IRGS: Image Segmentation Using Edge Penalties and Region Growing

Qiyao Yu and David A. Clausi, *Senior Member*, *IEEE*

Abstract—This paper proposes an image segmentation method named iterative region growing using semantics (IRGS), which is characterized by two aspects. First, it uses graduated increased edge penalty (GIEP) functions within the traditional Markov random field (MRF) context model in formulating the objective functions. Second, IRGS uses a region growing technique in searching for the solutions to these objective functions. The proposed IRGS is an improvement over traditional MRF-based approaches in that the edge strength information is utilized and a more stable estimation of model parameters is achieved. Moreover, the IRGS method provides the possibility of building a hierarchical representation of the image content and allows various region features and even domain knowledge to be incorporated in the segmentation process. The algorithm has been successfully tested on several artificial images and synthetic aperture radar (SAR) images.

Index Terms—Markov random field (MRF), Gaussian mixture, hybrid region and edge, region growing, Region Adjacency Graph (RAG).

✦

1 INTRODUCTION

IMAGE segmentation is a process that decomposes an image into disjoint regions and is a fundamental step for many image-processing tasks such as image understanding and content-based image retrieval. In general, image segmentation aims at producing regions that are homogeneous with respect to the extracted features, such as gray level or texture, and have significant different feature values across region boundaries. Due to the existence of noise in measurements and randomness of the features, pure feature-based segmentation approaches such as histogram thresholding [1] and clustering [2] often produce noisy results and there are numerous other methods that more or less utilize the spatial context information either implicitly or explicitly. Some examples are edge-based methods [3], [4], [5], [6], region splitting and merging [7], [8], and model-based methods such as curve evolution [9], [10], [11] and random fields [12], [13], [14], [15], [16], [17], [18]. These approaches all have certain attractive features but also have drawbacks. The edge-based methods are insensitive to image nonstationarity and are efficient in describing local behaviors, but are ineffective in producing results globally meaningful. The region splitting and merging category provides the possibility of incorporating a variety of regional features but often has difficulty in determining suitable merging and stopping criteria for a result that is neither oversegmented nor undersegmented. The model-based methods have an established mathematical foundation but require the model to be accurate and the optimization process

to be able to find a satisfactory solution efficiently, both of which are difficult for complex scenes.

In this paper, an image segmentation method named **iterative region growing using semantics** (IRGS) is proposed, based on our previous work in [19]. The IRGS is characterized by two aspects: 1) It uses a sequence of edge penalty functions to approximate the traditional **Markov random field** (MRF) context model in formulating the objective functions and 2) it uses a region growing technique in searching for the solutions to those objective functions. The MRF has been popular in modeling image spatial context and, with a feature model combined under the Bayesian framework, provides a statistically sound formulation of the segmentation problem. Starting from such an MRF-based formulation, the IRGS combines the attractive features of edge-based and region-growing methods. The motivation is fourfold.

1. Edge information should be efficiently used during the optimization process of the traditional MRF segmentation. For a complex scene, whether or not the model is sufficiently accurate to describe the image contents is often unclear. As a result, the global minimum of the objective function of model-based approaches may not correspond to the desired segmentation. In fact, a satisfactory solution is not necessarily the global minimum, but often local minima obtained by guiding the optimization process with good initial conditions or any helpful information. The edge strength is one such helpful feature and, hence, the IRGS incorporates the edge strength into the traditional MRF model. Such a model has something similar to the conditional random fields [15], [16] in that the interactions in labels are data dependent. However, the IRGS uses a sequence of models instead of a single one, which is fundamentally different from existing conditional random field approaches and other hybrid region and edge segmentation work [20], [21], [22] as well.

- *Q. Yu is with Eutrovision Inc., Room 13E, 468 Caoxi North, Shanghai, 20030, P.R. China. E-mail: qiyao.yu@eutrovision.com.*
- *D.A. Clausi is with the Department of Systems Design Engineering, University of Waterloo, Waterloo, ON, N2L 3G1, Canada. E-mail: dclausi@engmail.uwaterloo.ca.*

2. The role of the spatial context model should be adaptively weighted. The MRF spatial context model favors the configuration composed of large regions that are homogeneous in the defined features. Two related tasks are then important for obtaining a satisfactory solution to such an MRF-based formulation—the parameter estimation of the feature model and the optimization process for the segmentation. For the unsupervised segmentation problem, obtaining the feature model parameters a priori by training is not practical and those parameters are usually estimated and refined iteratively from intermediate segmentation results. However, simultaneous parameter estimation and segmentation often gives erroneous results due to the unpredictable role of the spatial context model on inaccurate configurations. At the initial stages of iterations, where the feature model parameters are far from the true values, incorporation of the spatial context model may produce meaningless segmentation, which in turn can make the resulting updated feature model parameters even worse. To alleviate this problem, Deng and Clausi [23] use a variable weighting scheme between the feature and spatial context model so that the influence of spatial context is minimum initially and then gradually increased. The IRGS provides a similar but different scheme.

3. Multiscale segmentation is desirable. Due to the different goals and models of the two tasks of image segmentation and interpretation, regions obtained by the segmentation process may not match the real objects well enough for an accurate subsequent interpretation. For example, it is often difficult to determine a suitable scale so that the segmentation result is neither oversegmented nor undersegmented with respect to the interpretation need. A possible solution to this is to generate multiple segmentation results [24] corresponding to multiple scales and let the interpretation process choose the best one. Such a multiscale segmentation can be provided by IRGS as it uses a sequence of objective functions with the solutions corresponding to segmentations of various coarseness.

4. Data should be taken into account at a region level. Pixel-based combinatorial optimization techniques [12], [14], [25], [26], [27] are either extremely slow or are easily trapped into local minima. Large moves in the searching space are thus desirable and can be achieved by performing configuration changes on groups of pixels instead of each single pixel during the optimization process. The grouping structure of pixels can be fixed [28], [29], [30], [31] or, more attractively, data-adaptive [32], [33], [34] so that the correspondence between the image structures and true objects can be possibly established. The IRGS uses a region growing technique to generate a hierarchical data-adaptive structure and is different from those approaches, as well as other model-based region growing work [34], [35], [36].

The organization of the paper is given as follows: Section 2 reviews the MRF-based image segmentations. Section 3 improves the MRF work using a sequence of edge penalty functions which, combined with a region growing technique, leads to the IRGS method presented in Section 4. Section 5 is the experiments and discussions. Summary and future work comprise Section 7.

2 IMAGE SEGMENTATION AND MRF CONTEXT MODEL

2.1 Problem Statement and Notations

Let $\mathcal{S}$ denote the discrete rectangular lattice on which images are defined. Suppose there are n different classes in the image to be segmented. $\mathbf{X} = \{X_s | s \in \mathcal{S}\}$ is a set of discrete valued random variables constituting a random field on $\mathcal{S}$, with each variable X_s taking a value in $\{1, \ldots, n\}$ representing the class to which the site s belongs. $\mathbf{Y} = \{Y_s | s \in \mathcal{S}\}$ is another random field somehow related to $\mathbf{X}$ and the observed image is a realization from $\mathbf{Y}$.

Let $\mathbf{x} = \{x_s | s \in \mathcal{S}\}$ and $\mathbf{y} = \{y_s | s \in \mathcal{S}\}$ denote the realizations of $\mathbf{X}$ and $\mathbf{Y}$, respectively. The image segmentation is an inverse process that attempts to estimate the best $\mathbf{x}$ given the observed image $\mathbf{y}$. With the obtained class labels $\mathbf{x}$, $\mathcal{S}$ is segmented to n classes, $\Omega_1, \ldots, \Omega_n$, such that

$$
\begin{aligned}
a) \quad & \Omega_i = \{s | X_s = i, s \in \mathcal{S}\}, \\
b) \quad & \bigcup_{i=1}^{n} \Omega_i = \mathcal{S}, \\
c) \quad & \forall i \neq j : \Omega_i \cap \Omega_j = \emptyset.
\end{aligned}
\tag{1}
$$

2.2 MRF-Based Formulation of Image Segmentation

The image segmentation task can be formulated as a maximum a posterior (MAP) problem for which maximizing the posterior $P(\mathbf{x}|\mathbf{y})$ gives a solution. By the Bayes' rule, this is equivalent to maximizing $p(\mathbf{y}|\mathbf{x})P(\mathbf{x})$. Two models are used for analytically representing $p(\mathbf{y}|\mathbf{x})$ (the feature model) and $P(\mathbf{x})$ (the spatial context model).

For the spatial context model $P(\mathbf{x})$, an MRF model named the **multilevel logistic model** (MLL) has been popular [13]. The MRF theory provides a way to model the joint probability distribution of the image sites in terms of local spatial interactions, which are analytically expressed by clique energy functions [17]. The clique energy of MLL is defined as

$$
V_2(x_s, x_t) = \begin{cases} \beta & \text{if } x_s \neq x_t \\ 0 & \text{otherwise,} \end{cases}
\tag{2}
$$

where s and t are neighboring sites forming a pair-site clique and β is a positive number. Such a model makes the prior $P(\mathbf{x})$ large if the local neighborhood is dominated by one single class and small otherwise and, hence, is effective in suppressing noisy configurations of class labels.

