A Curve Evolution Approach to Smoothing and Segmentation Using the Mumford-Shah Functional *

Andy Tsai† Anthony Yezzi, Jr.‡ Alan S. Willsky†

† Department of Electrical Engineering and Computer Science
Massachusetts Institute of Technology
Cambridge, MA 02139

‡ School of Electrical and Computer Engineering
Georgia Institute of Technology
Atlanta, GA 30332

Abstract

In this work, we approach the classic Mumford-Shah problem from a curve evolution perspective. In particular, we let a given family of curves define the boundaries between regions in an image within which the data are modeled by piecewise smooth functions plus noise as in the standard Mumford-Shah functional. The gradient descent equation of this functional is then used to evolve the curve. Each gradient descent step involves solving a corresponding optimal estimation problem which connects the Mumford-Shah functional and our curve evolution implementation with the theory of boundary-value stochastic processes. The resulting active contour model, therefore, inherits the attractive ability of the Mumford-Shah technique to generate, in a coupled manner, both a smooth reconstruction of the image and a segmentation as well. We demonstrate applications of our method to problems in which data quality is spatially varying and to problems in which sets of pixel measurements are missing. Finally, we demonstrate a hierarchical implementation of our model which leads to a fast and efficient algorithm capable of dealing with important image features such as triple points.

1. Introduction

In this paper, we propose a novel curve evolution approach to the Mumford-Shah problem of simultaneous image smoothing and segmentation. 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 in the standard Mumford-Shah problem, we may use the corresponding functional to determine gradient descent evolution equations to deform the active contour. Each gradient descent step involves 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 comes from the theory of boundary-value stochastic processes, which leads to decoupled partial differential equations (PDE's) in space whose solutions produce the optimal image estimates in each of the connected regions separated by the current curve estimate. This theory also gives 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.

This active contour model connects curve evolution and the Mumford-Shah functional with the theory of boundary-value stochastic processes. The resulting algorithm can be regarded as a curve evolution driven by solutions of a continuum of auxiliary spatial estimation problems. This work may be regarded as an extension of several recent region based approaches to curve evolution [2, 7, 8]. In particular, this work naturally generalizes the work of Chan and Vese in [2] who consider piecewise constant generalization of the Mumford-Shah functional within a level set framework.

An attractive feature of our model is that it inherits the original property of the Mumford-Shah formulation to segment and smooth images in a coupled manner. As such, our model generates both segmentations and smooth image reconstructions, including applications to problems in which data quality is spatially-varying and to images in which sets of pixel measurements are missing. While related methods require a priori knowledge of the number of region types, such as [2] which assumes exactly two region types with two different mean intensities, or [8] which requires separate curves and statistics to deal with more than two region types, our Mumford-Shah-based approach can automatically segment images with multiple region types (perhaps each with different mean intensities) without knowing a priori how many distinct regions are present in the image.

*This work was supported by ONR grant N00014-91-J-1004 and by AFOSR grant F49620-98-1-0349.

Direct implementation of our curve evolution starting from an initial curve which is far from the optimal curve can lead to slow convergence and substantial computation associated with the estimation PDE's to be solved at each step of the evolution. In this paper, we will therefore outline several approaches to obtain a fast and efficient implementation of our algorithm that is also capable of handling important image features such as multiple junctions without having to resort to sophisticated level set methods [3].

2. The Mumford-Shah formulation as a curve evolution problem

The point of reference for this paper is the Mumford-Shah functional

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

in which g denotes the data, f denotes the piecewise smooth approximating surface, $\vec{C}$ denotes the smooth, closed segmenting curve, and Ω denotes the image domain [4, 5]. The parameters α, β, and γ control the competition between the various terms above and determine the "scale" of the segmentation and smoothing. The Mumford-Shah problem is to minimize $E(f,\vec{C})$ over admissible f and $\vec{C}$. We now present how this problem is solved via a curve-evolution-based approach.

2.1. Linear estimation of boundary-value stochastic processes

Any arbitrary closed curve $\vec{C}$ partitions the domain Ω of the image into two regions R and R^c, corresponding to the interior and exterior of curve $\vec{C}$ respectively. If we fix this partitioning minimizing (1) corresponds to minimizing

$$E_{\vec{C}}(f_R, f_{R^c}) = \beta \iint_R (f_R - g)^2 \, dA + \alpha \iint_R |\nabla f_R|^2 \, dA + \beta \iint_{R^c} (f_{R^c} - g)^2 \, dA + \alpha \iint_{R^c} |\nabla f_{R^c}|^2 \, dA. \qquad (2)$$

