# Effective Level Set Image Segmentation With a Kernel Induced Data Term

Mohamed Ben Salah, Amar Mitiche, *Member, IEEE*, and Ismail Ben Ayed, *Member, IEEE*

***Abstract*—This study investigates level set multiphase image segmentation by kernel mapping and piecewise constant modeling of the image data thereof. A kernel function maps implicitly the original data into data of a higher dimension so that the piecewise constant model becomes applicable. This leads to a flexible and effective alternative to complex modeling of the image data. The method uses an active curve objective functional with two terms: an original term which evaluates the deviation of the mapped image data within each segmentation region from the piecewise constant model and a classic length regularization term for smooth region boundaries. Functional minimization is carried out by iterations of two consecutive steps: 1) minimization with respect to the segmentation by curve evolution via Euler-Lagrange descent equations and 2) minimization with respect to the regions parameters via fixed point iterations. Using a common kernel function, this step amounts to a mean shift parameter update. We verified the effectiveness of the method by a quantitative and comparative performance evaluation over a large number of experiments on synthetic images, as well as experiments with a variety of real images such as medical, satellite, and natural images, as well as motion maps.**

***Index Terms*—Kernel mapping, level set image segmentation, mean shift, multiphase, piecewise constant model.**

## I. Introduction

A central problem in computer vision, image segmentation has been the subject of a considerable number of studies [6]–[8], [10]–[13], [16]. Variational formulations [17], which express image segmentation as the minimization of a functional, have resulted in the most effective algorithms. This is mainly because they are amenable to the introduction of constraints on the solution. Conformity of region data to statistical models and smoothness of region boundaries are typical constraints. The Mumford–Shah variational model [17] is fundamental. Most variational segmentation algorithms minimize a variant of the piecewise constant Mumford–Shah functional

$$\mathcal{F}(M,K)=\int_{\Omega}|I-M|^2d\mathbf{x}+\lambda l(K) \tag{1}$$

where $M:\Omega\rightarrow\mathbb{R}^2$ is a piecewise constant approximation of the observed data $I$, and $K$ is the set of boundary points of $M$.

The piecewise constant image model [6]–[9], [17], [18], and its piecewise Gaussian generalization [3], [19], [20], have been the focus of most studies and applications because the ensuing algorithms reduce to iterations of computationally simple updates of segmentation regions and their model parameters. The more general Weibull model has also been investigated [21]. Although they can be useful, these models are not generally applicable. For instance, synthetic aperture radar (SAR) images, of great importance in remote sensing, require the Rayleigh distribution model [22]–[24] and polarimetric images, common in remote sensing and medical imaging, the Wishart or the complex Gaussian model [25], [26].

The use of accurate models in image segmentation is problematic for several reasons. First, modeling is notoriously difficult and time consuming [27]. Second, models are learned using a sample from a class of images and, therefore, are generally not applicable to the images of a different class. Finally, accurate models are generally complex and, as such, are computationally onerous, more so when the number of segmentation regions is large [25]. An alternative approach, which would not be prone to such problems, would be to transform the image data so that the piecewise constant model becomes applicable. This is typically what kernel functions can do, as several pattern classification studies have shown [28]–[31]. A kernel function maps implicitly the original data into data of a higher dimension so that linear separation algorithms can be applied [39]. This is illustrated in Fig. 1 with a 2-D data example. The mapping is implicit because the dot product, the Euclidean norm thereof, in the higher dimensional space of the transformed data can be expressed via the kernel function without explicit evaluation of the transform. Several studies [28], [29], [32], [33] have shown evidence that the prevalent kernels in pattern classification are capable of properly clustering data of complex structure. In the view that image segmentation is spatially constrained clustering of image data [34], kernel mapping should be quite effective in segmentation of various types of images.

This study investigates level set multiphase image segmentation by kernel mapping and piecewise constant modeling of the image data thereof. The method uses an active curve objective functional containing two terms: an original term which evaluates the deviation of the mapped image data within each segmentation region from the piecewise constant model and a classic length regularization term for smooth region boundaries. Functional minimization is carried out by iterations of two consecutive steps: 1) minimization with respect to the partition by curve evolution via the Euler-Lagrange descent equations and 2) minimization with respect to the regions parameters via fixed

Manuscript received January 07, 2009; revised August 26, 2009. First published September 22, 2009; current version published December 16, 2009. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Jenq-Neng Hwang.

M. B. Salah and A. Mitiche are with the Institut National de la Recherche Scientifique (INRS-EMT), Montréal, QC, H5A 1K6, Canada (e-mail: mitiche@emt.inrs.ca; bensalah@emt.inrs.ca).

I. B. Ayed is with General Electric (GE) Canada, 268 Grosvenor, E5-137, London, N6A 4V2, ON, Canada (e-mail: ismail.benayed@ge.com).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIP.2009.2032940

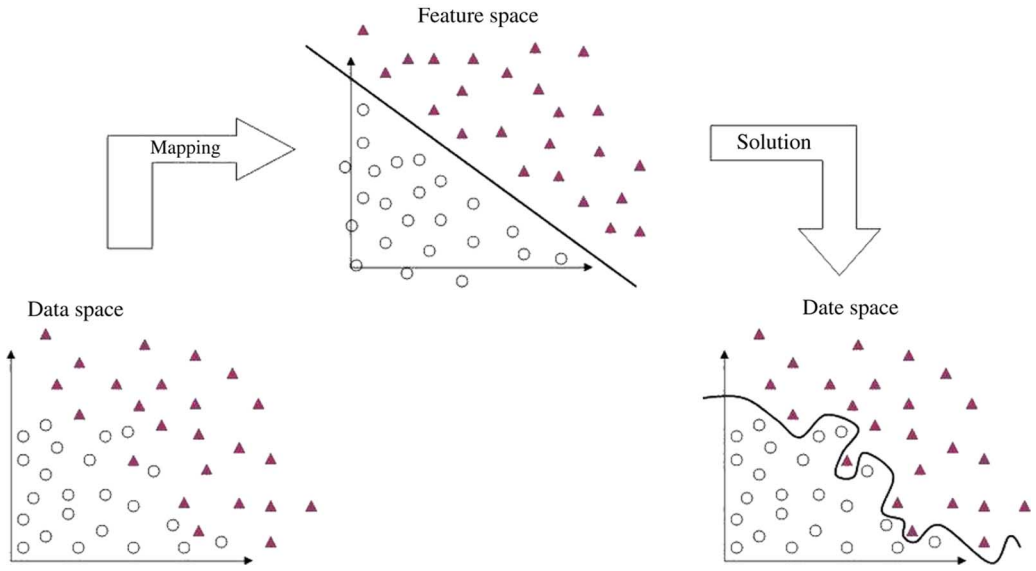



Fig. 1. Illustration of nonlinear 2-D data separation with mapping: The data is non linearly separable in the data space. Mapping the data to a feature (kernel) space and, then, separating it in the induced space with linear methods is possible. For the purpose of display, the feature space in this example is of the same dimension as the original data space. In general, however, the feature space is of higher dimension.

point iteration. The latter leads, interestingly, to a *mean shift* update of the regions parameters. Using a common kernel function, we verified the effectiveness of the method by a quantitative and comparative performance evaluation over a large number of experiments on synthetic images. In comparison to existing level set methods, the proposed method brings advantages with regard to segmentation accuracy and flexibility. To illustrate the flexibility of the method, we also show a representative sample of the tests we ran with various classes of real images including natural images from the Berkeley database, medical and satellite data, as well as motion maps.

The remainder of this paper is organized as follows. The next section reviews the Bayesian framework commonly used in level set segmentation. Section III contains the theoretical contribution. It describes an original kernel-based functional and derives the equations of its minimization in both two-region and multiregion cases. Section IV describes the validation experiments, and Section V contains a conclusion.

## II. Multiphase Image Segmentation

Let $I : \Omega \subset \mathbb{R}^2 \rightarrow \mathcal{I} \subset \mathbb{R}^+$ be an image function. Segmenting $I$ into $N$ regions consists of finding a partition $\{\mathbf{R}_i\}_{i=1}^N$ of the image domain so that each region is homogeneous with respect to some image characteristics commonly given in terms of statistical parametric models. In this case, it is convenient to cast segmentation in a Bayesian framework [12], [13], [25], [35]. The problem would then consist of finding a partition $\{\mathbf{R}_i\}_{i=1}^N$ which maximizes the a posteriori probability over all possible $N$-region partitions of $\Omega$

$$\begin{aligned}\{\hat{\mathbf{R}}_i\}_{i=1}^N &= \arg\max_{\mathbf{R}_i \subset \Omega} P\left(\{\mathbf{R}_i\}_{i=1}^N / I\right) \\ &= \arg\max_{\mathbf{R}_i \subset \Omega} P\left(I / \{\mathbf{R}_i\}_{i=1}^N\right) P\left(\{\mathbf{R}_i\}_{i=1}^N\right). \quad (2)\end{aligned}$$

Assuming that $I(\mathbf{x})$ is independent of $I(\mathbf{y})$ for $\mathbf{x} \neq \mathbf{y}$ and taking $-\log$ of (2), we have

$$\{\hat{\mathbf{R}}_i\}_{i=1}^N = \arg\min_{\mathbf{R}_i \subset \Omega} \mathbf{E}\left[\{\mathbf{R}_i\}_{i=1}^N\right] \quad (3)$$

where

$$\begin{aligned}\mathbf{E}\left[\{\mathbf{R}_i\}_{i=1}^N\right] = \sum_{i=1}^N \int_{\mathbf{x} \in \mathbf{R}_i} -\log P\left(I(\mathbf{x})/\mathbf{R}_i\right) d\mathbf{x} \\ -\log P\left(\{\mathbf{R}_i\}_{i=1}^N\right). \quad (4)\end{aligned}$$

The first term, referred to as the data term, measures the conformity of image data within each region $\mathbf{R}_i, i = 1, \ldots, N$, to a parametric distribution $P(I(\mathbf{x})/\mathbf{R}_i)$. The Gaussian distribution has been the focus of most studies because the ensuing algorithms are computationally simple [13]. The well known piecewise constant segmentation model [2], [6], [8], [34], [35] corresponds to a particular case of the Gaussian distribution. In this case, the data term is expressed as follows:

$$\mathcal{D} = \sum_{i=1}^N \int_{\mathbf{R}_i} e_i(\mathbf{x}) d\mathbf{x} \quad (5)$$

where $e_i(\mathbf{x}) = -\ln P_{\mu_i}(I(\mathbf{x})) = (I(\mathbf{x}) - \mu_i)^2$, and $\mu_i$ is the mean intensity of region $\mathbf{R}_i$. Although used most often, the Gaussian model is not generally applicable. For instance, natural images require more general models [21], and the specific, yet important, SAR and polarimetric images require the Rayleigh and Wishart models [22], [23], [25].

