent implementation of our active contour model using, as initializing curve, the one shown in Fig. 9(b). Equal penalty on the arc length of the curve is use in obtaining the curves shown in Fig. 9(b) and (d). For comparison to the original image, in Fig. 9(e) we show the optimal estimate produced by our algorithm with the segmenting curve suppressed. It is clear from these results that the segmentation of the cyst has been refined and that a denoised restoration of the image is obtained.

C. Multiresolution Approach

Though the two-step approach of above can substantially reduce the computational complexity of our algorithm, it is limited by its ability to handle only images with two distinct means. We now describe a multiresolution approach that not only speeds up our algorithm, but also adheres to the original capabilities of our model to segment images with multiple nonoverlapping regions.

The basic idea of the multiresolution approach is to use a coarsened representation of the image to obtain a good estimate of the segmenting curve, and then progressively refine this estimate of the curve as the resolution of the image is increased. Given an image, we repeatedly subsample it by a factor of two in both the x and the y direction to obtain a set of images of varying resolution. The subsampling process terminates before the relevant features within the image are lost. We begin our multiresolution approach by applying our technique at the coarsest scale. Operating at such a coarse scale, we decrease the number of curve evolution steps required by reducing the travel distance between the starting curve and the final curves. In addition, the computational requirements in updating each curve evolution step is also reduced due to the smaller image domain within which the curve is evolving. More importantly, we have substantially decreased the computation of $\hat{\mathbf{f}}_R$ and $\hat{\mathbf{f}}_{R^c}$ since the discretized versions of the estimation PDEs are much lower in dimension on the coarsened grid. The final curve obtained on each coarsened image is upsampled by a factor of two in both the x and the y direction to serve as the initializing contour for the image at the next higher resolution (which is easily accomplished in the level set framework by replication of the level set grid). At this new scale, because the initializing contour is already close in proximity to the edges of the image, the number of approximate gradient steps required to move the curve toward the edges of the image is small. We also end up with a good initial estimate of the smooth field at this new scale by up sampling. Only a few iterations is required in obtaining a segmentation and reconstruction of the image at this scale. This process of using the segmenting curve at one resolution as the initial curve for the next finer resolution is repeated until the finest resolution image is reached. As one can appreciate, the reductions in computation based on this multiresolution approach give rise to a much more efficient implementation of our algorithm. In Fig. 10, we demonstrate this multiresolution approach on a color photograph consisting of two different foreground regions (i.e., the two parakeets). A single contour successfully captured both foreground regions even though one foreground region has a different vector mean than the other.

D. Hierarchical Approach

We now propose an implementation of our active contour model, building on the preceding modifications, to enable our model to handle images with multiple junctions or holes (i.e., a region inside another region) without resorting to more sophisticated level set techniques [21]. This approach also allows the possibility of using the two-step approach of Section III-B to handle images with more than two distinct regional means.

Given an image, we apply our Mumford–Shah active contour model for segmentation and smoothing. After segmentation, if any of the resulting subregions require additional segmentation, apply our algorithm again, but this time, restricting the algorithm to operate only in that particular subregion. This approach has the natural notion of starting with a crude segmentation and refining the segmentation by telescoping down to the different subregions in order to capture finer and finer details in the image. The attractive feature associated with this implemen-

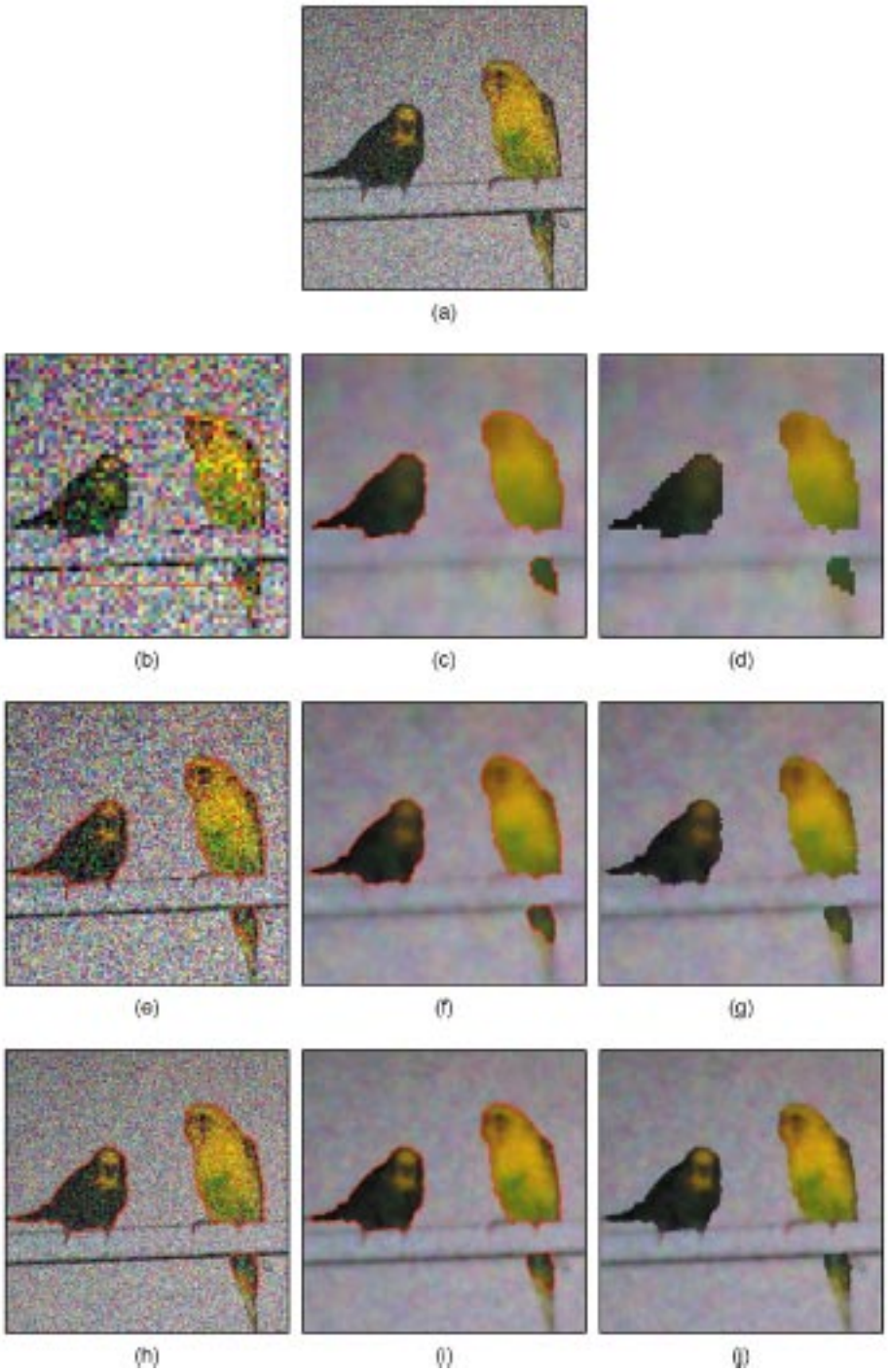

Fig. 10. Segmentation and smoothing of a noisy color image of two parakeets using a multiresolution approach. (a) The 240×240 original noisy image. (b) The 60×60 subsampled image of the original with initializing curve. (c) Final segmenting curve superimposed on top of the 60×60 reconstructed image. (d) Reconstruction of the 60×60 subsampled image. (e) The 120×120 original image with initializing curve obtained from the up sampled version of the curve from (c). (f) Final segmenting curve superimposed on top of the 120×120 reconstructed image. (g) Reconstruction of the 120×120 original image. (g) The 240×240 original image with initializing curve obtained from the up sampled version of the curve from (f). (i) Final segmenting curve superimposed on top of the 240×240 reconstructed image. (g) Reconstruction of the 240×240 original image.

tation is that it allows us to handle images with triple points by employing multiple curves to represent such junctions. Moreover, this nested implementation affords us better control as to what details we desire and what objects we would like to capture, in the segmentation and smoothing of our image.