In modeling $p(\mathbf{y}|\mathbf{x})$, y_s is often assumed to be a constant gray level (related to x_s), corrupted by additive independent noise. Furthermore, $p(y_s|x_s)$ is often assumed to be a Gaussian function for simplicity. With both the feature model and the spatial context model defined, the MAP formulation of the segmentation task is transformed into minimizing an energy

$$
E = E_f + E_s,
\tag{3}
$$

where

$$
E_f = \sum_{s \in \mathcal{S}} V_1(x_s)
\tag{4}
$$

and

$$E_s = \sum_{<s,t> \in \mathcal{C}} V_2(x_s, x_t), \tag{5}$$

where $\mathcal{C}$ is the set of all cliques on the entire lattice $\mathcal{S}$. In the above equations, we have

$$V_1(x_s) = \frac{1}{2}\ln(2\pi\sigma_{x_s}^2) + \frac{(y_s - \mu_{x_s})^2}{2\sigma_{x_s}^2}, \tag{6}$$

where μ_i and σ_i^2 are the mean and variance of the pixel values in class i, and $V_2(x_s, x_t)$ is defined in (2). Substituting $(6)^1$ and (2) into (3), the image segmentation problem is formulated as

$$\arg\min_{\{x_s, s \in \mathcal{S}\}} \left\{ \sum_{s \in \mathcal{S}} \left\{ \frac{1}{2}\ln(2\pi\sigma_{x_s}^2) + \frac{(y_s - \mu_{x_s})^2}{2\sigma_{x_s}^2} \right\} + \beta \sum_{<s,t> \in \mathcal{C}} \{1 - \delta(x_s, x_t)\} \right\}, \tag{7}$$

where $\delta(\cdot)$ is the Kronecker delta function.

2.3 Optimization and Parameter Estimation

Finding a solution for (7) represents a combinatorial optimization problem. Various combinatorial optimization techniques are known, including iterated conditional mode (ICM) [12], simulated annealing (SA) [14], mean field theory [27], genetic algorithm [25], belief propagation [26], and graph theoretic techniques [33]. The **ICM** [12] is the simplest among all of these. Utilizing the local dependence among pixels, the ICM iteratively refines the label configurations based on the provisional estimate of those configurations, accepting changes that decrease the energy E the most. Its greedy nature makes it easily trapped in local minima and sensitive to the initial estimate. A better and perhaps the most widely used approach is **simulated annealing** (SA) [14]. Unlike ICM, SA employs a random search scheme, for example, Metropolis-Hastings sampling [37], which accepts or rejects changes not deterministically but based on a probability:

$$e^{-\delta E/T},$$

where δE is the change of energy if the new configuration is accepted and T is a parameter named *temperature*, irrespective of the energy function involved. The SA process simulates the physical process of annealing by slowly decreasing the temperature T to force the system into lower energy states. With a certain cooling schedule (annealing schedule), the SA can be guaranteed to find the global minimum [14]. However, the schedules leading to this guarantee are impossibly slow and there is a trade-off between the cooling speed and the quality of the solution.

In the above ICM or SA optimization process for (7), several parameters are required to be known or estimated. These include the number of classes n, the Gaussian parameters μ_i and σ_i^2 for each class i, and the MLL parameter β.

The expectation-maximization (EM) algorithm [38] is typically used to determine the parameters for the Gaussian mixture and, hence, can be used to estimate μ_i and σ_i^2. Let π_i denote the prior for class i and w_{si} represent the probability of site s belonging to class i given the observed data and current estimate of all parameters. The EM algorithm computes the Gaussian parameters through iterating the E step,

$$w_{si} = \frac{p(y_s|\mu_i, \sigma_i)\pi_i}{\sum_{i=1}^{n} p(y_s|\mu_i, \sigma_i)\pi_i}, \tag{8}$$

and the M step,

$$\mu_i = \frac{\sum_{s \in \mathcal{S}} w_{si} y_s}{\sum_{s \in \mathcal{S}} w_{si}}, \tag{9}$$

$$\sigma_i^2 = \frac{\sum_{s \in \mathcal{S}} w_{si}(y_s - \mu_i)^2}{\sum_{s \in \mathcal{S}} w_{si}}, \tag{10}$$

$$\pi_i = \frac{\sum_{s \in \mathcal{S}} w_{si}}{\sum_i \sum_{s \in \mathcal{S}} w_{si}}. \tag{11}$$

In such an estimation of μ_i and σ_i^2 using (8)-(11), the spatial context has not been considered. Therefore, μ_i and σ_i^2 may seem necessary to refine based on intermediate segmentation results. That is, during each iteration of the optimization process, the parameters are updated based on the current segmentation result using the same M step, as in (9)-(11), but a different E step, where

$$w_{si} = \begin{cases} 1 & \text{if } x_s = i, \\ 0 & \text{otherwise.} \end{cases} \tag{12}$$

However, such an updating scheme has problems. At the initial stages of iterations where the feature model parameters are not close enough to the true values, incorporation of the spatial context model may produce poor segmentation, which in turn can make the resulting updated feature model parameters even worse. During the optimization process, the point at which the spatial context model begins to have a positive role and the point at which the parameter updating should begin is unclear. This is an issue addressed by the proposed method in Section 3.

Traditionally, β is a constant, set a priori using an experimentally satisfactory value. For example, Rignot and Chellappa [39] reported that [1.0-1.6] is the best range for β in polarimetric synthetic aperture radar (SAR) image segmentation. Yue [40] concluded that β could be set between 1 and 3 for SAR sea ice imagery. However, a single β cannot describe images of even a moderate range of complexity. The β should be set as a large value for simple scenes and small for complex scenes. Descombes et al. [41] derive the following mathematical formulation of the relationship between the β and the boundary length using maximum likelihood estimation:

$$N(\mathbf{x}) = \sum_{\mathbf{x}} N(\mathbf{x}) P(\mathbf{x}|\beta), \tag{13}$$

where $N(\mathbf{x})$ is the total number of pairs of neighboring sites that are of different labels. Equation (13) demonstrates that β should choose the value by which the expectation of the boundary length over all possible configurations $\mathbf{x}$ is equal to the current. A Monte Carlo scheme is then used to estimate β.

---

1. Please note that (4) alone defines a Gaussian mixture problem.

The number of classes n can be determined by explicitly incorporating into the objective function a penalty term for large n [18], [24], based on certain criterion such as the minimum description length [42]. However, establishing n is by no means a resolved issue. For the experiments in this paper, n is assumed to be known a priori.

2.4 Extending to Region-Based Segmentation

More generally, the MRF can be defined on irregular graphs rather than the regular image lattice. This allows the image segmentation problem formulated by (7) to be based on a set of interconnected groups of pixels (referred to here as "regions"), with the MRF spatial context model based on a region adjacency graph (RAG) [17]. Here, the labeling is not on single pixels but on regions, where the regions are commonly obtained by a deliberate oversegmentation. Each node in the RAG represents a region and a link between the nodes represents the existence of a common boundary between the regions. Defined on the RAG, the MRF models the behaviors of the regions in a similar way as for pixels. Let R_i denote node i in the graph and let x_i denote the label for all sites $s \in R_i$. The feature model energy for R_i can be defined as

$$V_1(x_i) = \sum_{s \in R_i} \left\{ \frac{1}{2} \ln(2\pi\sigma_{x_i}^2) + \frac{(y_s - \mu_{x_i})^2}{2\sigma_{x_i}^2} \right\} \qquad (14)$$

and the MRF pair site clique energy for two neighboring nodes R_i and R_j is

$$V_2(x_i, x_j) = \begin{cases} \sum_{\substack{<s,t> \in \mathcal{C} \\ s \in R_i, t \in R_j}} \beta & \text{if } x_i \neq x_j \\ 0 & \text{otherwise.} \end{cases} \qquad (15)$$

Summation of the above energies over the entire RAG, similarly to (3)-(5), gives exactly (7). A combinatorial optimization techniques is then applied to RAG nodes instead of pixels. Such a region-based segmentation method is advantageous in computation speed as the number of RAG nodes is usually significantly less than the number of pixels. The initial oversegmentation and RAG construction time can be relatively trivial.

3 GRADUATED INCREASED EDGE PENALTY (GIEP)

This section focuses on one aspect of the proposed IRGS method: using a sequence of edge penalty functions to approximate the MRF spatial context model. This aspect alone provides a basis for a novel segmentation method and it will be referred to as the **graduated increase edge penalty** (GIEP). The GIEP has two attractive features. First, it utilizes edge information to improve the segmentation on nonstationary situations. Second, it provides a simple and elegant method of simultaneously estimating model parameters and searching solutions for the MRF-based formulation in (7).