The second term in (4) embeds prior information on the segmentation [13]. The length prior, also called regularization term, is commonly used for smooth segmentation boundaries

$$\mathcal{R} = -\ln\left(P\left(\{\mathbf{R}_i\}_{i=1}^{N}\right)\right) = \lambda \sum_{i=1}^{N} \oint_{\partial \mathbf{R}_i} ds \tag{6}$$

where $\partial \mathbf{R}_i$ is the boundary of the region $\mathbf{R}_i$ and $\lambda$ is a positive factor.

In the next section, we will propose a data term which references the image data transformed via a kernel function, and explain the purpose and advantage of doing so.

## III. Level Set Segmentation in a Kernel-Induced Space

To explain the role of the kernel function in the proposed segmentation functional, and describe clearly the ensuing algorithm, we first treat the case of a segmentation into two regions (Sections III-A and III-B). In Section III-C, a multiregion extension is described.

### A. Two-Region Segmentation

The image data is generally non linearly separable. The basic idea in using a kernel function to transform the image data for image segmentation is as follows: rather than seeking accurate image models and addressing a non linear problem, we transform the image data implicitly via a kernel function so that the piecewise constant model becomes applicable and, therefore, solve a (simpler) linear problem.

Let $\phi(.)$ be a nonlinear mapping from the observation space $\mathcal{I}$ to a higher (possibly infinite) dimensional feature space $\mathcal{J}$. Let $\vec{\gamma}(s) : [0,1] \rightarrow \Omega$ be a closed planar parametric curve. $\vec{\gamma}$ divides the image domain into two regions: the interior of $\vec{\gamma}$ designated by $\mathbf{R}_1 = \mathbf{R}_{\vec{\gamma}}$, and its exterior $\mathbf{R}_2 = \mathbf{R}_{\vec{\gamma}}^c$. Solving the problem of segmentation in the kernel-induced space with curve evolution consists of evolving $\vec{\gamma}$ in order to minimize a functional corresponding to the mapped data. The functional we minimize, $\mathcal{F}_K$, measures a kernel-induced non Euclidean distance between the observations and the regions parameters $\mu_1$ and $\mu_2$ [see (7), shown at the bottom of the page].

In machine learning, the kernel trick [30], [31] consists of using a linear classifier to solve a nonlinear problem by mapping the original nonlinear data into a higher dimensional space. Following the Mercer's theorem [30], which states that any continuous, symmetric, positive semi-definite kernel function can be expressed as a dot product in a high-dimensional space, we do not have to know explicitly the mapping $\phi$. Instead, we can use a kernel function, $K(\mathbf{y},\mathbf{z})$, verifying

$$K(\mathbf{y},\mathbf{z}) = \phi(\mathbf{y})^T \cdot \phi(\mathbf{z}), \quad \forall (\mathbf{y},\mathbf{z}) \in \mathcal{I}^2 \tag{8}$$

where "$\cdot$" is the dot product in the feature space.

Substitution of the kernel functions in the data term yields the following non-Euclidean distance measure in the original data space:

$$\begin{aligned} J_K\left(I(\mathbf{x}),\mu_i\right) &= \left\|\phi\left(I(\mathbf{x})\right) - \phi(\mu_i)\right\|^2 \\ &= \left(\phi\left(I(\mathbf{x})\right) - \phi(\mu_i)\right)^T \cdot \left(\phi\left(I(\mathbf{x})\right) - \phi(\mu_i)\right) \\ &= \phi\left(I(\mathbf{x})\right)^T \cdot \phi\left(I(\mathbf{x})\right) - \phi(\mu_i)^T \cdot \phi\left(I(\mathbf{x})\right) \\ &\quad - \phi\left(I(\mathbf{x})\right)^T \cdot \phi(\mu_i) + \phi(\mu_i)^T \cdot \phi(\mu_i) \\ &= K\left(I(\mathbf{x}),I(\mathbf{x})\right) + K(\mu_i,\mu_i) \\ &\quad - 2K\left(I(\mathbf{x}),\mu_i\right). \end{aligned} \tag{9}$$

To minimize $\mathcal{F}_K$, which depends both on $\vec{\gamma}$ and on the regions parameters $\mu_1$ and $\mu_2$, we adopt an iterative two-step algorithm. The first step consists of fixing the curve and optimizing $\mathcal{F}_K$ with respect to the parameters. As the regularization term does not depend on regions parameters, this is equivalent to optimizing the data term, referred to as $\mathcal{D}_K$. The second step consists of evolving the curve with the parameters fixed.

*1) Step 1:* For a fixed partition of the image domain, the derivatives of $\mathcal{F}_K$ with respect to $\mu_i$, $i \in \{1,2\}$ yield the following equations:

$$\begin{aligned} \frac{\partial \mathcal{F}_K}{\partial \mu_i} &= \frac{\partial \mathcal{D}_K}{\partial \mu_i} \\ &= \int_{\mathbf{R}_1} \frac{\partial J_K\left(I(\mathbf{x}),\mu_1\right)}{\partial \mu_i} d\mathbf{x} + \int_{\mathbf{R}_2} \frac{\partial J_K\left(I(\mathbf{x}),\mu_2\right)}{\partial \mu_i} d\mathbf{x} \\ &= \int_{\mathbf{R}_i} \frac{\partial}{\partial \mu_i}\left[K(\mu_i,\mu_i) - 2K\left(I(\mathbf{x}),\mu_i\right)\right] d\mathbf{x}. \end{aligned} \tag{10}$$

Table I lists some common kernel functions. In all our experiments we used the radial basis function (RBF) kernel, a kernel which has been prevalent in pattern data clustering [28], [42], [43]. With an RBF kernel, the necessary conditions for a minimum of $\mathcal{F}_K$ with respect to region parameters are

$$\mu_i - g_{\mathbf{R}_i}(\mu_i) = 0, \quad i \in \{1,2\} \tag{11}$$

$$\begin{aligned} &\mathcal{F}_K(\mathbf{R}_1,\mathbf{R}_2,\mu_1,\mu_2) \\ &\quad = \underbrace{\int_{\mathbf{R}_1} \left\|\phi\left(I(\mathbf{x})\right) - \phi(\mu_1)\right\|^2 d\mathbf{x} + \int_{\mathbf{R}_2} \left\|\phi\left(I(\mathbf{x})\right) - \phi(\mu_2)\right\|^2 d\mathbf{x}}_{\text{Data term}} \\ &\qquad + \underbrace{\lambda \oint_{\vec{\gamma}} ds}_{\text{Regularization}}. \end{aligned} \tag{7}$$

TABLE I
EXAMPLES OF PREVALENT KERNEL FUNCTIONS

| | |
|---|---|
| RBF Kernel | $K(\mathbf{y},\mathbf{z}) = \exp(-\|\mathbf{y}-\mathbf{z}\|^2/\sigma^2)$ |
| Polynomial Kernel | $K(\mathbf{y},\mathbf{z}) = (\mathbf{y}\cdot\mathbf{z}+c)^d$ |
| Sigmoid Kernel | $K(\mathbf{y},\mathbf{z}) = \tanh(c(\mathbf{y}\cdot\mathbf{z})+\theta)$ |

where

$$g_{\mathbf{R}_i}(\mu_i) = \frac{\int_{\mathbf{R}_i} I(\mathbf{x}) K\left(I(\mathbf{x}),\mu_i\right) d\mathbf{x}}{\int_{\mathbf{R}_i} K\left(I(\mathbf{x}),\mu_i\right) d\mathbf{x}}, \quad i \in \{1,2\}. \tag{12}$$

The solution of (11) can be obtained by fixed point iterations. This consists of iterating

$$\mu_{i,n+1} = g_{\mathbf{R}_i}(\mu_{i,n}), \quad i \in \{1,2\}, n = 1,2,\ldots \tag{13}$$

For $i \in \{1,2\}$, sequence $\{\mu_{i,n}\}_{n=1,2,\ldots}$ converges. A detailed proof is given in Appendix A. Let $\mu_i^{opt}$ be its limit. Thus, $\mu_i^{opt}$ is a fixed point of function $g$ and, consequently, is a solution of (11).

The update of the region parameters obtained in (13) is a *mean-shift* update. Mean-shift corrections have traditionally appeared in data clustering and have been quite efficient [36], [37]. It is a mode search procedure which seeks the stationary points of the data distribution. It is quite interesting that a mean-shift correction appears in this context of active curve segmentation. This correction occurs in the minimization with respect to the region parameters due to the kernel induced data term, via the RBF kernel. The effectiveness and flexibility of this kernel formulation and the ensuing mean-shift update will be confirmed by an extensive experimentation in Section IV.

*2) Step 2:* With the region parameters fixed, this step consists of minimizing $\mathcal{F}_K$ with respect to $\vec{\gamma}$. The Euler-Lagrange descent equation corresponding to $\mathcal{F}_K$ is derived by embedding the curve $\vec{\gamma}$ into a family of one-parameter curves $\vec{\gamma}(s,t)$ : $[0,1]\times\mathbb{R}^+ \to \Omega$ and solving the following partial differential equation:

$$\frac{d\vec{\gamma}}{dt} = -\frac{\partial \mathcal{F}_K}{\partial \vec{\gamma}} \tag{14}$$

where $\partial\mathcal{F}_K/\partial\vec{\gamma}$ is the functional derivative of $\mathcal{F}_K$ with respect to $\vec{\gamma}$. Segmentation regions $\mathbf{R}_1$ and $\mathbf{R}_2$ are obtained from curve $\vec{\gamma}$ at convergence, i.e., when time $t \to \infty$.