Remark: It should be pointed out that this hierarchical approach has the limitation in that the boundaries of the regions detected later in the process meet the boundaries detected earlier at right angles. This follows from the theory of junctions as presented in [22]. Consequently, the triple points found by this approach are necessarily T-junctions.

We use the brain pathology image shown in Fig. 11(a) to demonstrate how the approximate gradient descent method of Section III-A is used in the hierarchical implementation of our

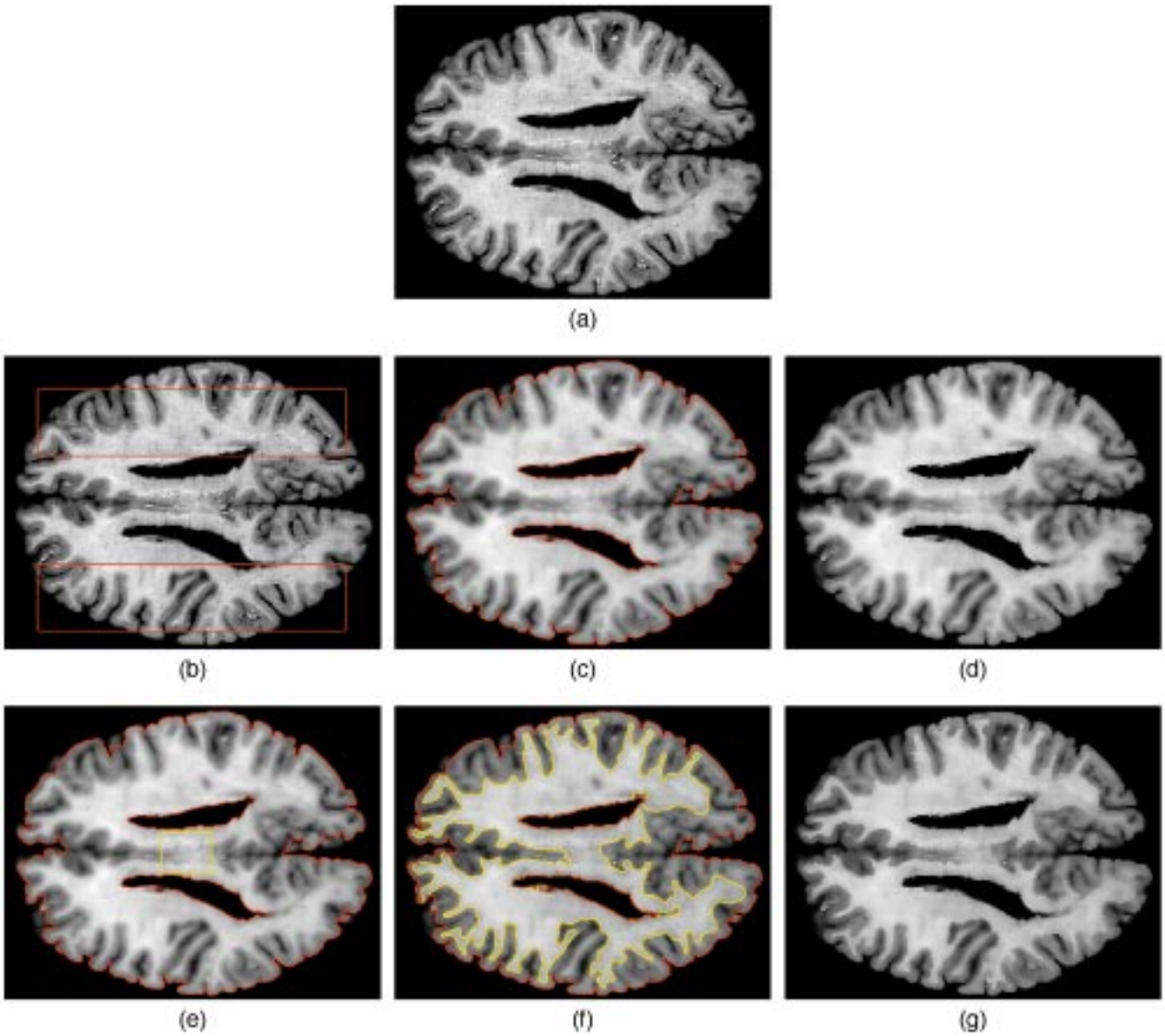

Fig. 11. Hierarchical segmentation and reconstruction of a pathology image of the human brain based on the approximate gradient descent method. (a) Original image. (b) Original image with the initialization scheme for the first curve. (c) Segmentation and reconstruction of the image based on the first curve. (d) Initialization scheme for the second curve. (e) Segmentation and reconstruction of the image based on both curves. (f) Reconstruction of the image based on both curves. Same set of parameters are used in obtaining the results shown in (c) and (e).

active contour model. First, the segmentation and reconstruction of the image shown in Fig. 11(c) is obtained based on the approximate gradient descent approach. The blurring across the boundary of the white and the gray matter is due to the erroneous implication of this coarse segmentation, namely that the inside of the brain is one region over which smoothing is performed. To provide better details within the brain, we again applied our technique to the interior region of the brain to obtain the segmentation and reconstruction of the image shown in Fig. 11(e). The yellow and the red curves segment the image into the background, white matter, and the gray matter. The piecewise smooth reconstruction of the image is shown in Fig. 11(f) without the segmenting curves for comparison to the original image.

Next, we demonstrate how the two-step method of Section III-B is used in the hierarchical implementation of our model for the segmentation and reconstruction of a color Doppler ultrasound image of the heart [Fig. 12(a)] using the vector-valued model in (15). To demonstrate the effectiveness of our technique for denoising/smoothing, we show in the next two frames a comparison between the blurry representation of the original image after applying isotropic smoothing (i.e., the same smoothing used in the Mumford–Shah framework but applied over the image as one single region) as shown in Fig. 12(b), and the smooth representation of the original image without blurring across edges using the Mumford–Shah model [Fig. 12(c)]. The reconstruction in Fig. 12(c) is obtained using four segmenting curves, each at a different hierarchical level. Fig. 12(d)–(g) shows the initialization technique for the four different curves at the different hierarchical levels. Notice that each region or subregion of interest is seeded with regularly spaced initial contours for automatic segmentation.