Since the last summation term in (7) is nonzero only at boundary sites (that is, having at least one neighbor belonging to a different region), the corresponding MRF spatial context model functions as a penalty for the existence of boundary site pairs. Instead of penalizing equally for all boundary site pairs, a greater penalty can be applied to weak edge and a lesser penalty to strong edge so that local statistic, such as edge strength, can be incorporated. Therefore, the penalty term can be replaced with

some monotonically decreasing function $g(\cdot)$ (defined in Section 3.1) of the strength of the edge between the two neighboring sites astride the boundary. The clique energy function in (2) is now changed to

$$V_2(x_s, x_t) = \begin{cases} \beta g(\nabla_{st}) & \text{if } x_s \neq x_t, \\ 0 & \text{otherwise,} \end{cases} \qquad (16)$$

where the edge strength $\nabla_{st} = |y_t - y_s|$. Correspondingly, the new objective function can thus be defined as

$$\arg \min_{\{x_s, s \in \mathcal{S}\}} \left\{ \sum_{s \in \mathcal{S}} \left\{ \frac{1}{2} \ln(2\pi\sigma_{x_s}^2) + \frac{(y_s - \mu_{x_s})^2}{2\sigma_{x_s}^2} \right\} + \beta \sum_{<s,t> \in \mathcal{C}} \left\{ (1 - \delta(x_s, x_t)) g(\nabla_{st}) \right\} \right\}. \qquad (17)$$

Although inspired by the MRF spatial context model, (17) can be viewed as a cost function by which the segmentation problem is formulated more generally without the Markov constraint and the Bayesian interpretation. Similar cost functions can also be found in the research literature, for example, [43]. However, the cost function (17) may produce biased results for image segmentations. This is clearer if we return to the Bayesian framework and view (16) as a spatial context prior model.[2] Obviously, this prior favors configurations in which the mean pixel values of the classes are more apart. Although a suitable $g(\cdot)$ can possibly be selected to make the bias under control for a given class of images, there is no clue as to how to choose this single $g(\cdot)$ for a variety of scenes and applications. Therefore, instead of using a single edge penalty function, a sequence of functions is designed to approach to the nonbiased MLL prior model, as shown in Section 3.1. This is one major difference between the proposed method and other published approaches.

3.1 Choosing Edge Penalty Function

The edge penalty function $g(\cdot)$ can be any monotonically decreasing function so that the greater the edge strength is, the smaller the penalty. Suppose the edge strength ∇_{st} for any boundary site pair s and t has been normalized to [0, 1]. Then, the penalty function can be formulated as

$$g(\nabla_{st}) = e^{-(\nabla_{st}/K)^2}. \qquad (18)$$

The parameter K (Fig. 1) defines how fast the edge penalty decays with the increase of edge strength. As K increases, the penalty difference between weak and strong edges decreases. When K approaches infinity, all edge penalties are equally 1.

Thus, the defined sequence of edge penalty functions is characterized by two aspects. With K gradually increased, we have the following:

- The sequence penalizes weak edges and strong edges differently, but the penalty difference is gradually decreased.
- The sequence gradually increases the penalty for any edge.

2. Here, we ignore the fact that a prior model, by definition, should not depend on the data.
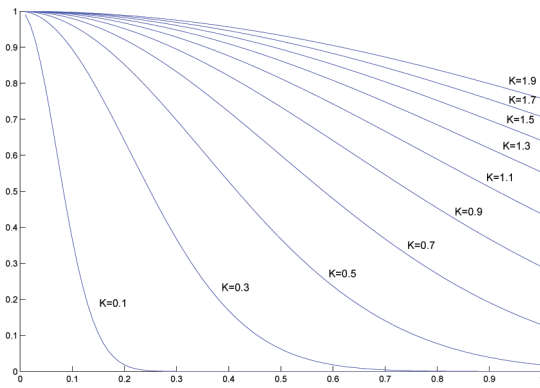
Fig. 1. Edge penalty functions defined by (18).

3.2 Overall Algorithm

A nice property of using such a sequence of edge penalty functions is that it provides a simple and elegant transition from the standard Gaussian mixture problem to the MRF-based formulation. From the point of view of a solution searching for (7), this allows an effective simultaneous parameter estimation and optimization, which is less influenced by the possibly negative role of spatial context model at the initial stages, as described earlier in Section 2.3. The overall process of GIEP segmentation is summarized in Table 1.

Starting from the extreme case of the Gaussian mixture problem,[3] the algorithm in Table 1 introduces more and more edge penalty during the solution searching. When K is sufficiently large, the energy function is close to (7) and an approximate solution to (7) is obtained by the algorithm.

The incremental schedule for the penalty function parameter K is desired to be very slow in order to carefully approximate the transition from the Gaussian mixture problem to the final objective function of (7). On the other hand, the processing time is required to be tolerable for the method to be practically useful. A trade-off has to be made. In all of the experiments in this paper, K is chosen to be

$$K^{(t+1)} = \begin{cases} f(2t) & \text{if } 0 \le t < 50, t \in I \\ 1.1 K^{(t)} & \text{if } t \ge 50, t \in I, \end{cases} \quad (19)$$

where $t = 0, 1, \ldots$ is the sequence number and $K^{(0)} = 0$. $f(i)$ is the value to which i percent of the site pairs have weaker edge strength in comparison. Such a schedule is experimentally satisfactory.

As in Section 2.4, a region-based GIEP is also possible. The same single-node clique energy as in (14) is used and a different pair-node clique energy is defined as

$$V_2(x_i, x_j) = \begin{cases} \beta \sum_{\substack{<s,t>\in \mathcal{C} \\ s\in R_i, t\in R_j}} g(\nabla_{st}) & \text{if } x_i \ne x_j \\ 0 & \text{otherwise.} \end{cases} \quad (20)$$

3. The extreme case corresponds to $K = 0$, where the edge penalty function has a nonzero value only at the zero edge strength point. Therefore, strictly speaking, it is not an exact Gaussian mixture problem but one with penalties on situations where neighboring sites with zero edge strength between them belong to different classes. Such situations will never happen in practice and the two problems are equivalent.

4 ITERATIVE REGION GROWING USING SEMANTICS

In this section, we extend the GIEP method in Table 1 in a hierarchical manner, leading to the IRGS method. Many researchers have applied the MRF model on a hierarchical structure of pixels instead of single pixels [28], [29], [30], [31], since pixel-based methods are either easily trapped in local minima or extremely slow in convergence. However, the fixed hierarchical structure in those approaches does not allow efficient descriptions of the underlying image contents and may produce undesirable artifacts. There are a number of papers that construct data-adaptive structures using graph cuts [32], [33] or region growing [34] techniques. The sequence of objective functions described in the preceding section are introduced naturally as the merging criteria corresponding to various scales and, here, a simple region growing technique is used to construct the hierarchical data-adaptive structures of the image for optimization purposes.

4.1 Region Growing

Similarly to the GIEP, the proposed IRGS method is also iterative, with each iteration an optimization process for finding a local minimum of the objective energy function (17) for a given K. Consider a hierarchical clustering process for such an optimization purpose. The process begins with a deliberate oversegmentation configuration with many classes and tries to reduce to the true number of classes. Examining each pair of classes, the energy (17) of the configuration obtained by merging the two classes is computed and compared with that before merging. If there is a decrease, the merging is justified. In computing such an energy difference, energy terms unrelated to the current two classes can be cancelled and a simple merging criterion can be derived from (17). Suppose two classes, Ω_i and Ω_j, are being investigated. Let $\Omega_k = \Omega_i \bigcup \Omega_j$ denote the class obtained by merging and N_i denote the number of pixels belonging to Ω_i. The energy difference

$$
\begin{aligned}
\delta E_{ij} =& \sum_{s\in\Omega_k} \ln(\sigma_k) + \sum_{s\in\Omega_k} \frac{(y_s - \mu_k)^2}{2\sigma_k^2} - \sum_{s\in\Omega_i} \ln(\sigma_i) \\
&- \sum_{s\in\Omega_i} \frac{(y_s - \mu_i)^2}{2\sigma_i^2} - \sum_{s\in\Omega_j} \ln(\sigma_j) \\
&- \sum_{s\in\Omega_j} \frac{(y_s - \mu_j)^2}{2\sigma_j^2} - \beta \sum_{\substack{<s,t>\in\mathcal{C} \\ s\in\Omega_i, t\in\Omega_j}} g(\nabla_{st}) \\
=& \sum_{s\in\Omega_k} \ln(\sigma_k) + \frac{1}{2}N_k - \sum_{s\in\Omega_i} \ln(\sigma_i) - \frac{1}{2}N_i \\
&- \sum_{s\in\Omega_j} \ln(\sigma_j) - \frac{1}{2}N_j - \beta \sum_{\substack{<s,t>\in\mathcal{C} \\ s\in\Omega_i, t\in\Omega_j}} g(\nabla_{st}) \\
=& \sum_{s\in\Omega_k} \ln(\sigma_k) - \sum_{s\in\Omega_i} \ln(\sigma_i) - \sum_{s\in\Omega_j} \ln(\sigma_j) \\
&- \beta \sum_{\substack{<s,t>\in\mathcal{C} \\ s\in\Omega_i, t\in\Omega_j}} g(\nabla_{st}).
\end{aligned}
\quad (21)
$$

If (21) gives a negative value, Ω_i and Ω_j can be merged. If (21) gives a nonnegative value, Ω_i and Ω_j cannot be merged. Such a hierarchical clustering does not guarantee to generate a given number of classes. A labeling process is

TABLE 1
Algorithm of
Graduated Increased Edge Penalty Segmentation (GIEP)

1.	Initially, $K = 0$. Assign each RAG node (representing a region or a pixel) with a random label.
2.	Based on the current segmentation result, compute the new feature model parameters using (9)-(12).
3.	Refine the segmentation. The process scans each RAG node in a random order, and for each scan chooses the label that produces the minimum energy for (17).
4.	If a maximum number of iterations has not been completed, increase K and go back to step 2.