Using the result in [12] which shows that, for a scalar function $f$, the functional derivative with respect to the curve $\vec{\gamma}$ of $\int_{\mathbf{R}_{\vec{\gamma}}} f(\mathbf{x})d\mathbf{x}$ is equal to $f(\mathbf{x})\vec{n}(\mathbf{x})$, where $\vec{n}$ is the outward unit normal to $\vec{\gamma}$, we have

$$\frac{\partial \int_{\mathbf{R}_i} J_K\left(I(\mathbf{x}),\mu_i\right) d\mathbf{x}}{\partial \vec{\gamma}} = J_K\left(I(\mathbf{x}),\mu_i\right)\vec{n}(\mathbf{x}), \quad i \in \{1,2\}. \tag{15}$$

The derivative of the length prior with respect to $\vec{\gamma}$ is [12]

$$\frac{\partial}{\partial \vec{\gamma}} \oint_{\vec{\gamma}} ds = \kappa\vec{n} \tag{16}$$

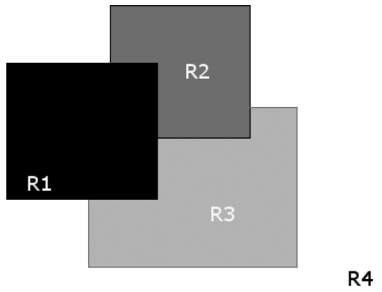



Fig. 2. Representation of a 4-region partition.

where $\kappa$ is the mean curvature function of $\vec{\gamma}$. The final evolution equation for a two-region segmentation in the kernel-induced space is

$$\frac{d\vec{\gamma}}{dt} = \left(J_K\left(I(\mathbf{x}),\mu_1\right) - J_K\left(I(\mathbf{x}),\mu_2\right) - \lambda\kappa\right)\vec{n}. \tag{17}$$

In the case of an RBF kernel, the expression (9) of $J_K(I,\mu_i)$ simplifies to $2(1-K(I,\mu_i))$, $i = 1, 2$.

### *B. Level Set Implementation*

To implement the curve evolution in (17), we use the well-known level set method [38]. The evolving curve $\vec{\gamma}(t)$ is *implicitly* represented by the zero level set of a function $u : \mathbb{R}^2\times\mathbb{R}^+ \to \mathbb{R}$ at time $t$, i.e., $\vec{\gamma}(t) = \{\mathbf{x}\in\Omega | u(\mathbf{x},t) = 0\}$. This representation is numerically stable and handles automatically topological changes of the evolving curve.

When the curve $\vec{\gamma}(t)$ is evolving following [38]:

$$\frac{d\vec{\gamma}}{dt} = F.\vec{n} \tag{18}$$

where $F : \mathbb{R}^2 \to \mathbb{R}$, the corresponding level set function $u$ evolves according to

$$\frac{\partial u}{\partial t}(\mathbf{x},t) = F.\|\vec{\nabla}u\|. \tag{19}$$

Using this result, the level set function evolution corresponding to (17) is given by

$$\frac{\partial u}{\partial t}(\mathbf{x},t) = \left(J_K\left(I(\mathbf{x}),\mu_1\right) - J_K\left(I(\mathbf{x}),\mu_2\right) - \lambda\kappa_u\right)\|\vec{\nabla}u\| \tag{20}$$

where the curvature function $\kappa_u$ is given by

$$\kappa_u = \vec{\nabla}\cdot\left(\vec{\nabla}u/\|\vec{\nabla}u\|\right) = \frac{u_{\mathbf{xx}}u_{\mathbf{y}}^2 - 2u_{\mathbf{x}}u_{\mathbf{y}}u_{\mathbf{xy}} + u_{\mathbf{yy}}u_{\mathbf{x}}^2}{(u_{\mathbf{x}}^2+u_{\mathbf{y}}^2)^{3/2}}. \tag{21}$$

It should be mentioned that (20) applies only for points $(\mathbf{x},t)$ on the curve $\vec{\gamma}$. We extend this evolution equation to the whole image domain [35]. The function $u$ evolves also for points outside its zero level according to (20) without affecting the process of segmentation and, us such, is more stable numerically. More details on the level set partial differential equation discretization schemes and fast resolution algorithms are available in [38].

### *C. Multiregion Segmentation*

Multiregion segmentation using several active curves can lead to ambiguity when two or more curves intersect. The

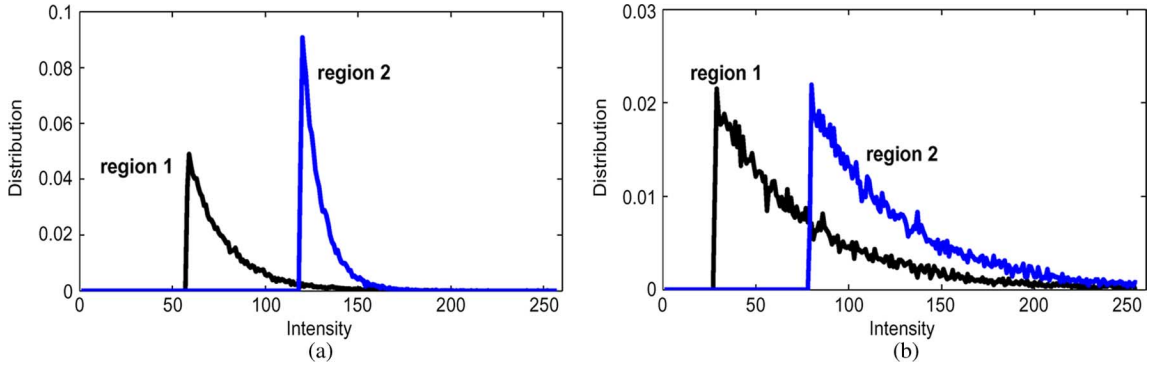



Fig. 3. Image intensity distributions: (a) small overlap; (b) significant overlap.

main issue is to guarantee that the curves converge to define a partition of the image domain. There are several ways of generalizing a two-region segmentation functional to a multi-region functional to guarantee such a partition. For instance, the generalization is done [3] via a term in the functional which draws the solution toward a partition, an explicit correspondence between the regions of segmentation and the interior of curves and their intersections in [8] and [35], and a partition constraint to use directly in the equations of minimization of the functional in [2]. These and other methods are reviewed in [35]. Here, we use the implementation of our generalization described in [35] and used in other applications [21], [22].

This generalization is based on the following definition of a partition. For a segmentation into $N$ regions, let $\{\vec{\gamma}_j\}_{j=1,\ldots,N-1}$ be $N-1$ simple closed plane curves and $\{\mathbf{R}_{\vec{\gamma}_j}\}$ the regions they enclose. Then, the following regions $\{\mathbf{R}_j\}$ form a partition: $\mathbf{R}_1 = \mathbf{R}_{\vec{\gamma}_1}; \mathbf{R}_2 = \mathbf{R}^c_{\vec{\gamma}_1} \cap \mathbf{R}_{\vec{\gamma}_2}; \ldots; \mathbf{R}_j = \mathbf{R}^c_{\vec{\gamma}_1} \cap \mathbf{R}^c_{\vec{\gamma}_2} \cap \ldots \cap \mathbf{R}^c_{\vec{\gamma}_{j-1}} \cap \mathbf{R}_{\vec{\gamma}_j}; \ldots;$ and $\mathbf{R}_N = \mathbf{R}^c_{\vec{\gamma}_1} \cap \mathbf{R}^c_{\vec{\gamma}_2} \cap \ldots \cap \mathbf{R}^c_{\vec{\gamma}_{j-1}} \cap \mathbf{R}^c_{\vec{\gamma}_{N-1}}$, where $\mathbf{R}^c_{\vec{\gamma}}$ is the complementary of $\mathbf{R}_{\vec{\gamma}}$ in $\Omega$. This is illustrated in Fig. 2 for four regions. The curve evolution equations and the corresponding level-set equations, using this definition of a partition, are given in Appendix B.

## IV. Experimental Results

To illustrate the effectiveness of the proposed method, we first give a *quantitative* and *comparative* performance evaluation over a large number of experiments on synthetic images with various noise models and contrast parameters. The percentage of misclassified pixels (PMP) was used as a measure of segmentation accuracy. To illustrate the flexibility of the method, we also show a representative sample of the tests with various classes of real images including natural images from the Berkeley database, medical and satellite data, as well as motion maps.

### A. Quantitative and Comparative Performance Evaluation

The piecewise constant segmentation method and the piecewise Gaussian generalization have been the focus of most studies and applications [6], [9] because of their tractability. In the following, evaluation of the proposed method, referred to as Kernelized Method (KM), is systematically supported by comparisons with the Piecewise Gaussian Method (PGM) [3], [19], [20]. The PGM method uses a Gaussian model in the data term of (5). In all our experiments, the KM method uses the RBF kernel (refer to Table I) with the same parameter $\sigma$.