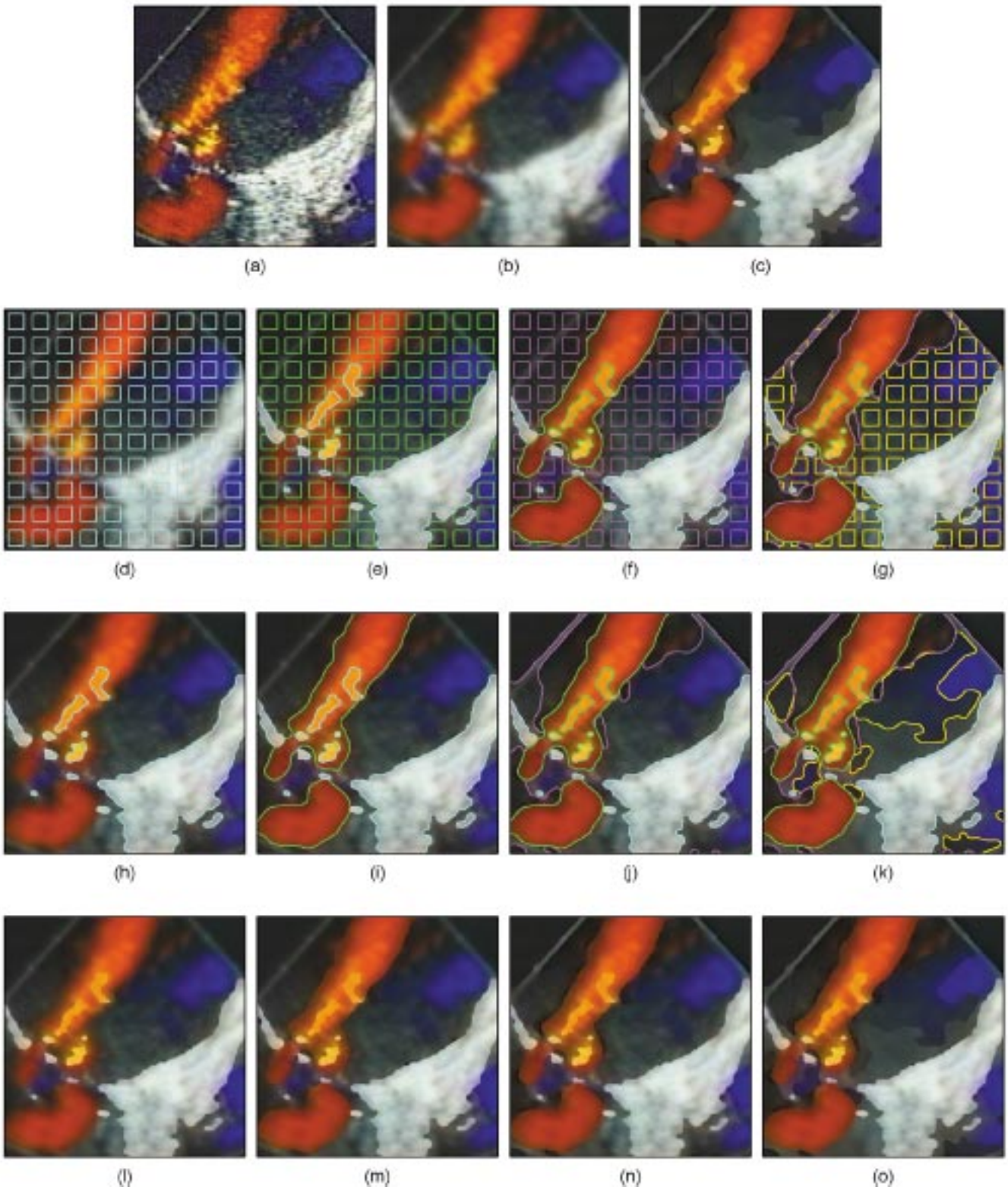

Fig. 12. Hierarchical segmentation and reconstruction of a color Doppler ultrasound image of the heart based on the two-step approach. (a) Original image. (b) Original image after isotropic smoothing. (c) Reconstruction of the original image based on the Mumford–Shah active contour [same image as (o)]. (d)–(g) Initialization scheme for the various levels of the hierarchy. (h)–(k) Segmentation and reconstruction of the image at various levels of the hierarchy. (l)–(o) Reconstruction of the image at various levels of the hierarchy.

Fig. 12(h)–(k) shows the segmentation and reconstruction of the image at the various hierarchical levels based on the two-step method. And finally, Fig. 12(l)–(o) show just the reconstruction of the image at the various hierarchical levels without the overlaying segmenting curve. At the first hierarchical level, the cyan curve captured the white myocardium and the yellow region corresponding to slow-velocity-flow moving away from the ultrasound transducer. The green curve at the second

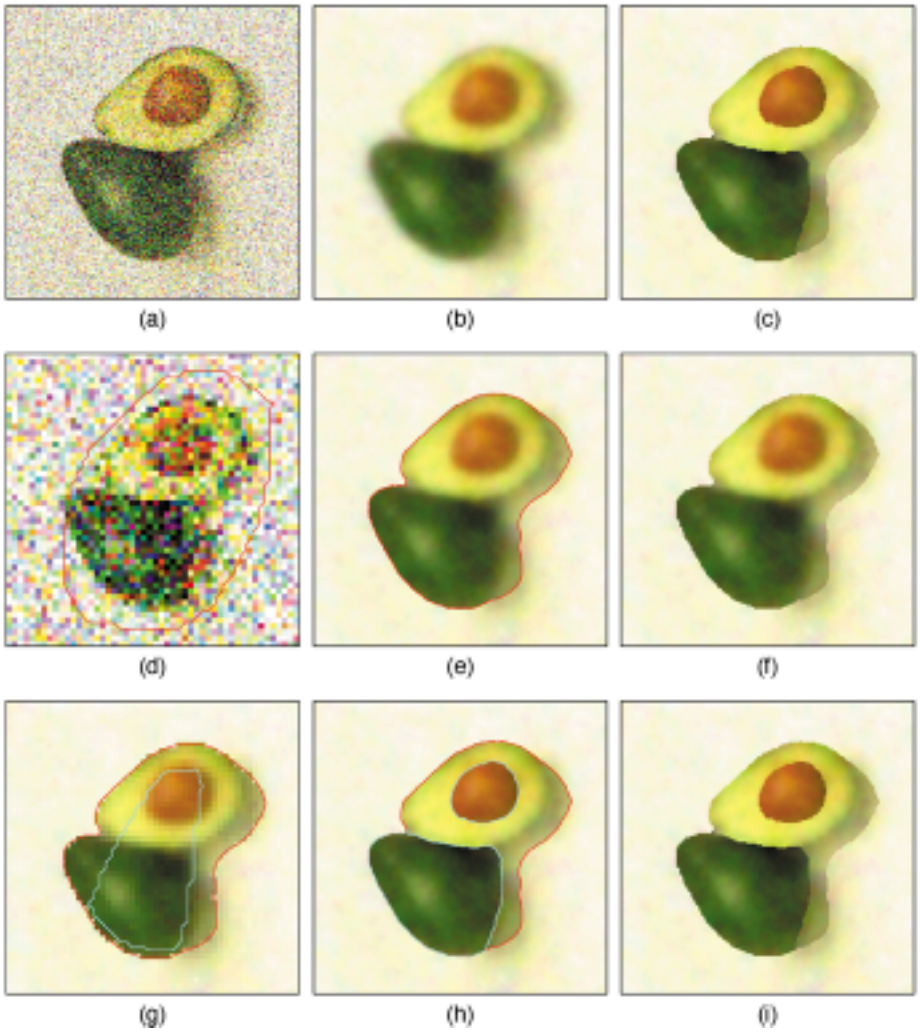

Fig. 13. Hierarchical segmentation and reconstruction of a noisy avocado image based on the multiresolution approach. (a) Original image. (b) Original image after isotropic smoothing. (c) Reconstruction of the original image based on the Mumford–Shah active contour [same image as (i)]. (d)–(f) Level 1 multiresolution segmentation and reconstruction. (g)–(i) Level 2 multiresolution segmentation and reconstruction.

hierarchical level captured the red high-velocity-flow moving away from the transducer near the heart valve. The magenta curve at the third hierarchical level captured the ultrasound beam (together with other brighter parts of the image). Finally, at hierarchical level four, the yellow curve separated the blue region (which corresponds to blood flow moving toward the transducer) from the brighter remaining areas within the image. As shown in this figure, these four curves enable the segmentation to handle various multiple junctions.