The estimates $\hat{\mathbf{f}}_R$ and $\hat{\mathbf{f}}_{R^c}$ that minimize (2) satisfy (decoupled) PDE's which can be obtained using standard variational methods [5]. Alternatively each of these estimates can also be obtained from the theory of optimal estimation. 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 $\frac{1}{\beta}$ and $\frac{1}{\alpha}$, 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 $\boldsymbol{\zeta} = \{\mathbf{v}, \mathbf{w}\}$ (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 damped Poisson equation with Neumann boundary condition for $\hat{f}_R$:

$$\hat{f}_R - \frac{\alpha}{\beta} \nabla^2 \hat{f}_R = g \qquad \text{on } R \qquad (5a)$$

$$\frac{\partial \hat{f}_R}{\partial \vec{\mathcal{N}}} = 0 \qquad \text{on } \vec{C}. \qquad (5b)$$

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

$$\hat{f}_{R^c} - \frac{\alpha}{\beta} \nabla^2 \hat{f}_{R^c} = g \qquad \text{on } R^c \qquad (6a)$$

$$\frac{\partial \hat{f}_{R^c}}{\partial \vec{\mathcal{N}}} = 0 \qquad \text{on } \vec{C}. \qquad (6b)$$

We will refer to equations (5) and (6) as the *estimation PDE's*. The conjugate gradient (CG) method is employed as a fast and efficient solver for these estimation PDE's.

2.2. 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

$$E_{\hat{f}_R, \hat{f}_{R^c}}(\vec{C}) = \beta \iint_R (\hat{f}_R - g)^2 \; dA + \alpha \iint_R |\nabla \hat{f}_R|^2 \, dA + \beta \iint_{R^c} (\hat{f}_{R^c} - g)^2 \; dA + \alpha \iint_{R^c} |\nabla \hat{f}_{R^c}|^2 \, dA + \gamma \oint_{\vec{C}} ds. \qquad (7)$$

The first four terms in (7) are of the form:

$$J = \iint_D \boldsymbol{\mathcal{H}} \, dA \qquad (8)$$

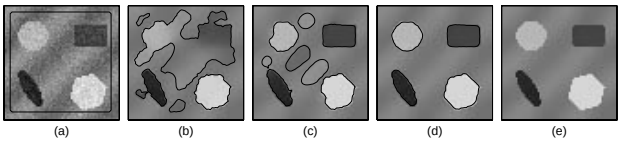


Figure 1. Four objects over spatially varying background.

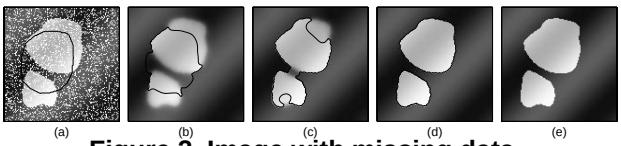


Figure 2. Image with missing data.

where D denotes either the interior or the exterior of $\vec{C}$, and $\mathcal{H} : \mathbf{R}^2 \rightarrow \mathbf{R}$ is a continuous function. The gradient flow to minimize (8) is given by

$$\vec{C}_t = -\mathcal{H}\vec{\mathcal{N}}. \tag{9}$$

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

$$\vec{C}_t = -\kappa\vec{\mathcal{N}} \tag{10}$$

where κ denotes the signed curvature of $\vec{C}$. Knowing gradient flows (9) and (10), the curve evolution that minimizes (7) is given by

$$\begin{aligned}\vec{C}_t &= \frac{\alpha}{2}\left(|\nabla \hat{f}_{R^c}|^2 - |\nabla \hat{f}_R|^2\right)\vec{\mathcal{N}} \\ &\quad + \frac{\beta}{2}\left((g-\hat{f}_{R^c})^2 - (g-\hat{f}_R)^2\right)\vec{\mathcal{N}} - \gamma\kappa\vec{\mathcal{N}}.\end{aligned} \tag{11}$$

This flow is implemented via the level set method [6] 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. Furthermore this flow together with the optimal estimation PDE's makes explicit the coupling between the optimal estimates and the curve evolution.

The class of imagery that our algorithm can handle is not restricted to just images with two distinct means but also images with multiple regions each having different means without knowing in advance the number of such regions or distinct means are present. As shown in Figure 1, segmentation and smoothing are performed on a noisy synthetic image with four objects of different means situated on a spatially varying background region. Multiple regions are captured by a single contour demonstrating the topological transitions allowed by the model's level set implementation.