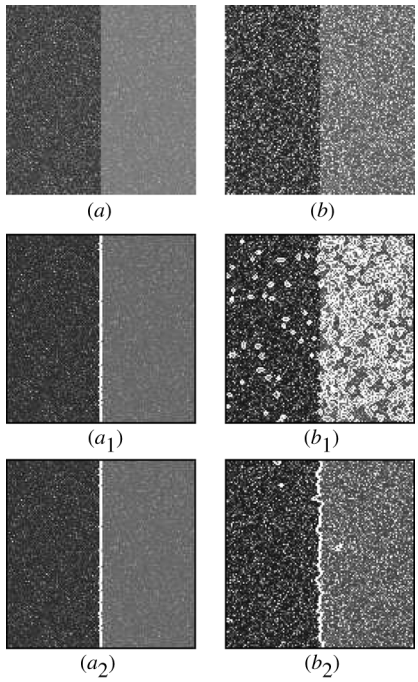


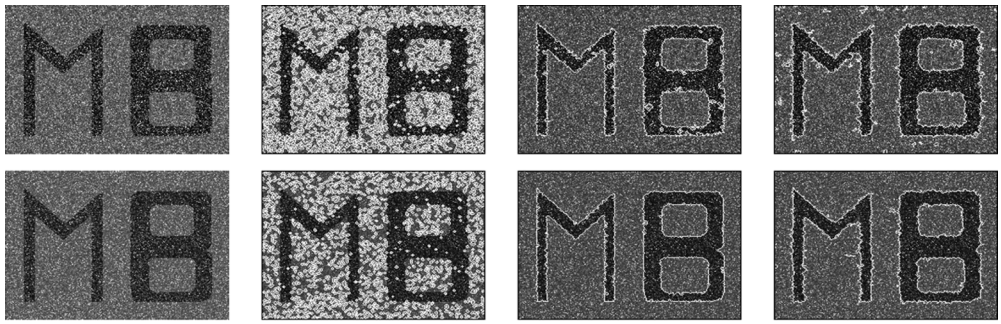

Fig. 4. Segmentation of two exponentially noisy images with different contrasts: (a), (b) noisy images with different contrasts; $(a_1)-(b_1)$ segmentation results with PGM; $(a_2)-(b_2)$ segmentation results with KM. Image size: 128 $\times$ 128. $\lambda = 1$.

We first show two typical examples of our extensive testing with synthetic images and define the measures we used for performance analysis: the contrast and the percentage of misclassified pixels (PMP). Fig. 4(a) and (b) depicts two versions of a two-region synthetic image, each perturbed with an exponential noise. Different noise parameters result in different amounts of overlap between the intensity distributions within the regions (Fig. 3). The larger the overlap, the more difficult the segmentation [4].

Figs. $4(a_1), (b_1)$ depict the segmentation results with the PGM. Because the actual noise model is exponential, segmentation quality obtained with the PGM was significantly affected with the second image in Fig. $4(b_1)$. However, the KM yielded approximately the same result for both images (Fig. $4a_2, b_2$), although the second image undergoes a relatively significant

TABLE II
PERCENTAGE OF CORRECTLY CLASSIFIED PIXELS IN FIG. 4 FOR DIFFERENT VALUES OF THE REGULARIZATION WEIGHT

| $\lambda$ | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|---|
| Percentage of PGM | 64.48 | 77.70 | 77.39 | 74.88 | 74.88 | 68.48 | 64.71 |



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

Fig. 5. (a) Synthetic images with Gamma (first row) and exponential (second row) noises; (b) segmentation results with PGM; (c) segmentation with the correct model: Gamma model for the first row image and exponential model for the second row image; and (d) the KM. Image size: 241 × 183. $\lambda = 2$ for both methods.

overlap between the intensity distributions within the two regions [Fig. 3(b)].

To demonstrate that the KM is a flexible and effective alternative to image modeling, we proceeded to a *quantitative* and *comparative* performance evaluation over a very large number of experiments. We run more than 100 experiments with a large set of synthetic tow-region images generated from various noise models and contrast values. The noise models we used include the Gaussian, the exponential, and the Gamma distributions. We recall the exponential distribution

$$f_\beta\left(I(\mathbf{x})\right) = \frac{1}{\beta} e^{-\frac{I(\mathbf{x})}{\beta}}, \quad \beta > 0 \tag{22}$$

and the Gamma distribution

$$f_{\beta,\eta}\left(I(\mathbf{x})\right) = I(\mathbf{x})^{\beta-1} \frac{e^{-\frac{I(\mathbf{x})}{\eta}}}{\eta^\beta \Gamma(\beta)}, \quad \eta, \beta > 0. \tag{23}$$

Each image was generated from a combination of a noise model and a contrast value. The latter was measured by the Bhattacharyya distance between the intensity distributions within the two regions of the actual image [4]

$$\mathcal{B} = -\ln \int\limits_{\mathbf{x}\in\mathcal{I}} \sqrt{P_1(\mathbf{x}) P_2(\mathbf{x})} d\mathbf{x} \tag{24}$$

where $P_1$ and $P_2$ denote the intensity distributions within the two regions and $\int_{\mathbf{x}\in\mathcal{I}} \sqrt{P_1(\mathbf{x})P_2(\mathbf{x})}d\mathbf{x}$ is the Bhattacharyya coefficient measuring the amount of overlap between these distributions. Note that the higher the overlap, the lower the contrast. Each image was segmented by three methods: the PGM, the KM and segmentation with the correct model, i.e., the noise model used to generate the actual image. A comparative performance analysis was carried out by assessing the effect of the contrast on segmentation accuracy for the three segmentation methods. We adopted the percentage of misclassified pixels (PMP) as a measure of segmentation accuracy

$$PMP = \left(1 - \frac{|B_O \cap B_S| + |F_O \cap F_S|}{|B_O| + |F_O|}\right) \times 100 \tag{25}$$

where $B_O$ and $F_O$ denote the background and foreground of the ground truth (correct segmentation) and $B_S$ and $F_S$ denote the background and foreground of the segmented image.

The value of parameter $\lambda$ for the PGM method is chosen by varying it in an interval about the value of 2. A value of approximately 2 was shown to be optimal for distributions from the exponential family such as the piecewise Gaussian model [4] and has an interesting minimum description length (MDL) interpretation. This was confirmed by experiments in the study [21] with other image data distributions. We used the percentage of misclassified pixels to chose $\lambda$, as illustrated in Table II for the example of Fig. 4.

A second typical example of our extensive testing with synthetic data is depicted in Fig. 5. It shows two different noisy versions of a piecewise constant two-region image perturbed with a Gamma (first row) and exponential noises (second row). The PGM yielded unsatisfying results [Fig. 5(b)]. Acceptable results can be obtained when the correct model is assumed [Fig. 5(c)]. Although no assumption was made as to the noise model, The KM yielded a competitive segmentation quality [Fig. 5(d)]. Thus, the proposed method allows much more flexibility in practice because the model distribution of image data does not have to be fixed.

We run more than 100 experiments to study the effect of the contrast on the PMP. Several synthetic two-region images were generated from the Gaussian, Gamma and exponential noises. For each noise model, we varied the contrast between the two regions, which yielded more than 30 images. For each image, we evaluated the segmentation accuracy for three level set methods: The KM, the PGM, and segmentation when the correct model is assumed, i.e., the model used to generate the current image.

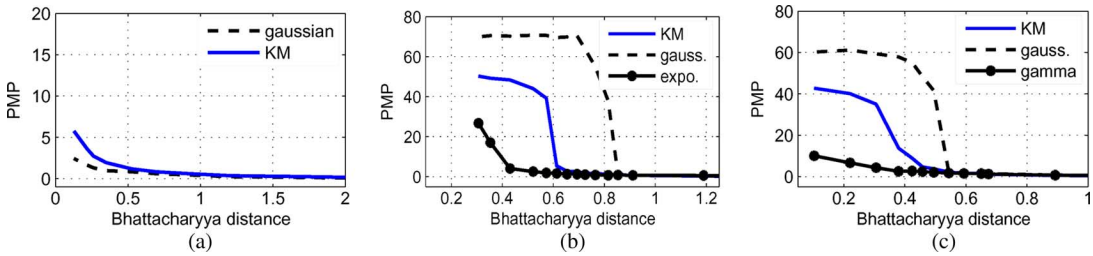



Fig. 6. Evaluation of segmentation error for different methods over a large number of experiments (PMP as a function of the contrast): comparisons over the subset of synthetic images perturbed with a Gaussinan noise in (a), the exponential noise in (b), and the Gamma noise in (c).

First, we applied both PGM and KM to the subset of images perturbed with a Gaussian noise, and plotted the PMP as a function of the contrast [Fig. 6(a)]. The higher the PMP, the higher the segmentation error. The KM yielded approximately the same error as segmentation with the correct model, i.e., the Gaussian model in this case. Second, we segmented all the images perturbed with an exponential noise with both PGM, KM and segmentation with the correct model, i.e., the exponential model in this case. We plotted the PMP as a function of the contrast in Fig. 6(b). Both segmentation methods undergo high error gradients at some Bhattacharyya distance. Those results are consistent with the experiments in [5]. When $\mathcal{B}$ is superior to 0.8, both methods yield a low segmentation error with a PMP less than 1%. However, the KM outperforms the PGM for a considerable range of Bhattacharyya distance values. Furthermore, the KM yielded a performance similar to segmentation with the correct model [refer to Fig. 6(b)] until the contrast becomes very small ($\mathcal{B} < 0.6$). Similar experiments were run with the subset of images perturbed with a Gamma noise, and a similar behavior was noticed [refer to Fig. 6(c)]. These results demonstrate the ability of the KM to deal with various classes of image noises for a large range of contrast values, which relaxes assumptions as to the correct noise model.

The ability of the KM to deal with different noise models allows segmenting regions which require different models.[1] To illustrate this important advantage of the KM, we consider a synthetic image of three regions with different noise models as shown in Fig. 7(a) with the initial curves in black and white. The clearer region is generated with a Gaussian noise, the gray region is derived from the Rayleigh distribution, and the darker region from the Poisson distribution. The final position of the curves following the KM is displayed in Fig. 7(b), and final segmentation, in Fig. 7(c), where each region is represented by its mean intensity value. Fig. 7(d)–(f) shows the segmentation regions separately. As shown in Fig. 7(g)–(i), the Gaussian model gives incorrect results as expected. The results demonstrate the ability of our kernel method to discriminate different distributions within the same image.

Fig. 8 depicts the results with a simulation of a Synthetic Aperture Radar (SAR) image. Segmentation of SAR images is a difficult task due to the presence of speckle which is known as strong and multiplicative noise. For single-look SAR images, the intensity is given by $I = a^2 + b^2$, where $a$ and $b$ denote the real and imaginary parts of the complex signal acquired from the radar. For multilook SAR images, the L-look intensity is the average of the L intensity images [22]. Fig. 8(a), with initial curves, is a synthetic four-region image simulating an amplitude multilook SAR image. Initial curves in black, white and gray are placed arbitrarily about the middle of the image. Fig. 8(b) shows final segmentation where each region is represented by its corresponding parameter (weighted mean). Fig. 8(c)–(f) shows segmentation regions separately.