Finally, by using the noisy avocado image in Fig. 13(a), we illustrate how the multiresolution approach of Section III-C can be used in the hierarchical implementation of our model. As before, in Figs. 13(b) and (c), we compare isotropic smoothing (with no segmentation) with smoothing based on the Mumford–Shah model. Fig. 13(d) shows the initialization scheme at the coarsest resolution for the first curve. Fig. 13(e) shows the curve successfully capturing the two avocado-halves, shown at the finest resolution. The reconstruction of the image based on this one curve is shown in Fig. 13(f). Notice the blurring across the shared edge of the two avocado-halves and across the edges of the avocado seed. To prevent this blurring, a second curve is used. Fig. 13(g) shows the initialization scheme for this second curve at the coarsest resolution. Shown at the finest resolution, Fig. 13(h) demonstrates how this new curve can capture *both* foreground regions-the seed and the dark outer covering of the avocado-despite the fact that these two foreground regions have distinct vector means. As advertised, the hierarchical approach enables our algorithm to handle triple points. Two such points can clearly be seen here. Fig. 13(i) shows the reconstruction of the noisy image without the segmentation curves.

In general, it is not always possible to automatically determine if a particular subregion requires further segmentation or not. However, in cases where one is only interested in the the reconstruction of a noisy image and not the location of the segmenting curve, the hierarchical approach can be employed in a fully automatic fashion. Specifically, we can use the two-step method of Section III-B within the hierarchical implementation to recursively segment and smoothly reconstruction the image. The recursive procedure terminates when further segmentation of a particular subregion does not substantially decrease the Mumford–Shah energy functional (i.e., by a predefined amount set equal to a percentage of the starting energy value).

IV. Extensions

So far, we have focused on developing our algorithm for the particular context in which the Mumford–Shah functional was originally designed, namely simultaneous image segmentation and denoising. However, the range of applications of our algorithm is much richer. In this section, we extend the approach of previous section to handle simultaneous segmentation, denoising, and interpolation by generalizing the original Mumford–Shah functional.

A. Segmentation, Denoising, and Interpolation of Images with Missing Data

Images with missing data are frequently encountered in many image processing problems. Various approaches have been proposed to restore these images [4], [8], [20], [24]. One approach in tackling this problem is the use of estimation-theoretic techniques to interpolate the values of these missing data. Standard estimation formalisms will generally produce smooth interpolations in regions of missing data, something that is not desirable if there are high-contrast boundaries in any of these missing data regions. However, since our formalism brings optimal estimation and curve evolution together, a very easily implemented generalization allows us to deal with segmentation, denoising, and interpolation in a simultaneous and naturally coupled manner.

Our model handles missing data through the parameter β. In the standard Mumford–Shah formulation (1), β is a constant scalar parameter reflecting our confidence in the measurements. To accommodate applications in which the data quality is spatially varying and even in the limiting such case in which there are missing pixel measurements distributed arbitrarily through the image domain, we replace the constant parameter β by a spatially varying function β whose value at each pixel is inversely proportional to the variance of the measured noise at that pixel. For example, in the situation where the data at pixel (x_o, y_o) is missing, we consider the variance of the data at that pixel as being infinite and accordingly set $\beta(x_o, y_o) = 0$. By introducing this spatially varying β, (1) becomes

$$E(\mathbf{f}, \vec{C}) = \int\int_{\Omega} \beta(\mathbf{f}-\mathbf{g})^2\, dA + \alpha \int\int_{\Omega\setminus\vec{C}} |\nabla \mathbf{f}|^2\, dA + \gamma \oint_{\vec{C}} ds. \quad (17)$$

The gradient flow that minimizes (17) is given by

$$\vec{C}_t = \frac{\alpha}{2}(|\nabla \hat{\mathbf{f}}_{R^c}|^2 - |\nabla \hat{\mathbf{f}}_R|^2)\vec{\mathcal{N}} + \frac{\beta}{2}((\mathbf{g}-\hat{\mathbf{f}}_{R^c})^2 - (\mathbf{g}-\hat{\mathbf{f}}_R)^2)\vec{\mathcal{N}} - \gamma\kappa\vec{\mathcal{N}} \quad (18)$$

where the optimal estimates $\hat{\mathbf{f}}_R$ and $\hat{\mathbf{f}}_{R^c}$ of (18) satisfy

$$\beta\hat{\mathbf{f}}_R - \alpha\nabla^2\hat{\mathbf{f}}_R = \beta\mathbf{g} \quad \text{on } R$$

$$\frac{\partial \hat{\mathbf{f}}_R}{\partial \vec{\mathcal{N}}} = 0 \quad \text{on } \vec{C}$$

and

$$\beta\hat{\mathbf{f}}_{R^c} - \alpha\nabla^2\hat{\mathbf{f}}_{R^c} = \beta\mathbf{g} \quad \text{on } R^c$$

$$\frac{\partial \hat{\mathbf{f}}_{R^c}}{\partial \vec{\mathcal{N}}} = 0 \quad \text{on } \vec{C}.$$

Over each region of missing data D, the estimation equation reduces to the Laplace equation with the same Neumann boundary condition

$$\nabla^2\hat{\mathbf{f}}_D = 0 \quad \text{on } D \quad (19a)$$

$$\frac{\partial \hat{\mathbf{f}}_D}{\partial \vec{\mathcal{N}}} = 0 \quad \text{on } \vec{C}. \quad (19b)$$

As solutions to the Laplace equation, the estimates obtained over any such missing data regions not containing part of $\vec{C}$ take the form of harmonic functions. As such, we can infer much about the smooth nature of these interpolated estimates as they are subject to both a maximum (and minimum) principle as well as the mean value property. However if the curve $\vec{C}$ intersects D, no such smoothing occurs across this boundary, allowing interpolation to be guided by the segmentation defined by $\vec{C}$. To illustrate this, we show in Fig. 14(a) a synthetic image of the United States with regions of missing data. The synthetic image is made in an attempt to simulate a satellite picture of the United States with regions of incomplete data as a result of obscuration by cloud coverings. The final curve estimate is depicted in Fig. 14(d), and the denoise and interpolated reconstruction is shown in Fig. 14(e).