3. Generalizations and extensions

We now extend the previous model to handle images with missing data; we do this through the parameter β. In the basic Mumford-Shah formulation (1), β is a 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 $\boldsymbol{\beta}$ whose value at each pixel is inversely proportional to the variance of the measured noise at that pixel. By introducing this spatially varying $\boldsymbol{\beta}$, equation (1) becomes:

$$\begin{aligned}E(\mathbf{f},\vec{C}) &= \iint_{\Omega}\boldsymbol{\beta}(\mathbf{f}-\mathbf{g})^2\,dA \\ &\quad + \alpha\iint_{\Omega\setminus\vec{C}}|\nabla\mathbf{f}|^2\,dA + \gamma\oint_{\vec{C}}ds.\end{aligned} \tag{12}$$

The gradient flow that minimizes (12) is given by

$$\begin{aligned}\vec{C}_t &= \frac{\alpha}{2}\left(|\nabla \hat{f}_{R^c}|^2 - |\nabla \hat{f}_R|^2\right)\vec{\mathcal{N}} \\ &\quad + \frac{\boldsymbol{\beta}}{2}\left((g-\hat{f}_{R^c})^2 - (g-\hat{f}_R)^2\right)\vec{\mathcal{N}} - \gamma\kappa\vec{\mathcal{N}}\end{aligned} \tag{13}$$

where, using the theory of boundary-value processes, we find that the optimal estimates $\hat{\mathbf{f}}_R$ and $\hat{\mathbf{f}}_{R^c}$ of (13) satisfy

$$\begin{aligned}\boldsymbol{\beta}\hat{\mathbf{f}}_R - \alpha\nabla^2\hat{\mathbf{f}}_R &= \boldsymbol{\beta}\mathbf{g} \qquad \text{on } R \\ \frac{\partial\hat{\mathbf{f}}_R}{\partial\vec{\mathcal{N}}} &= 0 \qquad \text{on } \vec{C}\end{aligned}$$

and

$$\begin{aligned}\boldsymbol{\beta}\hat{\mathbf{f}}_{R^c} - \alpha\nabla^2\hat{\mathbf{f}}_{R^c} &= \boldsymbol{\beta}\mathbf{g} \qquad \text{on } R^c \\ \frac{\partial\hat{\mathbf{f}}_{R^c}}{\partial\vec{\mathcal{N}}} &= 0 \qquad \text{on } \vec{C}.\end{aligned}$$

This ability to handle missing data is illustrated in Figure 2.

4. Implementation

There are two ways to improve the Mumford-Shah active contour model we presented so far. One, we can speed up the convergence of the algorithm by reducing the required number of curve evolution steps and/or reducing the computational cost of solving the estimation PDE's. 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 [3]). In this section, we present several progressive modifications to the implementation of the Mumford-Shah active contour model in order to make these improvements.

4.1. 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 equation (1). This approach consists of alternating between these 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 PDE's–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 $\mathbf{f}_{R^c}$. We illustrates the use of this approximate gradient descent implementation of our model to segment and smooth a noisy microscopic image of red blood cells shown in Figure 3. In Figure 4, we demonstrate the capabilities of our method in handling incomplete data by using a forward-looking infrared (FLIR) image of three tanks. The missing data in the image are due to intensity saturated and defective pixels of the infrared sensor.

4.2. The 2-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 [2] referred to earlier. Chan and Vese restricts 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 model. This reduces flow (11) to

$$\vec{C}_t = \frac{\beta}{2}(u - v)(g - u + g - v)\vec{\mathcal{N}} - \gamma\kappa\vec{\mathcal{N}} \qquad (14)$$

where u and v are the average intensities of R and R^c, respectively. This is precisely the flow presented in [2]. 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, 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 (14) in calculating the initial estimate of $\vec{C}$, this 2-step implementation approach of our Mumford-Shah active contour model can only handle images with two distinct means. In Section 4.4, we discuss ways to circumvent this restriction.

Figure 5 illustrates this 2-step implementation of our model. A noisy mammogram showing a cyst in the breast tissue is displayed in Figure 5(a). The starting curve is shown in Figure 5(b) with the next frame showing the initial estimate of $\vec{C}$ obtained by assuming piecewise constant regions; that is, obtained by employing flow (14). This curve is superimposed on top of the piecewise constant approximation of the image. In Figure 5(d), we superimpose this initial estimate of $\vec{C}$ on to the noisy mammogram image to illustrate the proximity of this initial curve in outlining the cyst within the mammogram. Figure 5(e) shows the final results of applying the approximate gradient descent implementation of our active contour model. It is clear from comparing the results shown in frames (d) and (e) that the segmentation of the cyst has been refined and a smooth estimate of the image is obtained. For comparison to the original image, we show the piecewise smooth approximation of the image in Figure 5(f) without the segmenting curve.