[1]In practice, the segmentation regions may require different models. For example, in synthetic aperture radar (SAR) images, the intensity follows a Gamma distribution in a zone of constant reflectivity and a K distribution in a zone of textured reflectivity [15]. The luminance within shadow regions in sonar imagery is well modeled by the Gaussian distribution while the Rayleigh distribution is more accurate in the reverberation regions [14].

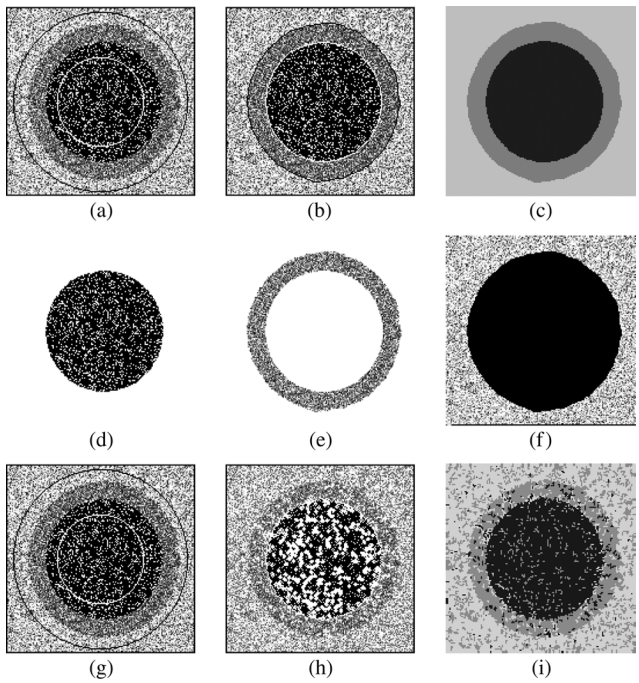



Fig. 7. Image with different noise models: (a) initialization; (b) final position of curves; (c) final segmentation; (d)–(f) segmentation regions separately; (g)–(i) results with the PGM. Image size: 163 × 158. $\lambda = 2$ for both methods.

### B. Real Data

In the following, we illustrate the flexibility of the proposed method by a representative sample of the tests with various classes of real images including natural images from the Berkeley database, medical and satellite data, as well as motion maps.

Segmentation of a natural plane image into two regions is depicted in Fig. 9. Initial, intermediate and final positions of the

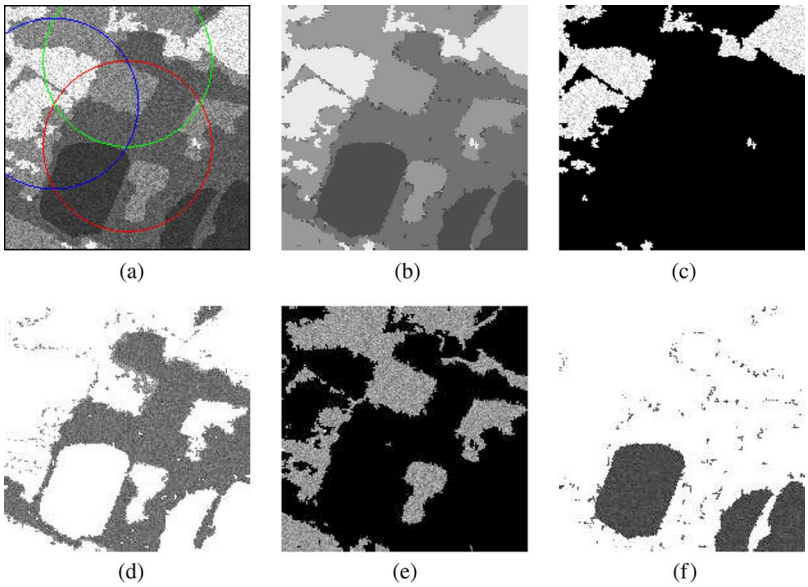



Fig. 8. Simulated multilook SAR image: (a) initialization; (b) final segmentation; (c)–(f) segmentation regions separately. Image size: 192 × 189. $\lambda = 2$.


Fig. 9. Real plane image: (a) initialization; (b) intermediate curve evolution step; (c) final position of the curve; (d) final segmentation. Image size: 110 × 70. $\lambda = 2$.

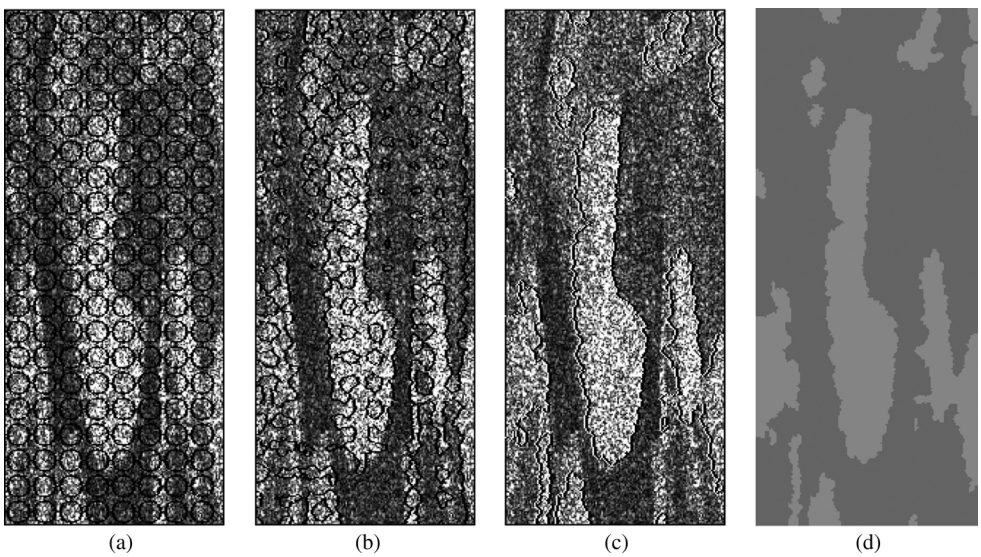
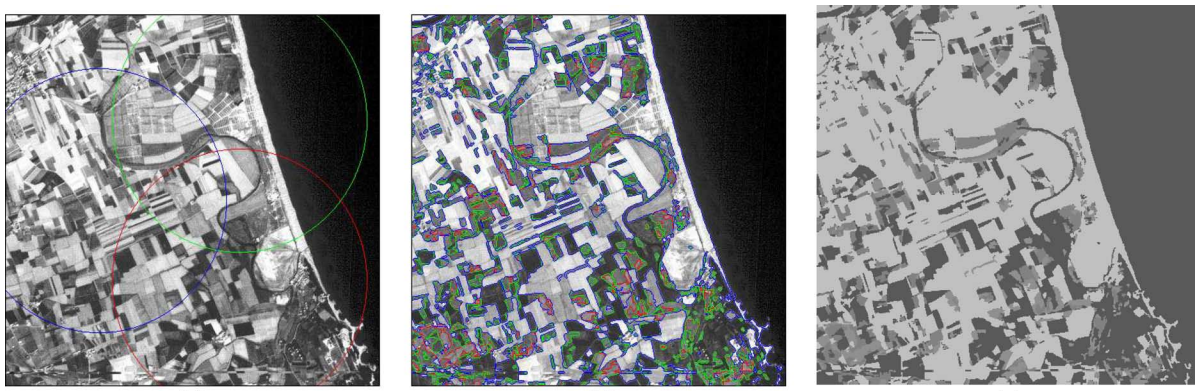



Fig. 10. Monolook SAR image: (a) initialization; (b) intermediate curve evolution step (c) final position of the curve; (d) final segmentation. Image size: 151 × 361. $\lambda = 2$.

evolving curve are displayed, respectively, in Fig. 9(a)–(c). The final segmentation regions, represented by their corresponding parameters at convergence, are illustrated in Fig. 9(d).

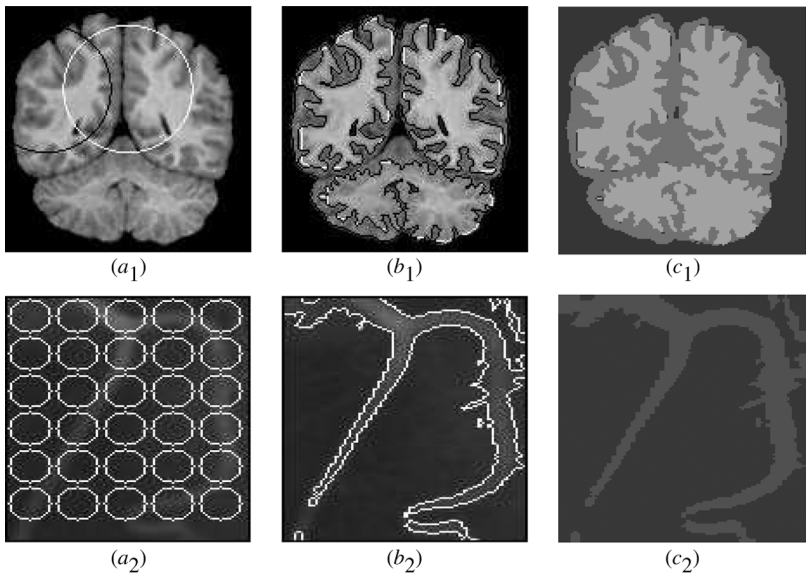
To illustrate the robustness of the method with respect to initial conditions, initial curves were either big circles placed arbitrarily about the middle of the image or tiny circles spread all over the image.