Our algorithm can also be used to segment and reconstruct images with isolated pixels of missing data distributed arbitrarily throughout the image as shown in Fig. 15(a). This often occurs in imaging modalities subject to speckle. By utilizing the prior smoothness constraint placed on the image, our model appropriately "fills in" the gaps created by the missing data. In Fig. 16, we demonstrate the capabilities of our method on a real image. A forward-looking infrared (FLIR) image of three tanks with missing data distributed throughout the image is shown in Fig. 16(a). The missing data in the FLIR image are due to intensity saturated and defective pixels of the infrared sensor. Using our method, we are able to segment out the tanks and also provide a denoised and complete reconstruction of the image.

B. Segmentation-Based Image Magnification

Image magnification or spatial resolution enhancement is required in a variety of applications including image compression, image coding, and HDTV. It deals with the problem of enlarging a small image to several times its size and often requires some sort of an interpolation scheme. The most straightforward approach for image enlargement is to use a zero-order interpolation technique, commonly known as replication, which may cause the resulting image to appear blocky [18]. Classical enlargement techniques such as bilinear or bicubic interpolation schemes tend to cause blurring across the edges when applied indiscriminantly to the image [18]. More sophisticated schemes may locate the edges first with local filters prior to interpolation so as to avoid the blurring artifacts [2], [31]. Three important shortcomings are evident in these types of algorithms. One,

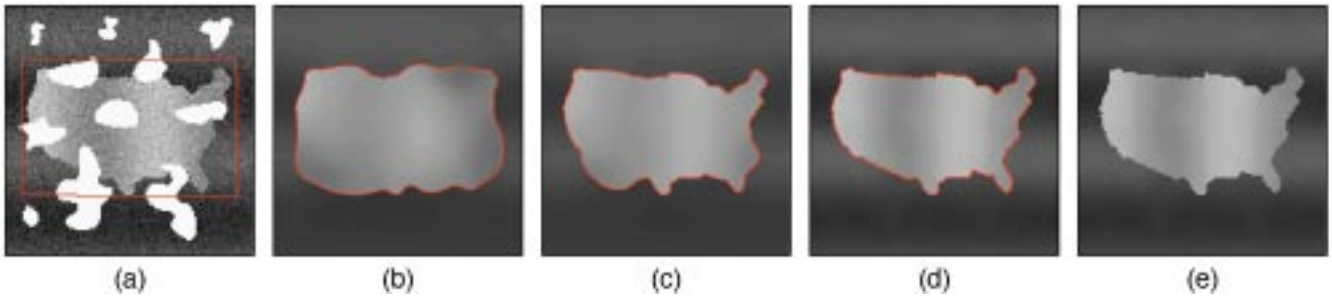

Fig. 14. Segmentation and smoothing of a synthetic image with regions of missing data. Missing measurement data points are shown as white pixels in (a).

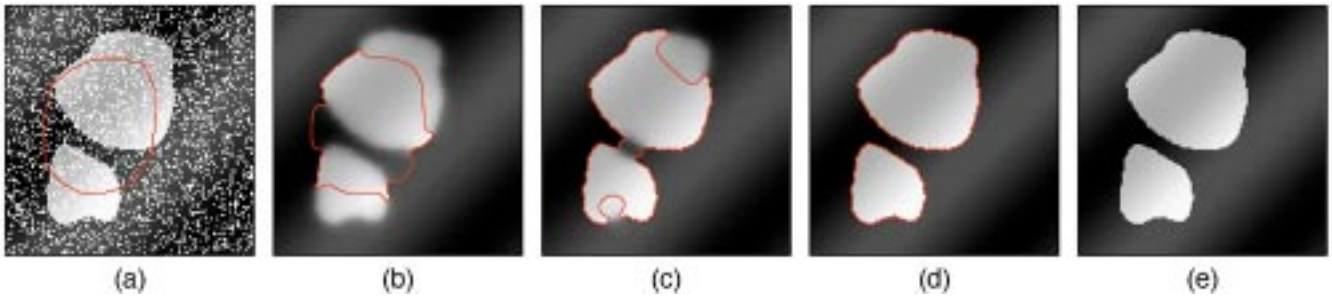

Fig. 15. Segmentation and smoothing of a synthetic image with missing data distributed throughout the image. Missing measurement data points are shown as white pixels in (a).

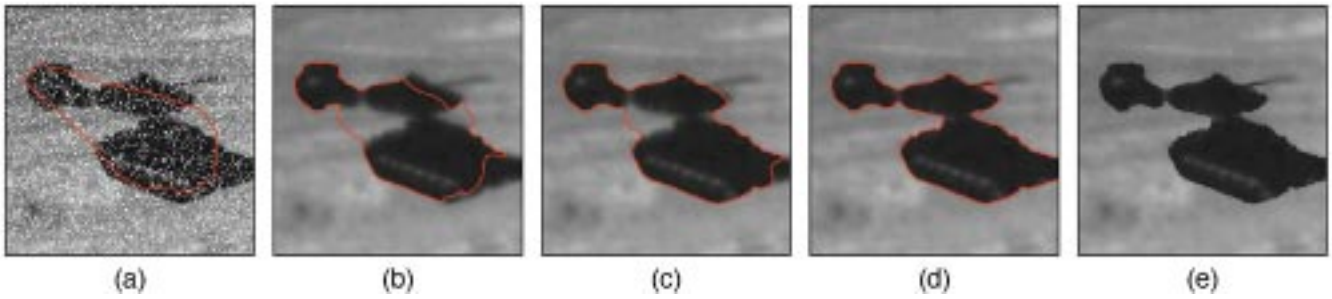

Fig. 16. FLIR image of M2, T62, and M60 tanks. White pixels in (a) denote locations with missing data.

the interpolation schemes used for magnification are local since they only utilize data values from neighboring pixels. This interpolation scheme becomes even more problematic when the observed image is noisy. Two, edge detection schemes employed prior to interpolation often only make use of local information (which are very susceptible to noise artifacts) and cannot guarantee continuous closed edge contours. Three, it is unclear in what order the three operations (smoothing, edge detection, and interpolation) should be performed since they are not commutative. Our approach for image magnification addresses the first deficiency by using an estimation-theoretic (PDE-based) model for interpolating the data which incorporates the use of all data values within each homogeneous region, not just neighboring pixels, to determine the interpolant. As a result, this interpolation scheme is much more robust to noise. The second deficiency is addressed by the use of our active contour model for boundary detection which is more global in nature than local filters (and therefore not as sensitive to noise) and is curve-based (hence providing a continuous closed edge contour). The third deficiency is addressed by using the Mumford–Shah model's principled approach to provide, in a single framework, a tight coupling for simultaneous image segmentation, denoising, and, under the extensions we have just presented, interpolation. In this manner, the ordering of the different operations is no longer an issue.

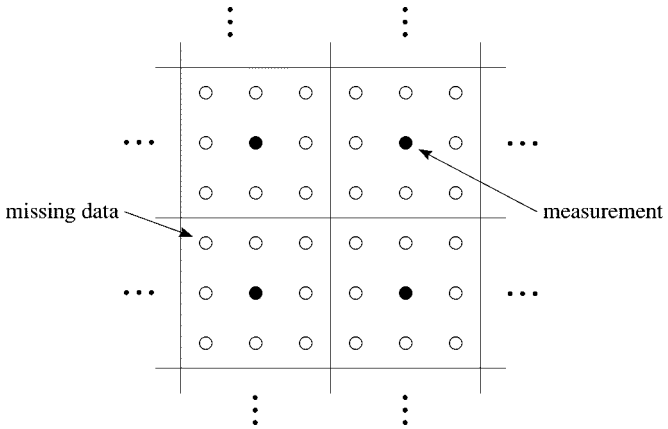

Fig. 17. Diagram showing the locations of missing data in relation to the measurements in our image magnification technique.