4.3. The multiresolution approach

Though the 2-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 non-overlapping 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 PDE's 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. 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 towards the edges of the image is small. 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.

We demonstrate this multiresolution approach on a 200×200 noisy picture (Figure 6(a)) consisting of two different foreground regions (i.e., the two pearl mushrooms of differing intensities). Figure 6(b) shows the initialization scheme at the coarsest scale (40×40 image) and Figure 6(c) shows the segmentation and reconstruction of the noisy image at that scale. Figure 6(d) shows the 200×200 original image with the curve obtained from the up sampled version of the curve from Figure 6(c) for visual comparison to the final segmenting curve shown in Figure 6(e). Figure 6(f) shows the reconstruction of the original image without the segmenting curve. As expected, a single contour successfully captured the two differing foreground regions.

4.4. The hierarchical approach

Finally, we propose an implementation of our active contour model, building on the preceding modifications, to enable our model to handle images with multiple junctions without resorting to more sophisticated level set techniques [3]. This approach also allows the possibility of using the 2-step approach of Section 4.2 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 (implemented either by approximate gradient descent, the 2-step approach, or the multiresolution approach). 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 implementation 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.

We use the brain pathology image shown in Figure 7(a) to demonstrate how the approximate gradient descent method of Section 4.1 is used in the hierarchical implementation of our active contour model. First, the segmentation and reconstruction of the image shown in Figure 7(c) is obtained based on the initialization curve shown in Figure 7(b) and implemented using the approximate gradient descent method. 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, using the initialization scheme shown in Figure 7(d), to obtain the segmentation and reconstruction of the image shown in Figure 7(e). The black and the white curves segment the image into the background, white matter, and the gray matter. For comparison, the piecewise smooth reconstruction of the image is shown in Figure 7(f) without the segmenting curves.

5. Conclusions

In this paper, we introduced a novel active contour model which ties together the theories of curve evolution, boundary-value stochastic processes, and the Mumford-Shah functional. The key attraction of this model is its ability to simultaneous segment and smooth images in a single framework and to be able to handle images with missing data. The various implementations of this model we outlined greatly increase the algorithm's convergence and extend its applicability to images with multiple junctions.

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. Automatic Control*, 29:803–811, 1984.

[2] T. Chan and L. Vese. Active contours without edges. *UCLA Technical Report*, 1999.

[3] B. Merriman, J. Bence, and S. Osher. Motion of multiple junctions: A level set approach. *Journal of Computational Physics*, 112:334–363, 1994.

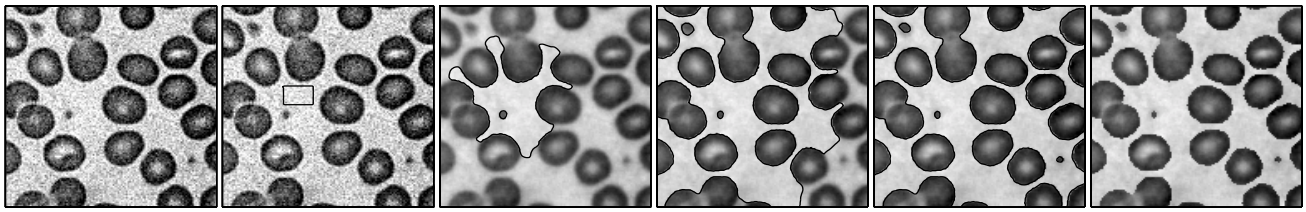
[4] D. Mumford and J. Shah. Boundary detection by minimizing functionals. *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 1985.

[5] D. Mumford and J. Shah. Optimal approximations by piecewise smooth functions and associated variational problems. *Communications in Pure and Applied Mathematics*, 42, 1989.

[6] S. Osher and J. Sethian. Fronts propagation with curvature dependent speed: Algorithms based on hamilton-jacobi formulations. *Journal of Computational Physics*, 79:12–49, 1988.