Fig. 10 depicts the result with a monolook SAR image characterized by a high multiplicative speckle noise. The noise level depends on the image data: the higher the intensity, the stronger the noise. Segmentation of this class of images is acknowledged as a difficult problem [21], [22]. An intermediate and the final positions of the evolving curve are shown respectively in Fig. 10(b) and Fig. 10(c). Both segmented regions, represented



(a) (b) (c)

Fig. 11. Panchromatic SPOT data of the Ampurias area: (a) initialization; (b) final position of the curve; (d) final segmentation. Image size: $512 \times 512$. $\lambda = 1$.



$(a_1)$ $(b_1)$ $(c_1)$

$(a_2)$ $(b_2)$ $(c_2)$

Fig. 12. Brain and Vessel images: $(a_i)$ initialization; $(b_i)$ final position of the curves; $(c_i)$ final segmentation.

by their parameters, are displayed in Fig. 10(d). This example demonstrates the ability of the method to deal effectively with difficult SAR images where the level of speckle results in a very low contrast within several locations.

Fig. 11 depicts the segmentation result into four regions of a high resolution panchromatic spot image of the region of Ampurias. This image shows several disjoint tiny blocks with different gray levels. Initial and final curves are displayed, respectively, in Fig. 11(a) and (b). Segmentation regions are represented in Fig. 11(c) by their parameters at convergence.

Medical image segmentation is challenging and of a rapidly growing interest. The brain image shown in Fig. $12(a_1)$ was segmented into three regions. The choice of the number of regions is based on prior medical knowledge. Fig. $12(a_2)$ depicts very narrow human vessels with very small contrast within some spots. The curves obtained at convergence are displayed, for both images, in Fig. $12(b_1)$ and $(b_2)$. Segmentation regions, represented by their parameters at convergence, are shown in Fig. $12(c_1)$ and $(c_2)$.

A representative sample of the experiments with natural images from the Berkeley database is depicted in Fig. 13. The number of regions was based on visual inspection. The proposed method yielded visually satisfying segmentation results with natural images.

### C. Motion

In this experiment, we segment optical flow images into motion regions. Optical flow at each pixel is a 2-D vector computed following the method in [44]. We show two examples. The first example uses the Road image sequence (first row of Fig. 14), which contains two regions: a moving vehicle and a background. The second example uses the Marmor sequence (second row), which contains three regions: two moving objects and a background. Motion maps, initial and final curves are depicted, respectively, in Fig. 14(a), (d), Fig. 14(b), (e), and Fig. 14(c), (f).

## V. Conclusion

This study investigated multiphase piecewise image segmentation in a kernel-induced space. This led to a flexible and effective alternative to complex modeling of image data. The method used an active curve objective functional containing an original term which references the image data transformed via a kernel

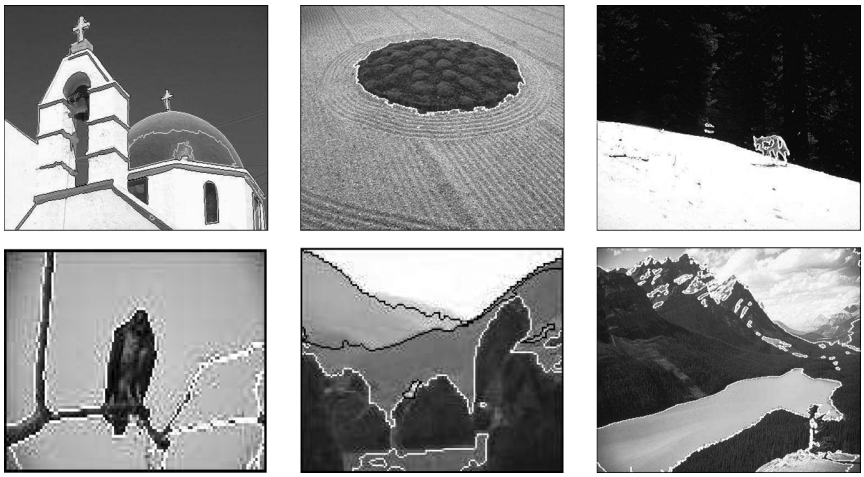


Fig. 13. Example of segmentation results with natural images from the Berkeley database.

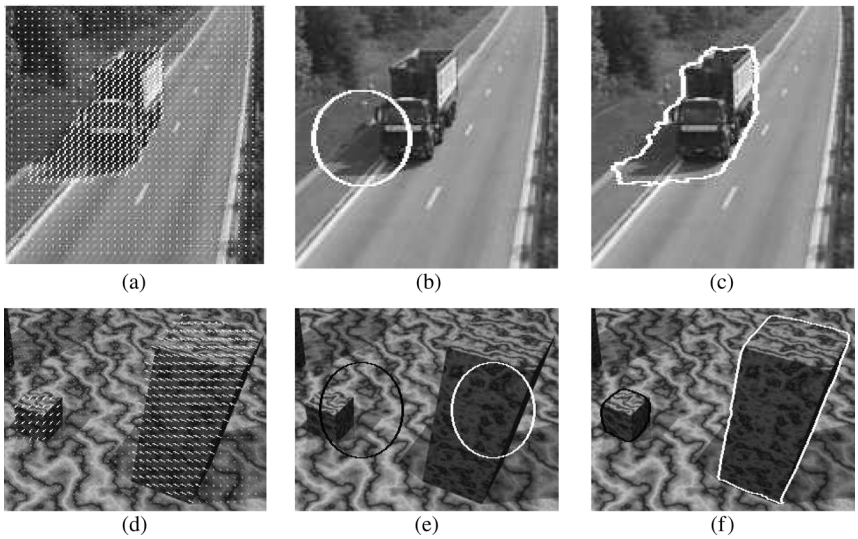



Fig. 14. Segmentation of the Road and Marmor sequences: (a) motion map of the Road sequence; (b), (e) initializations; (c), (f) final segmentations. $\lambda = 2$.

function. The algorithm iterated two consecutive steps: curve evolution and mean shift update of the regions parameters. We verified the effectiveness of the proposed method with a *quantitative* and *comparative* performance evaluation over a very large number of experiments on synthetic images. The flexibility of the method was illustrated with a representative sample of the experiments on various classes of real images including natural images, medical and satellite data, as well as motion maps.

## APPENDIX A

Let $g_{\mathbf{R}}$ be the mean shift update function defined in (12). In this Appendix, we prove the convergence of the following sequence:

$$\mathbf{t}_{j+1} = g_{\mathbf{R}}(\mathbf{t}_j) = \frac{\int_{\mathbf{R}} I(\mathbf{x}) K\left(I(\mathbf{x}), \mathbf{t}_j\right) d\mathbf{x}}{\int_{\mathbf{R}} K\left(I(\mathbf{x}), \mathbf{t}_j\right) d\mathbf{x}}, \quad j = 1, 2, \ldots$$

where $\mathbf{R}$ is a region within $\Omega$ and $K$ is the RBF kernel. A similar result was obtained for the discrete case in [36]. With $h(\mathbf{x}) = \exp(-\mathbf{x})$, sequence $\{\mathbf{t}_j\}_{j=1,2,\ldots}$ can be written as follows:

$$\mathbf{t}_{j+1} = \frac{\int_{\mathbf{R}} I(\mathbf{x}) h\left(\left\|\frac{I(\mathbf{x})-\mathbf{t}_j}{\sigma}\right\|^2\right) d\mathbf{x}}{\int_{\mathbf{R}} h\left(\left\|\frac{I(\mathbf{x})-\mathbf{t}_j}{\sigma}\right\|^2\right) d\mathbf{x}}, \quad j = 1, 2, \ldots \quad \text{(A1)}$$

Define sequence $\{f_\sigma(j)\}_{j=1,2,\ldots}$ as follows:

$$f_\sigma(j) = \int_{\mathbf{R}} h\left(\left\|\frac{I(\mathbf{x}) - \mathbf{t}_j}{\sigma}\right\|^2\right) d\mathbf{x}, \quad j = 1, 2, \ldots$$

First, we demonstrate that sequence $f_\sigma$ converges, and consequently, is a cauchy sequence. Because $0 < h(\|(I(\mathbf{x}) - \mathbf{t}_j)/\sigma\|^2) \leq 1 \forall \mathbf{x} \in \mathbf{R}$ and $\mathbf{R} \subset \Omega$, we have

$$0 < f_\sigma(j) \leq \int_{\mathbf{R}} d\mathbf{x} \leq \int_{\Omega} d\mathbf{x}. \quad \text{(A2)}$$

Thus, $f_\sigma$ is bounded by the area of the image domain and it suffices to show that $f_\sigma$ is strictly monotonic increasing.

Now for $\mathbf{t}_{j+1} \neq \mathbf{t}_j$, $j = 1, 2, \ldots$, consider the following expression:

$$f_\sigma(j+1) - f_\sigma(j) = \int_{\mathbf{R}} \left[ h\left( \left\| \frac{I(\mathbf{x}) - \mathbf{t}_{j+1}}{\sigma} \right\|^2 \right) - h\left( \left\| \frac{I(\mathbf{x}) - \mathbf{t}_j}{\sigma} \right\|^2 \right) \right] d\mathbf{x}. \quad \text{(A3)}$$