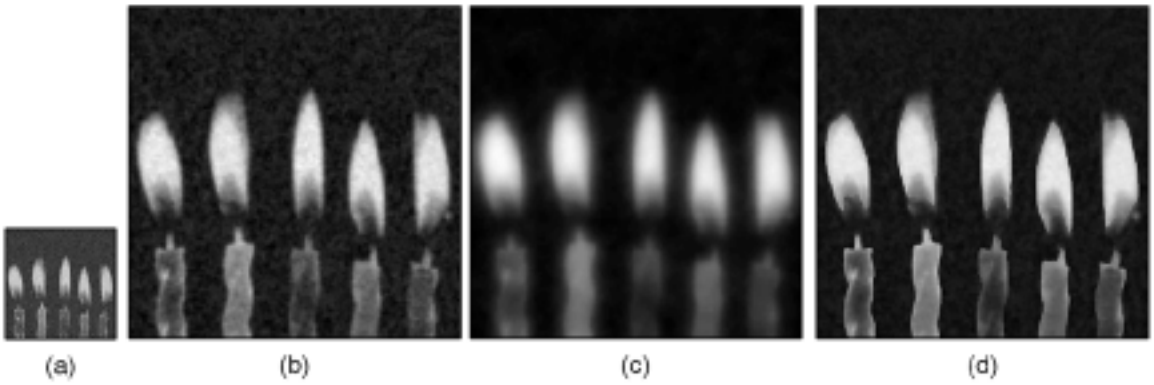

Fig. 18. Three-fold magnification of a photograph of birthday candles: (a) 160×160 original image; (b) 480×480 magnified image obtained by replication then smoothing; (c) 480×480 magnified image obtained by smoothing then replication; (d) 480×480 magnified image based on the Mumford–Shah model.

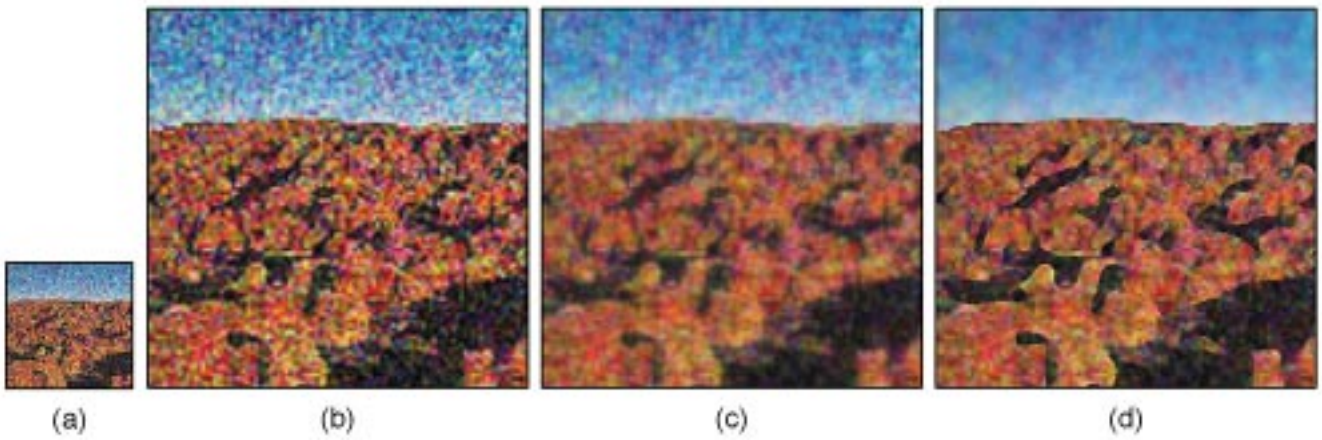

Fig. 19. Three-fold magnification of color photograph of a canyon region in Australia: (a) 100×100 original image; (b) 300×300 magnified image obtained by bilinear interpolation then smoothing; (c) 300×300 magnified image obtained by smoothing then bilinear interpolation; (d) 300×300 magnified image based on the Mumford–Shah model.

Image magnification capability is weaved into the Mumford–Shah active contour model by considering the image magnification problem as a very structured case of the missing data problem. Specifically, consider a new grid with three times as many pixels in each direction and assign the value of the original image to the "center" pixel in each 3×3 block on the grid and treat the remaining pixels as missing data points (see Fig. 17). From an estimation-theoretic standpoint, we can view these "center" pixels as sparse measurements on a much larger image domain. We then employ our generalized Mumford–Shah curve evolution procedure to interpolate to this finer grid, using the curve evolution portion of this procedure to partition the domain of the magnified image into different homogeneous subregions so as to provide smooth interpolations where appropriate without blurring across regions of high contrast.

In Fig. 18(a), we show a 160×160 noisy black-and-white photograph of 5 burning birthday candles, each of differing intensity. We show in Fig. 18(b) the image obtained by first magnifying the original noisy image using zero-order hold then smoothing it isotropically. Notice the magnified image is still noisy because the noise components within the original image have been exaggerated by the zero-order interpolation scheme. Fig. 18(c) shows the image obtained by first isotropically smoothing the original noisy image then magnifying it using zero-order hold. This image is blurry because the edges of the image were destroyed during the initial smoothing step. We show the magnification results based on our approach in Fig. 18(d). To demonstrate that this magnification technique is not just limited to scalar-valued images with nonoverlapping regions, in Fig. 19, we show the results of the various magnification technique (including ours) on a noisy vector-valued image of a canyon region in Australia [Fig. 19(a)]. Fig. 19(b) is obtained by first magnifying the original image using bilinear interpolation followed by isotropic smoothing while Fig. 19(c) is obtained by first isotropically smoothing the original image then magnifying it using bilinear interpolation. For comparison, the hierarchical implementation of the vector-valued counterpart of (17) is used to obtain our magnification result shown in Fig. 19(d).

V. Summary and Further Research Directions

In this paper, we have outlined an estimation-theoretic approach to curve evolution based on the Mumford–Shah functional. By viewing an active contour as the set of discontinuities in

the standard Mumford–Shah problem, we used the corresponding functional to determine gradient descent evolution equations to deform the active contour. Each gradient descent step involved solving a corresponding optimal estimation problem, namely the optimal estimate of the noise-free image given the noisy image data and the current estimate of the boundary curve. The solution of this estimation problem came from the theory of boundary-value stochastic processes, which leads to decoupled PDEs in space whose solutions produce the optimal image estimates in each of the connected regions separated by the current curve estimate. Very importantly, this theory also gave us boundary conditions for these estimates along the current estimate of the boundary curve which are directly used in evolving the curve to a local minimum of the Mumford–Shah functional. By connecting curve evolution and the Mumford–Shah functional with the theory of boundary-value stochastic processes, our algorithm can be regarded as a curve evolution driven by solutions of a continuum of auxiliary spatial estimation problems.