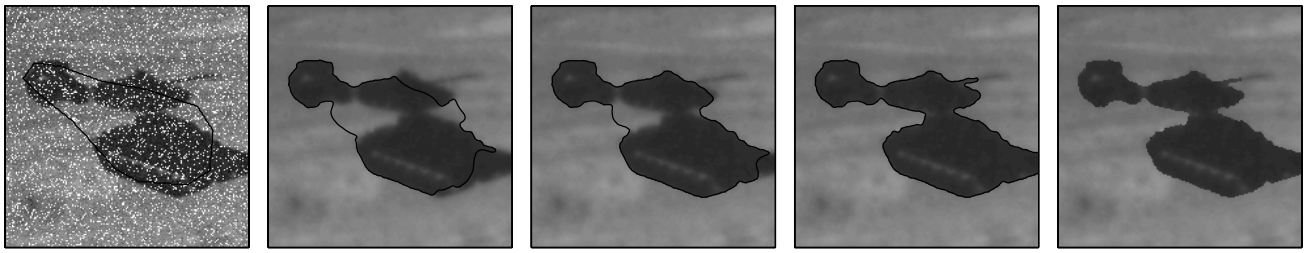
[7] N. Paragios and R. Deriche. Geodesic active regions for texture segmentation. *Research Report 3440*, 1998.

[8] A. Yezzi, A. Tsai, and A. Willsky. A statistical approach to snakes for bimodal and trimodal imagery. *Proc. IEEE Int. Conf. on Computer Vision*, 1999.



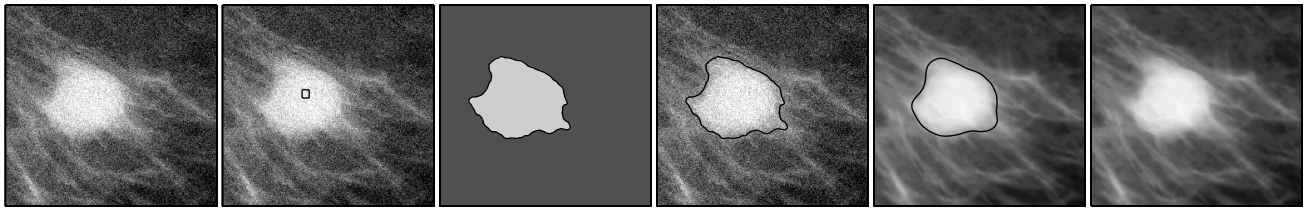
(a) (b) (c) (d) (e) (f)

Figure 3. Microscopic image of red blood cells.



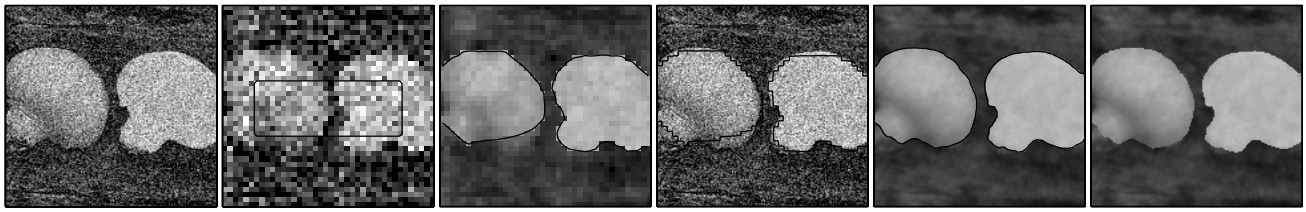
(a) (b) (c) (d) (e)

Figure 4. FLIR image of M2, T62, and M60 tanks. White pixels in (a) denote missing data locations.



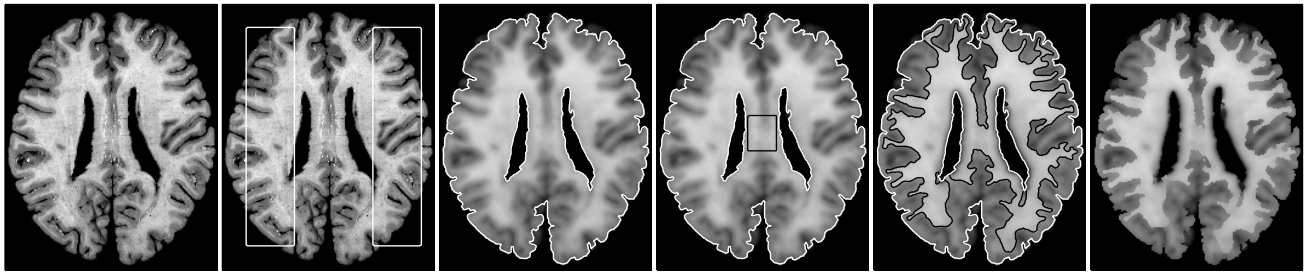
(a) (b) (c) (d) (e) (f)

Figure 5. Mammogram showing cyst in the breast tissue.



(a) (b) (c) (d) (e) (f)

Figure 6. Two pearl mushrooms on the bark of a tree.



(a) (b) (c) (d) (e) (f)

Figure 7. Axial pathology section of the human brain.