TABLE 2
Algorithm of the
Iterative Region Growing on Semantics (IRGS) Segmentation

1.	Initially, $K = 0$. Construct an initial RAG and assign each node with a random label.
2.	Based on the current segmentation result, compute the new feature model parameters using (9)-(12).
3.	Compute the energy difference δE using equation (22) for each pair of linked nodes in the RAG that have the same class label, and find the minimum δE_{min} for each.
4.	If $\delta E_{\min}$ is negative, merge the corresponding two nodes and go back to step 3.
5.	Perform the labeling on the new RAG. The process scans each node in the RAG in a random order, and for each node scanned chooses the label that produces the minimum energy for (17).
6.	If a maximum number of iterations has not been completed, increase K and go back to step 2.

then necessary to assign a given number of labels to the clusters obtained.

A sequential merging order needs to be defined since the merging cannot be performed simultaneously. A greedy way is to find the two classes of which the merging decreases the energy most and merge them first. At each iteration, the merging begins on a configuration obtained by a more conservative criterion from the previous iteration and continues until the energy cannot be reduced.

In practice, letting each region be a unique class for the initial oversegmented configuration is convenient. The IRGS in this paper always uses such an initial configuration and, hence, the hierarchical clustering is essentially a region growing process. The energy difference between the two neighboring regions R_i, R_j and its union $R_k = R_i \bigcup R_j$ is thus

$$\delta E_{ij} = \sum_{s \in R_k} \ln(\sigma_k) - \sum_{s \in R_i} \ln(\sigma_i) - \sum_{s \in R_j} \ln(\sigma_j) \\ - \beta \sum_{\substack{<s,t> \in \mathcal{C} \\ s \in R_i, t \in R_j}} g(\nabla_{st}). \quad (22)$$

4.2 Overall Algorithm

The overall algorithm is described in Table 2. Compared to the GIEP algorithm in Table 1, the IRGS is different only by an additional region merging process (corresponding to Steps 3 and 4) at each iteration, which reduces the dimension of the search space. In Step 3, energy differences are only computed between regions with the same class label. Two regions are not allowed to be merged if they belong to different classes, the purpose of which is to suppress the merging between parts of different objects that have weak boundaries in between. This concept is known as semantic region growing [44], [45] and similar ideas also appear in other approaches [32]. Here, the merging and labeling are iterative and, as such, are referred to as *IRGS*. Although the labeling process in this paper has no semantic meanings, it is possible to replace it with a domain specific labeling process and to integrate high-level knowledge into this system [46].

4.3 Estimating the MLL Parameter

Equation (13) demonstrates that the MLL parameter β should only be determined by the boundary length expected in the image. This seems doubtful because β also needs to be adaptive to the noise strength of the image to obtain a good segmentation. For example, if there is sufficient overlapping of the pixel intensities between the classes in the image (that is, noise strength is high), it is quite likely that the global minimum corresponds to an

undersegmentation (for example, all of the sites belong to one single class) result if the β is not reduced to a sufficiently small value. However, this phenomenon does not mean that the estimation of β is wrong, but, rather, indicates the inaccuracy of the MLL as the prior. Nevertheless, it is possible to adjust the MLL parameter to achieve a good segmentation.

Let β represent the MLL parameter estimated by the method in [41] and a different notation $\hat{\beta}$ represent its adjustment. In the IRGS, the MLL parameter is estimated by the following:

$$\hat{\beta} = C_1 \frac{J/C_2}{1 + J/C_2} \beta, \quad (23)$$

where C_1 and C_2 are two constants and J is the minimum of the Fisher criterion for any two classes in the image:

$$J = \min_{i,j} \frac{(\mu_i - \mu_j)^2}{\sigma_i^2 + \sigma_j^2}. \quad (24)$$

J gives a measure of the separability of the classes in the image with respect to the gray level. The equation gives a small $\hat{\beta}$ when J is small (corresponding to high noise strength) and vice versa. At the initial stages of the segmentation, when spatial context information does not play an important role and the regions are obtained mostly based on features, J is relatively large (compared to those obtained later). When the merging proceeds, J is generally decreasing, causing $\hat{\beta}$ to be smaller. Therefore, the spatial correlations decreases as the scales increases, which agrees with intuition.

On the other hand, it seems counterintuitive that the strength of the interactions in the spatial context model should be small when the noise strength is high. However, such an intuition is based on considering the problem in a local manner. For example, methods that update labels pixel by pixel generally do not suffer the undersegmentation problem and produce similar results using large MLL parameters [47]. However, such a nice property of pixel-based methods is due to its being trapped into a local minimum, which is not theoretically desirable.

5   EXPERIMENTS AND DISCUSSION

5.1   Experimental Setup

Five methods are tested and compared. The methods include

1. the **Gaussian mixture model**-based clustering (GMM),
2. the classical MLL approach that uses (7) as the objective function and SA for optimization with a constant (C) weighting between the feature and context model (C-MLL),
3. Deng and Clausi's method [23] that uses a variable (V) weighting between the feature and context model (V-MLL),
4. the GIEP described in Table 1, and
5. the IRGS described in Table 2.

Although there is no reason why the IRGS segmentation in Table 2 cannot begin with each pixel being a node of the RAG,[4] the construction of such a RAG has a high memory requirement. Therefore, a watershed segmentation [6] is performed first to provide an oversegmentation result upon which the initial RAG is constructed. However, this implies that, in addition to the overall model for merging and labeling, a different model has been introduced in the process. It is desirable that the initial watershed does not greatly influence the robustness and accuracy of the overall process and, hence, oversegmentation that produces very tiny regions is deliberately chosen. The degree of over-segmentation is determined by the variance of the Gaussian filter for smoothing the gradient magnitude, of which the local minima act as the seeds for the watershed process. In this paper, the Gaussian filter has a variance of 1.0 and highly oversegmented results are obtained initially. For comparison, the C-MLL, V-MLL, and GIEP here are also region-based using the same initial watershed result.

The annealing schedule, if applied, is always a simple geometric cooling $T^{(t)} = 0.98^t$ [48]. For all of the methods, the segmentation is initialized with random labels and multiple results have been obtained by multiple runs. For each method, the quality of those results is generally very similar and we present the best result.

The C-MLL first uses the feature model parameters estimated without considering the spatial context by (8)-(11), as in a pure GMM approach. Due to the unpredictable role of the spatial context model at the initial stages, the parameters are updated only when a sufficient number of iterations have been completed, achieving low randomness in the segmentation, i.e., the temperature is lower than 0.1. This implementation improved these segmentation results.

The $\hat{\beta}$ of (23) is used as the MLL model parameter for all methods if applied. The constants C_1 and C_2 in (23) are application dependent and obtained by trial and error. For synthetic images, $C_1 = 2$ and $C_2 = 0.4$. For SAR sea ice imagery, which has more textures and structures, $C_1 = 5$ and $C_2 = 0.4$. In computing $\hat{\beta}$, the original unadjusted MLL parameter β is estimated and refined iteratively by the method in [41] based on intermediate results along the segmentation process. Since all of the methods except GMM

can be viewed as trying to find the global minimum of the energy function (7), computing this energy to evaluate the quality of the result is helpful. However, different $\hat{\beta}$ are obtained by each different method, making any comparison of the final energy meaningless. Therefore, we choose the $\hat{\beta}$ obtained by the IRGS as the standard and run the other methods using this fixed $\hat{\beta}$ instead of estimating it by themselves in the segmentation process. The reason why $\hat{\beta}$ is estimated by the IRGS as opposed to any of the other methods is based on the observation that other methods (C-MLL, V-MLL, and GIEP) tend to give underestimates of the appropriate $\hat{\beta}$ when heavy noise exists in the image. A noisier image often corresponds to a noisier initial config-uration, from which an oversegmentation is likely to occur if the configuration changes are on single pixels or small groups of pixels. The $\hat{\beta}$ computed from the segmentation result is thus smaller than it should be and will then favor oversegmented configurations in turn. Except for the proposed IRGS method, all other tested methods are likely to give underestimates of $\hat{\beta}$ and, hence, $\hat{\beta}$ is set a priori for them for the purposes of comparing energies for final solutions, as per (7).

Percentages of correctly labeled pixels are also computed for comparisons of the segmentation methods. This requires correspondence between the segmentation labels and truth labels and we establish such a correspondence by making the classes in the obtained result have the same brightness order as that of the corresponding classes in the truth. That is, the brightest of obtained classes should correspond to the brightest class in the truth, the second brightest of obtained classes should correspond to the second brightest in the truth, and so on.