We have also demonstrated that our approach extends well beyond the basic Mumford–Shah formulation including two important applications in which data quality is spatially-varying or in which sets of pixel measurements are missing throughout the image. The segmentation and restoration of these missing data problems is handled seamlessly in our estimation-theoretic/curve evolution framework. In addition, our technique is also applicable to the image magnification problem by considering it as a special case of the missing data problem in which the missing data occur in a very structured manner. This segmentation-based approach for image magnification is more global, is much less susceptible to blurring or blockiness artifacts as compared to other more traditional techniques, and has the additional attractive denoising capability.

Direct implementation of our curve evolution starting from an initial curve which is far from the optimal curve can lead to very slow convergence and substantial computation associated with the estimation PDEs to be solved (or at least approximated) at each step of the evolution. In this paper, we outlined several approaches to obtain a fast and efficient implementation of our algorithm that is also capable of handling important image features such as triple-points and other multiple junctions without having to resort to sophisticated level set methods [21].

Our algorithm suggests possible extensions to the Mumford–Shah model. In particular the estimation-theoretic interpretation of the Mumford–Shah functional as specifying particular random field models within each region segmented by the curve opens up the idea of replacing the specific random field models in the Mumford–Shah formulation with other models more appropriate to other contexts. One such application is that of surface and shape estimation in which the so-called thin-plate prior model (involving the norm-squared of the image Laplacian rather than the image gradient) is often used. The generalization of our formalism to such a prior, which will allow the simultaneous segmentation and shape estimation of multiple objects and regions in an image, is currently under investigation.

While we have not pursued it in this paper, one of the potential advantages of our estimation-theoretic formalism is that we can also use the same formalism to compute statistics (e.g., covariances) of the errors in these estimates. These statistics can then be used for a variety of purposes. For example statistical optimal fusion of estimates from multiple images (as is commonly encountered in remote sensing) requires error covariances in order to weight each of the images to be fused according to its quality. Also, error variances can be used to detect and correct for anomalous measurements or to detect changes between the estimate produced from one set of imagery and a later set of imagery of the same region. More generally, since (1) can be interpreted as the log-likelihood of the observed imagery given the random field model implied by the Mumford–Shah functional, error statistics can in principle be used both for model validation and parameter estimation. These topics will be the subject of future research.

ACKNOWLEDGMENT

The authors would like to kindly thank Prof. J. Shah at Northeastern University for some very helpful conversations and for suggesting some of the references. The authors would also like to thank the anonymous reviewers for their valuable comments.

REFERENCES

[1] M. Adams, A. Willsky, and B. Levy, "Linear estimation of boundary value stochastic processes—Part 1: The role and construction of complementary models," *IEEE Trans. Automat. Contr.*, vol. 29, pp. 803–811, 1984.

[2] J. Allebach and W. Wong, "Edge directed interpolation," in *IEEE Int. Conf. Image Processing*, vol. 3, 1996, pp. 707–711.

[3] L. Ambrosio and V. M. Tortorelli, "Approximation of functionals depending on jumps by elliptic functionals via Γ-convergence," *Commun. Pure Appl. Math*, vol. 43, no. 8, 1990.

[4] M. Bertalmio, G. Sapiro, V. Caselles, and C. Ballester, "Image inpainting," in *Proc. SIGGRAPH 2000*, July 2000.

[5] A. Blake and A. Zisserman, *Visual Reconstruction*. Cambridge, MA: MIT Press, 1987.

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

[7] V. Caselles, R. Kimmel, and G. Sapiro, "Geodesic snakes," *Int. J. Comput. Vis.*, 1998.

[8] T. Chan and J. Shen, "Non-texture inpainting by curvature-driven diffusions,", UCLA Tech. Rep., 2000.

[9] T. Chan and L. Vese, "Active contours without edges," *IEEE Trans. Image Processing*, submitted for publication.

[10] ——, "A level set algorithm for minimizing the Mumford–Shah functional in image processing," Univ. California, Los Angeles, Tech. Rep., 2000.

[11] ——, "An efficient variational multiphase motion for the Mumford–Shah segmentation model," in *Proc. Asilomar Conf. Signals, Systems, Computers*, Nov. 2000.

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

[13] A. I. El-Fallah and G. E. Ford, "Mean curvature evolution and surface area scaling in image filtering," *IEEE Trans. Image Processing*, vol. 6, pp. 750–753, May 1997.

[14] M. Kass, A. Witkin, and D. Terzopoulos, "Snakes: Active contour models," *Int. J. Comput. Vis.*, 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. Ration. Mech. Anal.*, vol. 134, pp. 275–301, 1996.

[16] B. B. Kimia and K. Siddiqi, "Geometric heat equation and nonlinear diffusion of shapes and images," in *Proc. IEEE Conf. Computer Vision Pattern Recognition*, 1994, pp. 113–120.

[17] R. Kimmel, "Affine differential signatures for gray level images of planar shapes," in *Proc. Int. Conf. Computer Vision*, 1996.

[18] J. Lim, *Two-Dimensional Signal and Image Processing*. Englewood Cliffs, NJ: Prentice-Hall, 1992.