Using the fact that $h$ is convex, then for all $u, v \in \mathbb{R}^+$ such that $u \neq v$, we have

$$h(v) \geq h(u) + h'(u)(v-u). \quad \text{(A4)}$$

As function $h$ verifies $-h'(u) = h(u)$, then (A4) becomes

$$h(v) - h(u) \geq h(u)(u-v). \quad \text{(A5)}$$

Now, combining (A3) and (A5) gives

$$\begin{aligned} f_\sigma(j+1) - f_\sigma(j) &\geq \frac{1}{\sigma^2} \int_{\mathbf{R}} h\left( \left\| \frac{I(\mathbf{x}) - \mathbf{t}_j}{\sigma} \right\|^2 \right) \times \left[ \|I(\mathbf{x}) - \mathbf{t}_j\|^2 - \|I(\mathbf{x}) - \mathbf{t}_{j+1}\|^2 \right] d\mathbf{x} \\ &= \frac{1}{\sigma^2} \int_{\mathbf{R}} h\left( \left\| \frac{I(\mathbf{x}) - \mathbf{t}_j}{\sigma} \right\|^2 \right) \times \left[ 2\mathbf{t}_{j+1}^T I(\mathbf{x}) - \|\mathbf{t}_{j+1}\|^2 - 2\mathbf{t}_j^T I(\mathbf{x}) + \|\mathbf{t}_j\|^2 \right] d\mathbf{x} \\ &= \frac{1}{\sigma^2} \left[ 2\mathbf{t}_{j+1}^T \int_{\mathbf{R}} I(\mathbf{x}) h\left( \left\| \frac{I(\mathbf{x}) - \mathbf{t}_j}{\sigma} \right\|^2 \right) d\mathbf{x} \right. \\ &\quad - \|\mathbf{t}_{j+1}\|^2 \int_{\mathbf{R}} h\left( \left\| \frac{I(\mathbf{x}) - \mathbf{t}_j}{\sigma} \right\|^2 \right) d\mathbf{x} \\ &\quad - 2\mathbf{t}_j^T \int_{\mathbf{R}} I(\mathbf{x}) h\left( \left\| \frac{I(\mathbf{x}) - \mathbf{t}_j}{\sigma} \right\|^2 \right) d\mathbf{x} \\ &\quad \left. + \|\mathbf{t}_j\|^2 \int_{\mathbf{R}} h\left( \left\| \frac{I(\mathbf{x}) - \mathbf{t}_j}{\sigma} \right\|^2 \right) d\mathbf{x} \right]. \end{aligned} \quad \text{(A6)}$$

Using (A1) in (A6) yields, after some manipulations

$$f_\sigma(j+1) - f_\sigma(j) \geq \frac{1}{\sigma^2} \|\mathbf{t}_{j+1} - \mathbf{t}_j\|^2 \int_{\mathbf{R}} h\left( \left\| \frac{I(\mathbf{x}) - \mathbf{t}_j}{\sigma} \right\|^2 \right) d\mathbf{x}. \quad \text{(A7)}$$

Because $\mathbf{t}_{j+1} \neq \mathbf{t}_j$ for $j = 1, 2, \ldots$, the right term in (A7) is strictly positive, and consequently, sequence $f_\sigma$ is strictly increasing. With inequalities (A2), this concludes that $f_\sigma$ is convergent.

Summing both sides in inequality (A7) over $j, j+1, \ldots, j+m-1$ gives

$$\begin{aligned} f_\sigma(j+m) - f_\sigma(j) &\geq \frac{1}{\sigma^2} \|\mathbf{t}_{j+m} - \mathbf{t}_{j+m-1}\|^2 \int_{\mathbf{R}} h\left( \left\| \frac{I(\mathbf{x}) - \mathbf{t}_{j+m-1}}{\sigma} \right\|^2 \right) d\mathbf{x} \\ &\quad + \ldots + \frac{1}{\sigma^2} \|\mathbf{t}_{j+1} - \mathbf{t}_j\|^2 \int_{\mathbf{R}} h\left( \left\| \frac{I(\mathbf{x}) - \mathbf{t}_j}{\sigma} \right\|^2 \right) d\mathbf{x} \\ &\geq \frac{1}{\sigma^2} \left[ \|\mathbf{t}_{j+m} - \mathbf{t}_{j+m-1}\|^2 + \ldots + \|\mathbf{t}_{j+1} - \mathbf{t}_j\|^2 \right] M \\ &\geq \frac{M}{\sigma^2} \|\mathbf{t}_{j+m} - \mathbf{t}_j\|^2 \end{aligned} \quad \text{(A8)}$$

where $M$ is the minimum of the integral $\int_{\mathbf{R}} h(\|(I(\mathbf{x}) - \mathbf{t}_j)/\sigma\|^2) d\mathbf{x}$ with respect to $\{t_j\}_{j=1,2,\ldots}$. Note that $M$ is strictly positive.

Sequence $f_\sigma$ is convergent, and consequently, is a Cauchy sequence. This result combined with (A8) concludes that $\{t_j\}_{j=1,2,\ldots}$ is a Cauchy sequence. As a result, $\{t_j\}_{j=1,2,\ldots}$ converges in the Euclidean space.

## APPENDIX B

For multiphase segmentation into $N$ regions ($N > 2$), the correspondence in Section III-C leads to coupled evolution equations:

$$\begin{aligned} \frac{d\vec{\gamma}_1}{dt} &= -\left(e_1(\vec{\gamma}_1) - \Phi_1(\vec{\gamma}_1) + \lambda\kappa_1\right)\vec{n}_1 \\ &\vdots \\ \frac{d\vec{\gamma}_j}{dt} &= -\Big(\chi_{\mathbf{R}_1^c}(\vec{\gamma}_j) \cdots \chi_{\mathbf{R}_{j-1}^c}(\vec{\gamma}_j)\left[e_j(\vec{\gamma}_j) - \Phi_j(\vec{\gamma}_j)\right] + \lambda\kappa_j\Big)\vec{n}_j \\ &\vdots \\ \frac{d\vec{\gamma}_{N-1}}{dt} &= -\Big(\chi_{\mathbf{R}_1^c}(\vec{\gamma}_{N-1}) \cdots \chi_{\mathbf{R}_{N-2}^c}(\vec{\gamma}_{N-1}) \times \left[e_{N-1}(\vec{\gamma}_{N-1}) - \Phi_{N-1}(\vec{\gamma}_{N-1})\right] + \lambda\kappa_{N-1}\Big) \times \vec{n}_{N-1} \end{aligned} \quad \text{(A9)}$$

where, for $j = 1, \ldots, N-1$, $e_j(\mathbf{x}) = J_K(I(\mathbf{x}), \mu_j)$, $\mu_j$ is the parameter of region $\mathbf{R}_j$, $\vec{n}_j$ is the outward unit normal to $\vec{\gamma}_j$ and $\kappa_j$ its curvature function, $\chi_{\mathbf{R}_j}$ is the characteristic function of $\mathbf{R}_j$, and $\Phi_j(\mathbf{x})$ is given by

$$\begin{aligned} \Phi_j(\mathbf{x}) = &\, e_{j+1}(\mathbf{x})\chi_{\mathbf{R}_{\vec{\gamma}_{j+1}}}(\mathbf{x}) + \ldots \\ &+ e_{N-1}(\mathbf{x})\chi_{\mathbf{R}^c_{\vec{\gamma}_{j+1}}}(\mathbf{x}) \cdots \chi_{\mathbf{R}^c_{\vec{\gamma}_{N-2}}}(\mathbf{x})\chi_{\mathbf{R}_{\vec{\gamma}_{N-1}}}(\mathbf{x}) \\ &+ e_N(\mathbf{x})\chi_{\mathbf{R}^c_{\vec{\gamma}_{j+1}}}(\mathbf{x}) \cdots \chi_{\mathbf{R}^c_{\vec{\gamma}_{N-2}}}(\mathbf{x})\chi_{\mathbf{R}^c_{\vec{\gamma}_{N-1}}}(\mathbf{x}). \end{aligned}$$

Note that the update of the regions parameters follows directly from the tow-region case.

As shown in Section III-B, deriving the level-set form of the evolution equations corresponding to system (A9) is straightforward following (19). Indeed, each curve $\vec{\gamma}_j$ is represented by the zero level set of a function $u_j : \mathbb{R}^2 \to \mathbb{R}$, where $u_j > 0$ only

inside $\vec{\gamma}_j$. As in [35], level-set evolution equations follow this system of coupled partial differential equations

$$\begin{aligned}
\frac{\partial u_1}{\partial t}(\mathbf{x},t) &= -\left(e_1(\mathbf{x}) - \Phi_1(\mathbf{x}) + \lambda\kappa_{u_1}\right)\left\|\vec{\nabla}u_1(\mathbf{x},t)\right\| \\
&\vdots \\
\frac{\partial u_j}{\partial t}(\mathbf{x},t) &= -\left(\chi_{\{u_1(\mathbf{x},t)\le 0\}}\cdots\chi_{\{u_{j-1}(x,t)\le 0\}}\right. \\
&\quad \left.\times\left[e_j(\mathbf{x}) - \Phi_j(\mathbf{x})\right] + \lambda\kappa_{u_j}\right)\left\|\vec{\nabla}u_j(\mathbf{x},t)\right\| \\
&\vdots \\
\frac{\partial u_{N-1}}{\partial t}(\mathbf{x},t) &= -\left(\chi_{\{u_1(\mathbf{x},t)\le 0\}}\cdots\chi_{\{u_{N-2}(\mathbf{x},t)\le 0\}}\right. \\
&\quad \left.\times\left[e_{N-1}(\mathbf{x}) - \Phi_{N-1}(\mathbf{x})\right] + \lambda\kappa_{u_{N-1}}\right) \\
&\quad \times\left\|\vec{\nabla}u_{N-1}(\mathbf{x},t)\right\|
\end{aligned}$$

where $\kappa_{u_j}$ is the curvature function of the level set $u_j$, $\chi_A$ is the indicator function of the set $A$, and $\Phi_j(\mathbf{x})$ is given by

$$\begin{aligned}
\Phi_j(\mathbf{x}) = {} & e_{j+1}(\mathbf{x})\chi_{\{u_{j+1}(\mathbf{x},t)>0\}}(\mathbf{x}) + \cdots \\
& + e_{N-1}(\mathbf{x})\chi_{\{u_{j+1}(\mathbf{x},t)\le 0\}}(\mathbf{x})\cdots \\
& \chi_{\{u_{N-2}(\mathbf{x},t)\le 0\}}(\mathbf{x})\chi_{\{u_{N-1}(\mathbf{x},t)>0\}}(\mathbf{x}) \\
& + e_N(\mathbf{x})\chi_{\{u_{j+1}(\mathbf{x},t)\le 0\}}(\mathbf{x})\cdots\chi_{\{u_{N-2}(\mathbf{x},t)\le 0\}}(\mathbf{x}) \\
& \times\chi_{\{u_{N-1}(\mathbf{x},t)\le 0\}}(\mathbf{x}).
\end{aligned}$$

## REFERENCES


[1] M. Mignotte, "Segmentation by fusion of histogram-based K-means clusters in different color spaces," *IEEE Trans. Image Process.*, vol. 17, no. 5, pp. 780–787, May 2008.