The methods have been tested on several synthetic images and numerous SAR sea ice images and some of these are presented here. All of the SAR sea ice images are captured by RADARSAT in ScanSAR (C-band and HH) mode and have 100-m pixel spacing. Ground truth is provided by ice maps produced by professional ice analysts at the Canadian Ice Service (CIS). Manual segmentations are produced based on the ice analysts observations. Sea ice is notoriously difficult to classify on a pixel-by-pixel basis, but manual segmentations based on ice analyst observations provide sufficient truth for quantitative evaluations here. The authors have a working relationship with CIS and a strong understanding of the challenges for SAR sea ice image segmentation that helps to motivate the practical aspects of this research.

5.2   Synthetic Images

The first experiment is on a synthetic image shown in Fig. 2a. The image has three different gray levels: 96, 144, and 160. The zero mean Gaussian noise of standard deviation 48 is then added, with the resulting image shown in Fig. 2b and the unimodal histogram shown in Fig. 2c. The noise causes a significant overlapping of the intensity ranges among the objects and background. A Gaussian mixture segmentation is first applied and the corresponding result shown in Fig. 2d is very noisy. Results of the various methods using spatial context models are then presented in Figs. 2e, 2f, 2g, and 2h. In Fig. 2e, noisy configurations have been greatly reduced by the C-MLL approach, but most of the triangle area is labeled incorrectly. The V-MLL method and the GIEP give improved labeling in Figs. 2f and 2g, respectively, with the feature

---

4. The only difficulty is the computation of σ_i in (22) if there is only one pixel in Ω_i. However, this can be solved by assigning a certain initial value to σ_i. For example, we can use the square root of the noise variance, which can be estimated by the mode of local variances of the entire image.
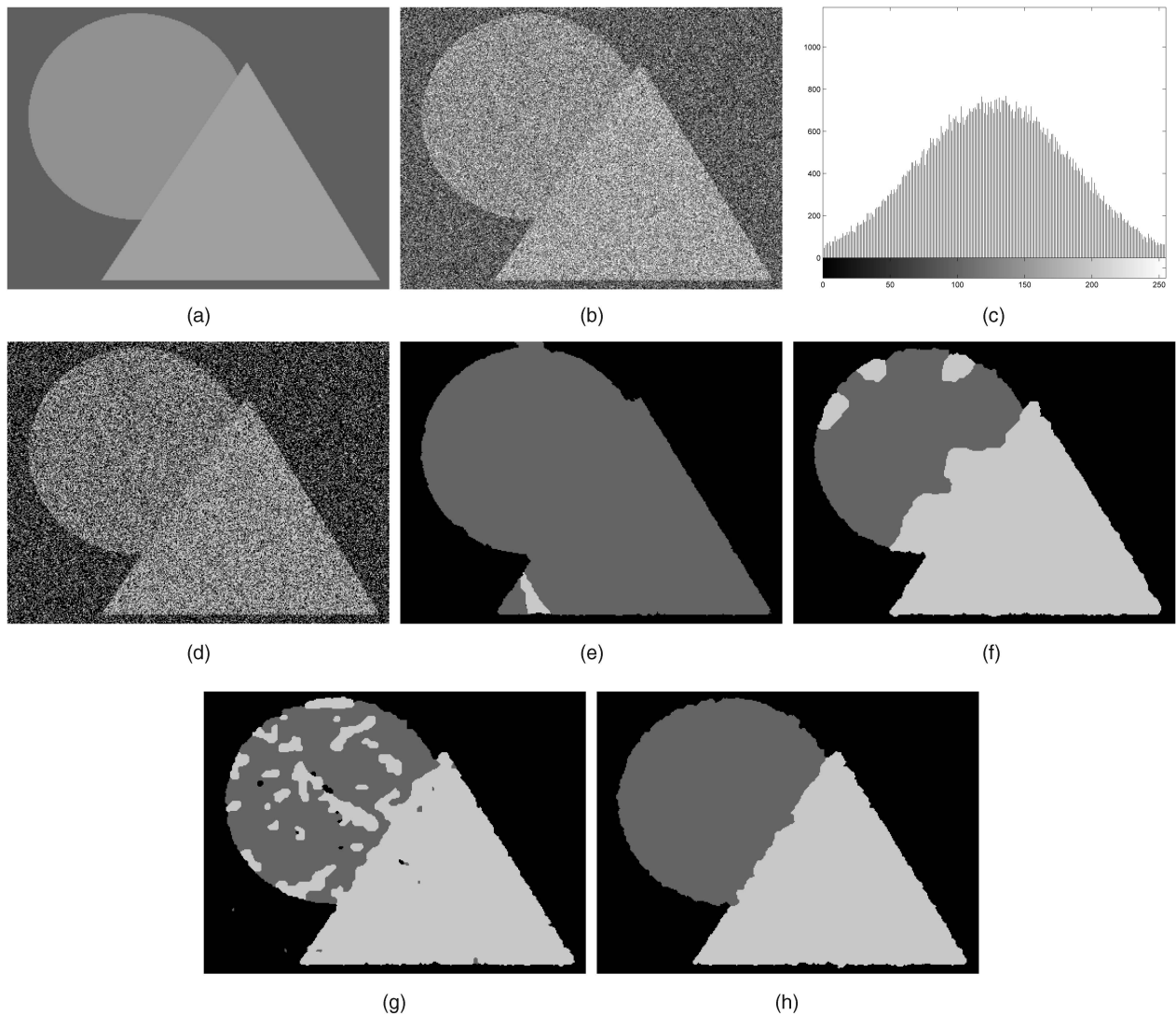
Fig. 2. Segmentations on an example synthetic noisy image. The three gray levels in the original image are 96, 144, 160, respectively. $\hat{\beta}$ obtained by IRGS is 0.27. (a) An example image. (b) With zero mean Gaussian white noise ($\sigma = 48$) added. (c) Histogram of the image in (b). (d) By GMM. Estimated mean: 60.0, 132.1, 205.4; Accuracy: 50.0 percent. (e) By C-MLL. Estimated mean: 96.6, 151.7, 169.8; Energy: 558363; Accuracy: 71.9 percent; Kappa: 0.57. (f) By V-MLL. Estimated mean: 96.6, 144.0, 157.7; Energy: 557647; Accuracy: 93.7 percent; Kappa: 0.90. (g) By GIEP. Estimated mean: 96.7, 142.7, 158.6; Energy: 558553; Accuracy: 93.1 percent; Kappa: 0.89. (h) By IRGS. Estimated mean: 96.4, 144.3, 160.1; Energy: 557210; Accuracy: 99.0 percent; Kappa: 0.98.

model parameters estimated more accurately at the same time (the means shown in the figure). However, their results still have obvious regional errors. Both methods update labels on groups of a limited number of pixels and without a very slow annealing schedule are trapped in a local minimum. A successful segmentation has been given by IRGS in Fig. 2h. The IRGS method is clear of any regional errors with only minor boundary errors. The IRGS has achieved an accuracy of 99.0 percent and a Kappa coefficient close to 1. The IRGS result also generates a lower energy than each of the other methods.

To show the necessity of using the adjusted MLL parameter $\hat{\beta}$ rather than β, we estimate from the truth the β, which is 4.49, and compute the corresponding energies for the truth and the obtained configurations in Figs. 2e, 2f, 2g, and 2h. The results are shown in Table 3. The minimum energy across all tests corresponds to the C-MLL result in Fig. 2e and, hence, indicates that the objective function

using the unadjusted MLL parameter β is inappropriate for the segmentation of the tested image.

Experiments are then performed on a set of synthetic images. We are particularly interested in the segmentation of SAR sea ice imagery and, hence, we try to synthesize images that are similar to SAR sea ice scenes. In the synthesis of these images, two different intensities are chosen, one relatively dark for the background and one relatively bright for the foreground objects. Those foreground objects have round shape with a limited amount of deformation and are generated at random locations in the

TABLE 3
Energies of the Configurations in Fig. 2
Using the Estimated β (4.49) from the Truth