[19] 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, Feb. 1995.
[20] S. Masnou and J. Morel, "Level lines based disocclusion," *Proc. IEEE Int. Conf. Image Processing*, vol. 3, pp. 259–263, 1998.
[21] B. Merriman, J. Bence, and S. Osher, "Motion of multiple junctions: A level set approach," *J. Comput. Phys.*, vol. 112, no. 2, pp. 334–363, 1994.
[22] D. Mumford and J. Shah, "Optimal approximations by piecewise smooth functions and associated variational problems," *Commun. Pure Appl. Math.*, vol. 42, no. 4, 1989.
[23] ——, "Boundary detection by minimizing functionals," *Proc. IEEE Conf. Computer Vision Pattern Recognition*, 1985.
[24] M. Nitzberg, D. Mumford, and T. Shiota, *Filtering, Segmentation and Depth*, Berlin, Germany: Springer-Verlag, 1993.
[25] S. Osher, "Riemann solvers, the entropy condition, and difference approximations," *SIAM J. Numer. Anal.*, vol. 21, pp. 217–235, 1984.
[26] S. Osher and J. Sethian, "Fronts propagation with curvature dependent speed: Algorithms based on Hamilton-Jacobi formulations," *J. Comput. Physics*, vol. 79, pp. 12–49, 1988.
[27] N. Paragios and R. Deriche, "Geodesic active regions for texture segmentation," INRIA, France, Res. Rep. 3440, 1998.
[28] P. Perona, "Orientation diffusions," *IEEE Trans. Image Processing*, vol. 7, pp. 457–467, Mar. 1998.
[29] P. Perona and J. Malik, "Scale-space and edge detection using anisotropic diffusion," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 12, pp. 629–639, July 1990.
[30] P. Perona, T. Shiota, and J. Malik, "Anisotropic diffusion," in *Geometry-Driven Diffusion in Computer Vision*, B. M. ter Harr Romeny, Ed. Norwell, MA: Kluwer, 1994.
[31] K. Ratakonda and N. Ahuja, "POC based adaptive image magnification," in *IEEE Int. Conf. Image Processing*, vol. 3, 1998, pp. 203–207.
[32] T. J. Richardson, "Scale independent piecewise smooth segmentation of images via variational methods," Ph.D. dissertation, Dept. Elect. Eng. Comput. Sci., Mass. Inst. Technol., 1989.
[33] B. M. ter Harr Romeny, Ed., *Geometry-Driven Diffusion in Computer Vision*. Norwell, MA: Kluwer, 1994.
[34] R. Ronfard, "Region-based strategies for active contour models," *Int. J. Comput. Vis.*, vol. 13, pp. 229–251, 1994.
[35] G. Sapiro and A. Tannenbaum, "Area and length preserving geometric invariant scale-space," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 17, pp. 67–72, Jan. 1995.
[36] J. Sethian, *Level Set Methods: Evolving Interfaces in Geometry, Fluid Mechanics, Computer Vision, and Material Science*. Cambridge, U.K.: Cambridge Univ. Press, 1996.
[37] J. Shah, "A common framework for curve evolution, segmentation and anisotropic diffusion," *Proc. IEEE Conf. Computer Vision Pattern Recognition*, 1996.
[38] 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, Mar. 1998.
[39] G. Sapiro and D. L. Ringarch, "Anisotropic diffusion of multivalued images with applications to color filtering," *IEEE Trans. Image Processing*, vol. 5, pp. 1582–1585, Nov. 1996.
[40] N. Sochen, R. Kimmel, and R. Malladi, "A general framework for low level vision," *IEEE Trans. Image Processing*, vol. 7, pp. 310–318, Feb. 1998.
[41] H. Tek and B. Kimia, "Image segmentation by reaction diffusion bubbles," in *Proc. Int. Conf. Computer Vision*, 1995, pp. 156–162.
[42] D. Terzopoulos and A. Witkin, "Constraints on deformable models: Recovering shape and nonrigid motion," *Artif. Intell.*, vol. 36, pp. 91–123, 1988.
[43] A. Tsai, A. Yezzi, and A. Willsky, "A curve evolution approach to smoothing and segmentation using the Mumford-Shah functional," in *Proc. IEEE Conf. Computer Vision Pattern Recognition*, June 2000.
[44] J. Weickert, "Scale-space properties of nonlinear diffusion filtering with a diffusion tensor," Lab. Technomath., Univ. Kaiserslautern, Germany, Tech. Rep. 110, 1994.
[45] R. Whitaker and G. Gerig, "Vector-valued diffusion," in *Geometry Driven Diffusion in Computer Vision*, B. ter Haar Romeny, Ed. Norwell, MA: Kluwer, 1994, pp. 93–133.
[46] A. Yezzi, "Modified curvature motion for image smoothing and enhancement," *IEEE Trans. Image Processing*, vol. 7, pp. 345–352, Mar. 1998.
[47] A. Yezzi, S. Kichenassamy, A. Kumar, P. Olver, and A. Tannenbaum, "A geometric snake model for segmentation of medical imagery," *IEEE Trans. Med. Imag.*, vol. 16, pp. 199–209, Apr. 1997.
[48] A. Yezzi, A. Tsai, and A. Willsky, "A statistical approach to snakes for bimodal and trimodal imagery," in *Proc. Int. Conf. Computer Vision*, 1999.
[49] ——, "A fully global approach to image segmentation via coupled curve evolution equations," *J. Vis. Commun. Image Represent.*, submitted for publication.
[50] Y. L. You, W. Zhu, A. Tannenbaum, and M. Kaveh, "Behavioral analysis of anisotropic diffusion," *IEEE Trans. Image Processing*, vol. 5, pp. 1539–1553, Nov. 1996.
[51] S. C. Zhu, T. S. Lee, and A. L. Yuille, "Region competition: Unifying snakes, region growing, and Bayes/MDL for multiband image segmentation," in *Proc. Int. Conf. Computer Vision*, 1995, pp. 416–423.
[52] S. Zhu and A. Yuille, "Region competition: Unifying snakes, region growing, and Bayes/MDL for multiband image segmentation," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 18, pp. 884–900, Sept. 1996.

Andy Tsai received the B.S. degree in electrical engineering from the University of California, San Diego, in 1993, and the S.M., E.E., and Ph.D. degrees from the Massachusetts Institute of Technology (MIT), Cambridge, in 1995, 1999, and 2000, respectively. He is currently pursuing the M.D. degree at Harvard Medical School, Boston, MA.

He was a Postdoctoral Research Associate with the Laboratory for Information and Decision Systems, MIT. His research interests are computer vision and statistical image processing.

Anthony Yezzi, Jr. (M'99) was born in Gainesville, FL, in 1972 and received the Ph.D. degree in 1997 from the Department of Electrical Engineering, University of Minnesota, Minneapolis.

After completing a postdoctoral research position in the Laboratory for Information and Decision Systems (LIDS), Massacusetts Institute of Technology, Cambridge, he began his current position at Georgia Institute of Technology as an Assistant Professor in 1999. He has consulted for a number of medical imaging companies including GE, Picker, and VTI. His research lies primarily within the fields of image processing and computer vision. His work within these fields includes anisotropic diffusion for image smoothing and adaptive thresholding, edge-detection, active contours, segmentation, multiframe shape from shading and stereo matching, and shape analysis. His work in anisotropic smoothing and segmentation has been largely motivated and directed toward problems in medical imaging applied to MRI, ultrasound, CT, and OCT modalities. Two central themes of his research in general are curve/surface evolution theory from differential geometry and partial differential equations.

Alan S. Willsky (S'70–M'73–SM'82–F'86) received the S.B. and Ph.D. degrees from the Massachusetts Institute of Technology (MIT), Cambridge, in 1969 and 1973 respectively.

He joined the MIT faculty in 1973 and his present position is as the Edwin S. Webster Professor of electrical engineering. From 1974 to 1981, he served as Assistant Director of the MIT Laboratory for Information and Decision Systems. He is also a Founder and member of the board of directors of Alphatech, Inc., and is currently a member of the U.S. Air Force Scientific Advisory Board. He has held visiting positions at Imperial College, London, U.K., L'Université de Paris-Sud, Paris, France, and the Institut de Recherche en Informatique et Systémes Aléatoires, Rennes, France. He is the author of the research monograph *Digital Signal Processing and Control and Estimation* and is the co-author of the undergraduate text *Signals and Systems*. His research interests are in the development and application of advanced methods of estimation and statistical signal and image processing. Methods he has developed have been successfully applied in a wide variety of applications including failure detection in high-performance aircraft, advanced surveillance and tracking systems, electrocardiogram analysis, computerized tomography, and remote sensing.

Dr. Willsky was awarded the 1979 Alfred Nobel Prize by the ASCE and the 1980 Browder J. Thompson Memorial Prize Award by the IEEE for a paper excerpted from his monograph. In 1975, he received the Donald P. Eckman Award from the American Automatic Control Council. He was program chairman for the 17th IEEE Conference on Decision and Control, has been an associate editor of several journals and special guest editor for several special issues, and has served as a member of the Board of Governors and Vice President for Technical Affairs of the IEEE Control Systems Society. In 1988, he was made a Distinguished Member of the IEEE Control Systems Society. He has given plenary and keynote lectures at a number of major scientific meetings, including the 20th IEEE Conference on Decision and Control, the 1991 IEEE International Conference on Systems Engineering, the 1991 SIAM Conference on Applied Linear Algebra, the 1992 Inaugural Workshop for the National Centre for Robust and Adaptive Systems, Canberra, Australia, the 1992 INRIA 25th Anniversary Symposium in Paris, the 1993 IEEE Symposium on Image and Multidimensional Signal Processing in Cannes, and the 1997 Wavelet Applications in Signal and Image Processing Conference.