[2] I. B. Ayed and A. Mitiche, "A region merging prior for variational level set image segmentation," *IEEE Trans. Image Process.*, vol. 17, no. 12, pp. 2301–2311, Dec. 2008.

[3] C. Samson, L. Blanc-Féraud, G. Aubert, and J. Zerubia, "A level set model for image classification," *Int. J. Comput. Vis.*, vol. 40, no. 3, pp. 187–197, Mar. 2000.

[4] P. Martin, P. Réfrégier, F. Goudail, and F. Guérault, "Influence of the noise model on level set active contour segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, no. 6, pp. 799–803, Jun. 2004.

[5] G. Delyon, F. Galland, and P. Réfrégier, "Minimal stochastic complexity image partitioning with unknown noise model," *IEEE Trans. Image Process.*, vol. 15, no. 10, pp. 3207–3212, Oct. 2006.

[6] T. F. Chan and L. A. Vese, "Active contours without edges," *IEEE Trans. Image Process.*, vol. 10, no. 2, pp. 266–277, Feb. 2001.

[7] T. F. Chan, B. Y. Sandberg, and L. A. Vese, "Active contours without edges for vector-valued images," *J. Vis. Commun. Image Represent.*, vol. 11, no. 2, pp. 130–141, 2000.

[8] L. A. Vese and T. F. Chan, "A multiphase level set framework for image segmentation using the Mumford ans Shah model," *Int. J. Comput. Vis.*, vol. 50, no. 3, pp. 271–293, 2002.

[9] X.-C. Tai, O. Christiansen, P. Lin, and I. Skjaelaaen, "Image segmentation using some piecewise constant level set methods with MBO type of projection," *Int. J. Comput. Vis.*, vol. 73, no. 1, pp. 61–76, 2007.

[10] J. M. Morel and S. Solimini, *Variational Methods in Image Segmentation*. Boston, MA: Birkhauser, 1995.

[11] G. Aubert and P. Kornprobst, *Mathematical Problems in Image Processing: Partial Differential Equations and the Calculus of Variations*. New York: Springer-Verlag, 2006.

[12] S. C. Zhu and A. Yuille, "Region competetion: Unifying snakes, region growing, and bayes/MDL for multiband image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 18, no. 6, pp. 884–900, Jun. 1996.

[13] D. Cremers, M. Rousson, and R. Deriche, "A review of statistical approches to level set segmentation: Integrating color, texture, motion and shape," *Int. J. Comput. Vis.*, vol. 72, no. 2, 2007.

[14] M. Mignotte, C. Collet, P. Pérez, and P. Bouthemy, "Sonar image segmentation using an unsupervised hierarchical MRF model," *IEEE Trans. Image Process.*, vol. 9, no. 7, pp. 1216–1231, Jul. 2000.

[15] R. Fjortoft, Y. Delignon, W. Pieczynski, M. Sigelle, and F. Tupin, "Unsupervised classification of radar images using hidden Markov chains and hidden Markov random fields," *IEEE Trans. Geosci. Remote Sens.*, vol. 41, no. 3, pp. 675–686, Mar. 2003.

[16] V. Caselles, R. Kimmel, and G. Sapiro, "Geodesic active contours," in *Proc. Int. Conf. Computer Vision*, 1995, pp. 694–699.

[17] D. Mumford and J. Shah, "Optimal approximation by piecewise smooth functions and associated variational problems," *Comm. Pure Appl. Math*, vol. 42, pp. 577–685, 1989.

[18] Y. G. Leclerc, "Constructing simple stable descriptions for image partitioning," *Int. J. Comput. Vis.*, vol. 3, no. 1, pp. 73–102, 1989.

[19] Y. Dong, B. C. Forster, and A. K. Milne, "Comparison of radar image segmentation by Gaussian- and Gamma-Markov random field models," *Int. J. Remote Sens.*, vol. 24, no. 4, pp. 711–722, 2003.

[20] C. Lenglet, M. Rousson, and R. Deriche, "DTI segmentation by statistical surface evolution," *IEEE Trans. Med. Imag.*, vol. 25, no. 6, pp. 685–700, Jun. 2006.

[21] I. B. Ayed, N. Hennane, and A. Mitiche, "Unsupervised variational image segmentation/classification using a weibull observation model," *IEEE Trans. Image Process.*, vol. 15, no. 11, pp. 3431–3439, Nov. 2006.

[22] I. B. Ayed, A. Mitiche, and Z. Belhadj, "Multiregion level-set partitioning of synthetic aperture radar images," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 5, pp. 793–800, May 2005.

[23] E. E. Kuruoglu and J. Zerubia, "Modeling SAR images with a generalization of the Rayleigh distribution," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 527–533, Apr. 2004.

[24] A. Achim, E. E. Kuruoglu, and J. Zerubia, "SAR image filtering based on the heavy-tailed Rayleigh model," *IEEE Trans. Image Process.*, vol. 15, no. 9, pp. 2686–2693, Sep. 2006.

[25] I. B. Ayed, A. Mitiche, and Z. Belhadj, "Polarimetric image segmentation via maximum-likelihood approximation and efficient multiphase level-sets," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 9, pp. 1493–1500, Sep. 2006.

[26] A. A. Nielsen, K. Conradsen, and H. Skriver, "Polarimetric synthetic aperture radar data and the complex wishart distribution," in *Proc. 13th Scandinavian Conf. Image Analysis*, Jun. 2003, pp. 1082–1089.

[27] S. C. Zhu, "Statistical modeling and conceptualization of visual patterns," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, pp. 691–712, 2003.

[28] I. S. Dhillon, Y. Guan, and B. Kulis, "Weighted graph cuts without eigenvectors: A multilevel approach," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 11, pp. 1944–1957, Nov. 2007.

[29] A. Jain, P. Duin, and J. Mao, "Statistical pattern recognition: A review," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 1, pp. 4–37, Jan. 2000.

[30] K. R. Muller, S. Mika, G. Ratsch, K. Tsuda, and B. Scholkopf, "An introduction to kernel-based learning algorithms," *IEEE Trans. Neural Netw.*, vol. 12, no. 2, pp. 181–202, Feb. 2001.

[31] B. Scholkopf, A. J. Smola, and K. R. Muller, "Nonlinear component analysis as a kernel eigenvalue problem," *Neural Comput.*, vol. 10, pp. 1299–1319, 1998.

[32] M. Girolami, "Mercer kernel based clustering in feature space," *IEEE Trans. Neural Netw.*, vol. 13, pp. 780–784, 2001.

[33] D. Q. Zhang and S. C. Chen, "Fuzzy clustering using kernel methods," in *Proc. Int. Conf. Control Automation*, Xiamen, China, Jun. 2002, pp. 123–127.

[34] C. Vazquez, A. Mitiche, and I. B. Ayed, "Image segmentation as regularized clustering: A fully global curve evolution method," in *Proc. IEEE Int. Conf. Image Processing*, Oct. 2004, pp. 3467–3470.

[35] A.-R. Mansouri, A. Mitiche, and C. Vazquez, "Multiregion competition: A level set extension of region competition to multiple region partitioning of images and image sequences," *Comput. Vis. Image Understand.*, vol. 101, no. 3, pp. 137–150, Mar. 2006.

[36] D. Comaniciu and P. Meer, "Mean shift: A robust approach toward feature space analysis," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 5, pp. 603–619, May 2002.

[37] Y. Cheng, "Mean shift, mode seeking, and clustering," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 17, no. 8, pp. 790–799, Aug. 1995.

[38] J. A. Sethian, *Level Set Methods and Fast Marching Methods*, 2nd ed. Cambridge, U.K.: Cambridge Univ. Press, 1999.

[39] T. M. Cover, "Geomeasureal and statistical properties of systems of linear inequalities in pattern recognition," *Electron. Comput.*, vol. EC-14, pp. 326–334, Mar. 1965.

[40] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York: Wiley-Interscience, 1991.
[41] F. Goudail, P. Réfrégier, and G. Delyon, "Bhattacharyya distance as a contrast parameter for statistical processing of noisy optical images," *J. Opt. Soc. Amer. A*, vol. 21, no. 7, pp. 1231–1240, 2004.
[42] P. J. Hubert, *Robust Statistics*. New York : Wiley, 1981.
[43] K. L. Wu and M. S. Yang, "Alternative c-means clustering algorithms," *Pattern Recognit.*, vol. 35, pp. 2267–2278, 2002.
[44] C. Vazquez, A. Mitiche, and R. Laganiere, "Joint multiregion segmentation and parametric estimation of image motion by basis function representation and level set evolution," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, pp. 782–793, 2006.

**Amar Mitiche** (M'03) received the Licence Ès Sciences degree in mathematics from the University of Algiers, Algeria, and the Ph.D. degree in computer science from the University of Texas at Austin.

He is currently a Professor at the Institut National de Recherche Scientifique (INRS), Department of Telecommunications, Montreal, QC, Canada. His research is in computer vision. His current interests include image segmentation, motion analysis in monocular and stereoscopic image sequences (detection, estimation, segmentation, tracking, 3-D interpretation) with a focus on level set methods, and written text recognition with a focus on neural networks methods.

**Mohamed Ben Salah** was born in 1980. He received the State engineering degree from the École Polytechnique, Tunis, Tunisia, in 2003, the M.S. degree in computer science from the National Institute of Scientific Research (INRS-EMT), Montreal, Quebec, QC, Canada, in January 2006, where he is currently pursuing the Ph.D. degree in computer science.

His research interests are in image and motion segmentation with focus on level set and graph cut methods and shape priors.

**Ismail Ben Ayed** (M'04) received the Ph.D. degree in computer science from the Institut National de la Recherche Scientifique, Montreal, Quebec, QC, Canada.

His research is in computer vision, pattern recognition, and medical image analysis. He is currently a Scientist with GE Healthcare, London, ON, Canada.