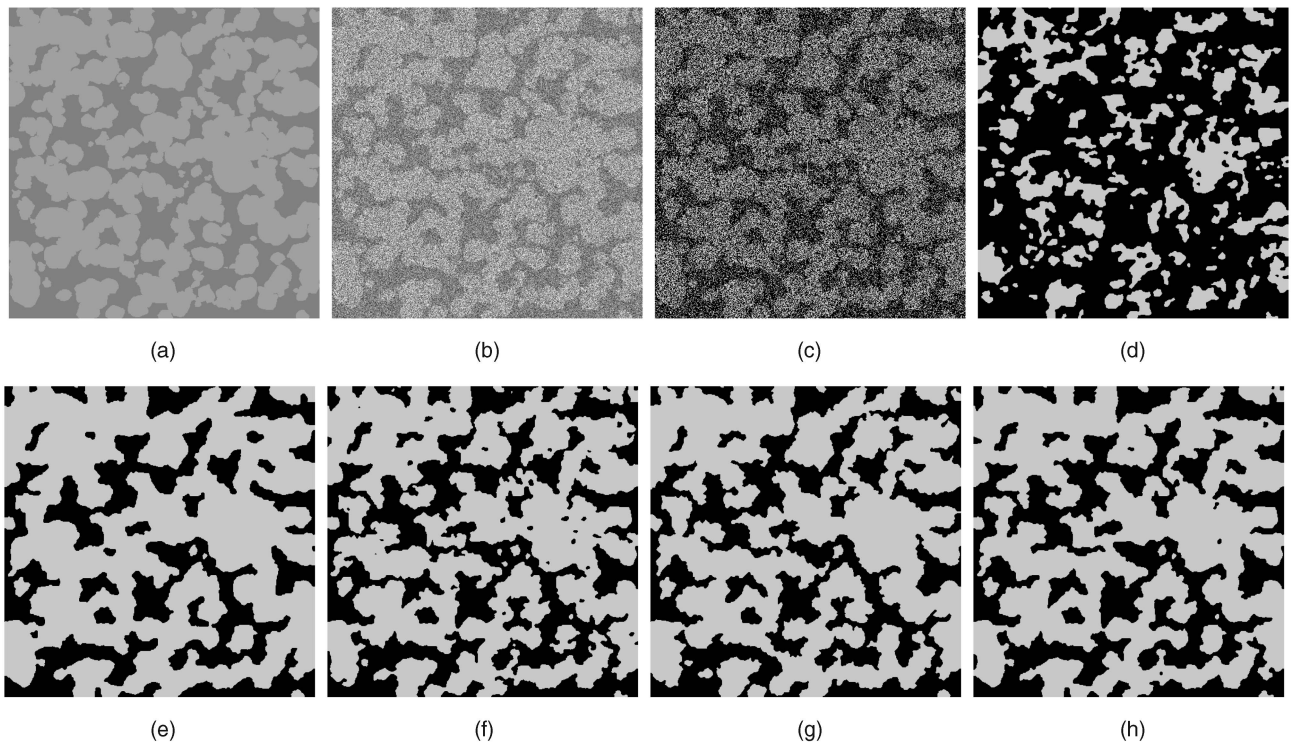| C-MLL	V-MLL	GIEP	IRGS	truth
573584	577274	595626	575841	574297

Fig. 3. Segmentation on an example 512 $\times$ 512 synthetic noisy images. The two gray levels in the original image are 128 and 160. The mean diameter of the simulated ice floes is 20 pixels. The $\hat{\beta}$ obtained by IRGS is 0.23. (a) Original. (b) Synthetic noisy image (unit mean multiplicative Gaussian noise, $\sigma_n^2 = 0.08$). (c) By GMM. Estimated means: 122.9, 194.0; Accuracy: 60.1 percent; Kappa: 0.23. (d) By C-MLL. Estimated means: 140.9, 164.9; Energy: 1.3891×10^6; Accuracy: 66.0 percent; Kappa: 0.384. (e) By V-MLL. Estimated means : 128.5, 157.8; Energy: 1.3791×10^6; Accuracy: 91.6 percent; Kappa: 0.817. (f) By GIEP. Estimated means: 129.5, 160.1; Energy: 1.3797×10^6; Accuracy: 92.4 percent; Kappa: 0.840. (g) By IRGS. Estimated means: 128.9, 159.5; Energy: 1.3792×10^6; Accuracy: 93.2 percent; Kappa: 0.855. (h) By C-MLL with true means. Energy: 1.3795×10^6; Accuracy: 92.2 percent; Kappa: 0.834.

TABLE 4
Summary of Segmentation Results (Accuracy/Kappa) for Synthetic Floe Images

σ_n^2	SNR	GMM	C-MLL	V-MLL	GIEP	IRGS
0.01	1.986	77.6% / 0.574	98.7% / 0.972	98.6% / 0.971	98.7% / 0.972	98.6% / 0.970
0.02	1.511	69.9% / 0.418	96.3% / 0.926	96.8% / 0.932	97.3% / 0.942	96.9% / 0.933
0.03	1.352	66.8% / 0.368	94.7% / 0.888	95.8% / 0.909	96.4% / 0.922	96.1% / 0.916
0.04	1.264	65.0% / 0.318	94.1% / 0.887	94.8% / 0.888	95.6% / 0.905	95.2% / 0.897
0.05	1.216	63.6% / 0.285	93.1% / 0.854	93.8% / 0.864	94.7% / 0.887	94.3% / 0.879
0.06	1.181	62.4% / 0.260	90.8% / 0.810	92.6% / 0.839	93.8% / 0.868	93.8% / 0.867
0.07	1.154	61.6% / 0.247	89.7% / 0.789	92.3% / 0.832	93.3% / 0.858	93.3% / 0.855
0.08	1.149	60.4% / 0.234	74.6% / 0.520	91.8% / 0.821	92.5% / 0.842	93.2% / 0.854
0.09	1.129	59.4% / 0.222	42.5% / 0.063	91.5% / 0.814	92.5% / 0.841	92.7% / 0.844
0.1	1.128	58.4% / 0.208	45.9% / 0.105	90.9% / 0.802	91.7% / 0.824	91.4% / 0.816
0.2	1.091	46.1% / 0.090	CF	88.0% / 0.740	75.4% / 0.526	88.6% / 0.755
0.3	1.083	48.7% / 0.119	CF	85.3% / 0.678	52.8% / 0.187	85.8% / 0.695
0.4	1.071	CF	CF	CF	CF	83.2% / 0.641
0.6	1.045	CF	CF	CF	CF	80.1% / 0.581

The gray levels in the original image are 128 and 160, respectively. The mean diameter of the simulated ice floes is 20 pixels. The image size is 512×512*. In the table, CF represents that at least one of the 20 samples has "convergence failure," meaning that the method cannot converge to a solution that segments the image into the specified number of regions (that is, all of the sites are labeled as one class).*

image. With speckle noise added, the resulting images are expected to give a rough simulation of SAR sea ice scenes characterized by water and well-defined ice floes. For adding the speckle noise common in SAR, Gaussian noise with mean of 1.0 is multiplied with the image intensities. Although SAR speckle is generally modeled as a Gamma distribution [49], multiplicative Gaussian noise is convenient to use and gives a reasonable approximation to SAR speckle. Fig. 3a shows an example simulated scene and Fig. 3b is the one with multiplicative Gaussian noise of variance $\sigma_n^2 = 0.08$ added.

For the experiments, 14 different levels of noise are chosen, and 20 scenes different in floe shape and location are generated, with each having an average floe diameter of 20 pixels. By combining each of the 20 scenes and each of the 14 different levels of noise, a total of 280 synthetic images are available for testing. Segmentations are then performed on these images using each method. Table 4 gives a summary of the average percentage accuracies and the Kappa coefficients of those methods at each noise level. When the noise variance is relatively low ($\sigma_n^2 \le 0.07$), all of the methods except GMM give satisfactory results. As noise
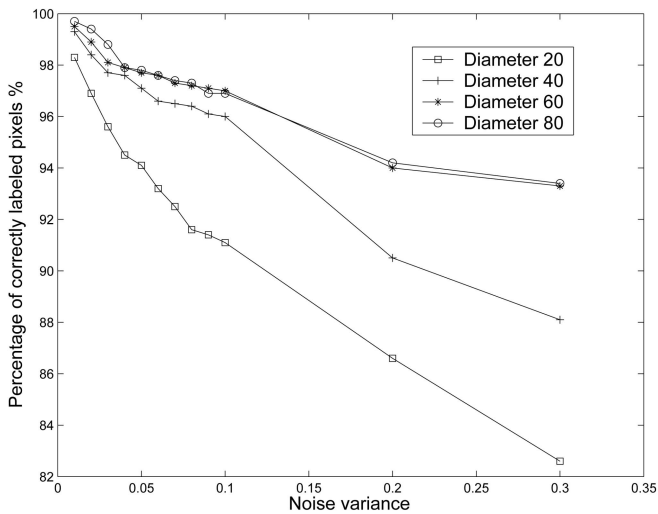
Fig. 4. Percentage of correctly labeled pixels versus noise variance for images of various floe diameters using IRGS.

variance increases, the performance of C-MLL drops quickly, whereas some others (V-MLL, GIEP, and IRGS) are more stable since they are less influenced by the unpredictable role of the spatial context model at the initial stages. The GIEP is deterministic compared to V-MLL and does not allow configuration changes on a hierarchical structure of pixels as IRGS does. As a result, GIEP is significantly inferior to V-MLL and IRGS when the noise variance increases beyond 0.2. The IRGS performs well for a wide range of noise strength and is comparable to the best method, if itself not the best, at all noise levels.

The segmentations on the image with $\sigma_n^2 = 0.08$ are shown in Figs. 3c, 3d, 3e, 3f, and 3g. At such a noise level, the C-MLL performs poorly due to the inaccurate feature model parameter estimation. In fact, accurate results can be obtained by C-MLL if the feature model means are fixed with the true values of 128 and 160, as shown in Fig. 3h. Although the IRGS has achieved the highest accuracy in Fig. 3g, the corresponding energy is higher than that of the V-MLL result in Fig. 3e. We have also found that the energy corresponding to the true configuration is 1.3793×10^6 and, hence, is also higher than the V-MLL result. Such a phenomenon implies that the MLL context model is not appropriate for describing the behaviors of the tested image. The MLL context model penalizes the existence of boundary site pairs and thus favors the configurations of large regions with smooth boundaries. Unfortunately, the image in Fig. 3a contains many small simulated floes and the corresponding boundaries are irregular and have many sharp corners. Therefore, the more energetically favorable solution obtained by IRGS does not correspond to a better configuration for the tested image.

Segmentations of images having larger floe sizes are also investigated. Three additional floe sizes (40, 60, and 80 pixels in diameter) are chosen for testing. Again, for each of the three sizes, 20 scenes are simulated, with different levels of noise added. Fig. 4 gives the IRGS segmentation accuracies over the noise variances. As expected, the IRGS have higher accuracies with the growth of the floe sizes. However, such a tendency is not obvious for other methods as those pixel-based methods typically locate local minima and are less sensitive to the correctness of the prior model.

5.3 SAR Sea Ice Images

Two operational SAR sea ice images are tested. The first image is shown in Fig. 5a, which was captured over the Gulf of St. Lawrence on 20 February 1998. In the image, the center of the bottom is water and land. Those regions that are relatively dark and have brighter lines (ice ridges caused by pressure) inside belong to an ice type called "gray-white" ice and the rest are "gray ice." The image is difficult to segment because of the heavy noise and the large intraclass variations of the gray-white ice. A manual segmentation is provided in Fig. 5b. Figs. 5c, 5d, 5e, and 5f show the results by the various methods. The highest percentage (92.2 percent) of correctly labeled sites is produced by the IRGS method. Although this number may not be convincing, since the manual segmentation is inaccurate due to the ambiguity of the ice types in the image, a subjective evaluation also indicates that IRGS is superior than others in producing large homogeneous regions consistent with human interpretation.

Fig. 6a shows another SAR sea ice image captured over Baffin Bay on June 24, 1998. This image is associated with a two-class segmentation (ice versus water) problem, with the water being dark and the ice being bright. It should be noted that there is an area of rough water toward the top right part of the image that appears a little brighter than the rest of the water and, hence, the image is not stationary with respect to the statistics of water.

The C-MLL result in Fig. 6c is accurate for a majority part of the image. However, due to the stationary model assumed, many sites belonging to rough water are incorrectly grouped into the other region (see the center right part of the image). Such a phenomenon also occurs in the V-MLL result in Fig. 6d. The GIEP and IRGS have incorporated the edge information that is helpful in describing local behaviors and, hence, improves significantly on such nonstationary situations, as shown in Figs. 6e and 6f. In the results, the rough water region is much cleaner, with very few sites incorrectly labeled.

6 ALGORITHMIC SPEED

The computational complexity of the region-based CMLL, VMLL, and GIEP are all $O(mnk)$, where k is the number of the nodes of the constructed RAG, m is the number of iterations, and n is the number of classes. Therefore, the algorithmic speed depends on the complexity of the image contents as it determines the number of RAG nodes. Analyzing the computational complexity of IRGS is difficult. Although the number of nodes keeps decreasing with increased iterations and, hence, reducing the time for the labeling process, the additional computation of the merging process and the updating of the edge penalties are not ignorable. Depending on the complexity of the image, the IRGS can be much faster than other methods if the number of iterations is large and most merging occurs early (that is, the RAG nodes decrease significantly by the first several runs of iterations). Table 5 gives a summary of the runtime for the various algorithms on those images in the previous experiments.
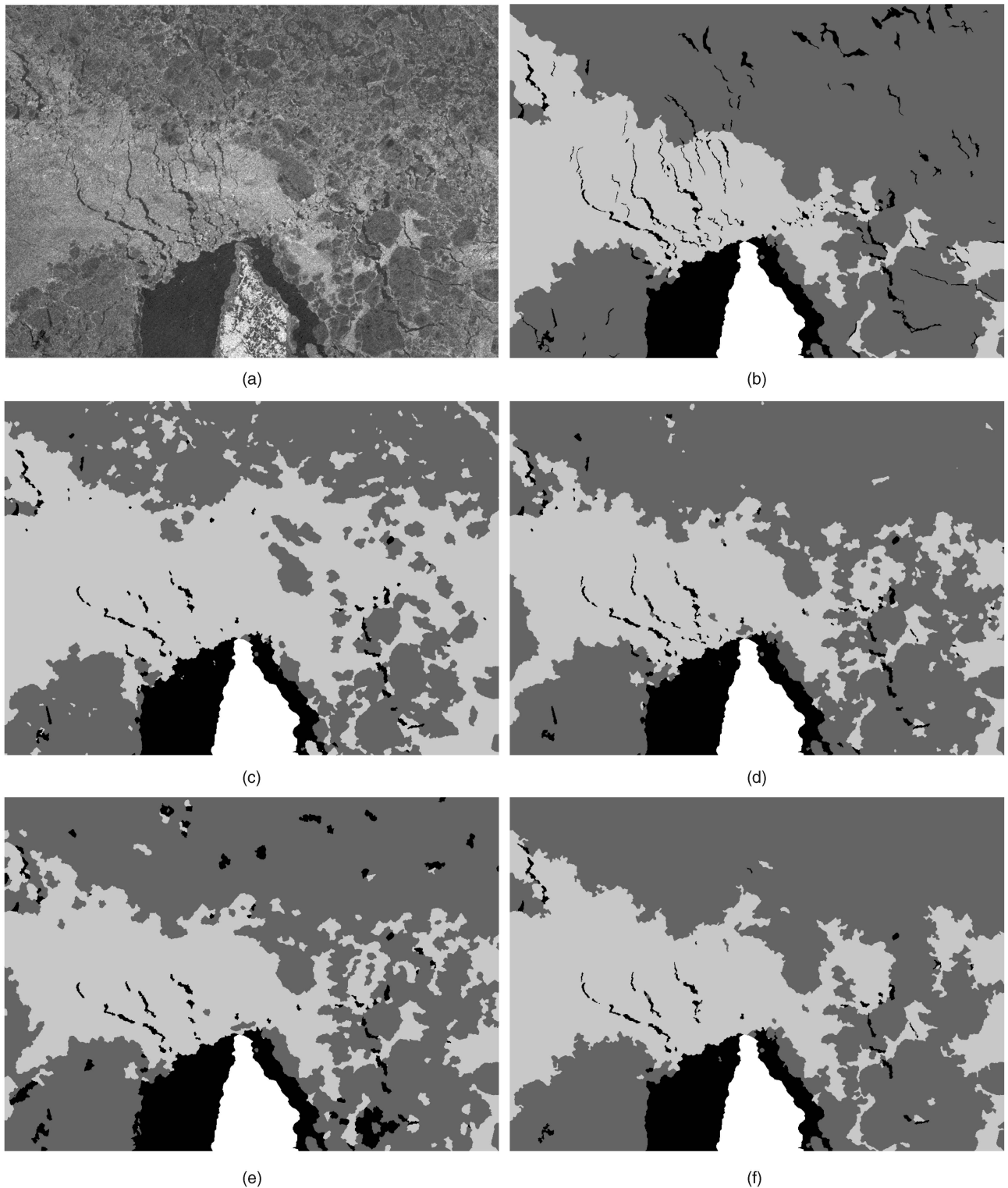
Fig. 5. Segmentation of a SAR image (1,209 × 865) captured over the Gulf of St. Lawrence on 20 February 1998. $\hat{\beta} = 1.68$. (a) Original. (b) Manual segmentation. (c) By C-MLL. Energy: 4.7475×10^6; Accuracy: 76.7 percent; Kappa: 0.60. (d) By V-MLL. Energy: 4.7318×10^6; Accuracy: 88.9 percent; Kappa: 0.789. (e) By GIEP. Energy: 4.7439×10^6; Accuracy: 87.9 percent; Kappa: 0.772. (f) By IRGS. Energy: 4.7222×10^6; Accuracy: 92.2 percent; Kappa: 0.849.

7   SUMMARY

This paper has presented an image segmentation method that is based on the classical MRF formulation of the segmentation problem. The strength of the proposed IRGS method lies in three aspects. First, it incorporates the edge information in a very simple manner without changing the ultimate formulation of the problem. Second, it provides a transition from the standard Gaussian mixture problem to the more difficult MRF one, allowing a more elegant and accurate estimation of the feature model parameters. Third, it uses a region growing technique to generate a hierarchical organization of the pixels
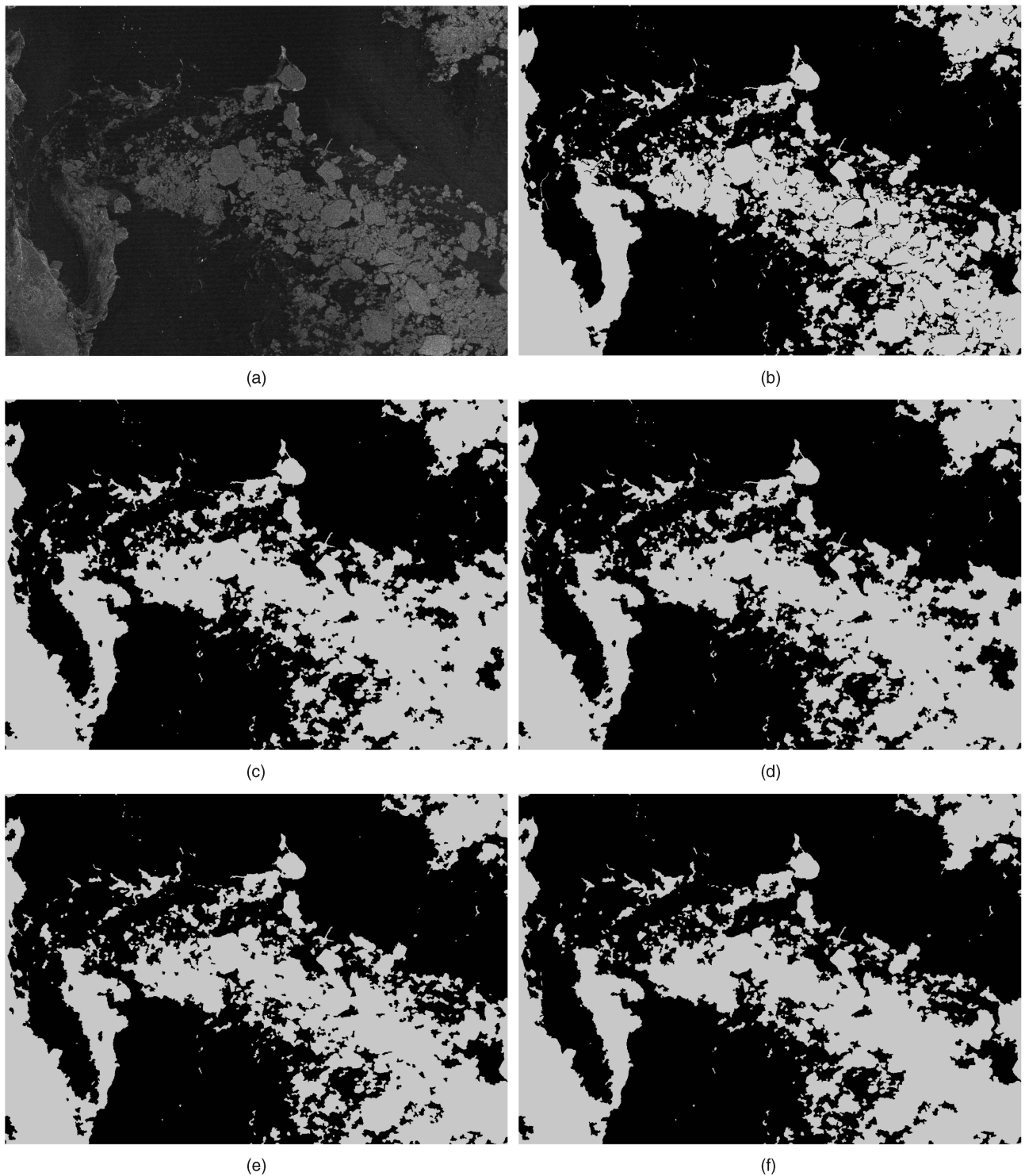
Fig. 6. Segmentation of a SAR image ($1,252 \times 873$) captured over the Baffin Bay on 24 June 1998. $\hat{\beta} = 1.35$. (a) Original. (b) Manual segmentation. (c) By C-MLL. Energy: 4.3067×10^6; Accuracy: 94.2 percent; Kappa: 0.875. (d) By V-MLL. Energy: 4.3008×10^6; Accuracy: 95.3 percent; Kappa: 0.898. (e) By GIEP. Energy: 4.2949×10^6; Accuracy: 97.3 percent; Kappa: 0.941. (f) By IRGS. Energy: 4.3050×10^6; Accuracy: 97.1 percent; Kappa: 0.936.

on which the optimization is more efficient. The proposed method is superior to the compared methods in obtaining visually better segmentations and, for most of the tested images, finds energetic better local minima.

Based on the hierarchical structure established during the process, an application specific interpretation is possible and is also allowed to influence the region growing. This can be easily achieved by replacing the region-based labeling step in the algorithm with the interpretation process [46]. Such a bidirectional interaction between the segmentation and interpretation is also a nice property of the proposed method, but has not been discussed in this paper as this paper deals with general image segmentations only.

TABLE 5
Image Content Complexities and Average Running Time
(in seconds (s)) of the Algorithms (in C++) on PC
(P4 1.8G, 1 Gbyte of RAM)

	Fig. 2(b)	Fig. 3(b)	Fig. 5(a)	Fig. 6(a)
Image size	378×279	512×512	1209×865	1252×873
Number of RAG nodes	11113	27701	86405	89257
C-MLL (s)	11.3	28.7	89.7	93.3
V-MLL (s)	12	29.3	92	97.7
GIEP (s)	19.3	47	142	146.3
IRGS (s)	12.7	37	98.7	96.3

All methods have 300 iterations.

The proposed IRGS method is deterministic and the merging step is not reversible. Fast speed is achieved at the expense of accuracy. More accurate results may be obtained with the incorporation of a region splitting process and adaptive updating of the hierarchical structure. This aspect is part of future work.

ACKNOWLEDGMENTS

REFERENCES

[1] N. Otsu, "A Threshold Selection Method from Gray-Level Histograms," *IEEE Trans. Systems, Man, and Cybernetics,* vol. 9, pp. 62-66, 1979.

[2] A.K. Jain and R.C. Dubes, *Algorithms for Clustering Data.* Prentice Hall, 1988.

[3] J.F. Canny, "A Computational Approach to Edge Detection," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 8, no. 6, pp. 679-698, 1986.

[4] W.Y. Ma and B.S. Manjunath, "Edgeflow: A Technique for Boundary Detection and Image Segmentation," *IEEE Trans. Image Processing,* vol. 9, no. 8, 2000.

[5] H.T. Nguyen, M. Worring, and R. van den Boomgaard, "Watersnakes: Energy-Driven Watershed Segmentation," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 25, no. 3, Mar. 2003.

[6] L. Vincent and P. Soille, "Watershed in Digital Spaces: An Efficient Algorithm Based on Immersion Simulations," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 13, no. 6, pp. 583-598, June 1991.

[7] R. Adams and L. Bischof, "Seeded Region Growing," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 16, no. 6, June 1994.

[8] K. Haris, S.N. Efstratiadis, N. Maglaveras, and A.K. Katsaggelos, "Hybrid Image Segmentation Using Watersheds and Fast Region Merging," *IEEE Trans. Image Processing,* vol. 7, no. 12, pp. 1684-1699, 1998.

[9] T.F. Cootes, C.J. Taylor, D.H. Cooper, and J. Graham, "Active Shape Models—Their Training and Application," *Computer Vision and Image Understanding,* vol. 61, no. 1, pp. 38-59, 1995.

[10] M. Kass, A. Witkin, and D. Terzopoulos, "Snakes: Active Contour Models," *Int'l J. Computer Vision,* vol. 1, no. 4, pp. 321-331, 1987.

[11] J.A. Yezzi, A. Tsai, and A. Willsky, "A Fully Global Approach to Image Segmentation via Coupled Curve Evolution Equations," *J. Visual Comm. and Image Representation,* vol. 13, pp. 195-216, 2002.

[12] J. Besag, "On the Statistical Analysis of Dirty Pictures," *J. Royal Statistical Soc. B,* vol. 48, no. 3, pp. 259-302, 1986.

[13] H. Derin and H. Elliott, "Modeling and Segmentation of Noisy and Textured Images Using Gibbs Random Fields," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 9, no. 1, pp. 39-55, Jan. 1987.

[14] S. Geman and D. Geman, "Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 6, no. 6, pp. 721-741, 1984.

[15] X. He, R.S. Zemel, and M.A. Carreira-Perpinan, "Multiscale Conditional Random Fields for Image Labeling," *Proc. Computer Vision and Pattern Recognition,* vol. 2, pp. 695-702, June-July 2004.

[16] S. Kumar and M. Hebert, "Discriminative Random Fields," *Int'l J. Computer Vision,* vol. 68, no. 2, pp. 179-202, 2006.

[17] S.Z. Li, *Markov Random Field Modeling in Image Analysis.* Springer, 2001.

[18] C.S. Won and H. Derin, "Unsupervised Segmentation of Noisy and Textured Images Using Markov Random Fields," *CVGIP: Graphical Models and Image Processing,* vol. 54, no. 4, pp. 308-328, 1992.

[19] Q. Yu and D.A. Clausi, "Combining Local and Global Features for Image Segmentation Using Iterative Classification and Region Merging," *Proc. Second Canadian Conf. Computer and Robot Vision,* May 2005.

[20] S.E. Hernandez and K.E. Barner, "Joint Region Merging Criteria for Watershed-Based Image Segmentation," *Proc. Int'l Conf. Image Processing,* vol. 2, pp. 108-111, Sept. 2000.

[21] R. Huang, V. Pavlovic, and D.N. Metaxas, "A Graphical Model Framework for Coupling MRFs and Deformable Models," *Proc. Computer Vision and Pattern Recognition,* vol. 2, pp. 739-746, June-July 2004.

[22] B. McCane and T. Caelli, "Multi-Scale Adaptive Segmentation Using Edge and Region Based Attributes," *Proc. First Int'l Conf. Knowledge-Based Intelligent Electronic Systems,* vol. 1, pp. 72-80, May 1997.

[23] H. Deng and D.A. Clausi, "Unsupervised Image Segmentation Using a Simple MRF Model with a New Implementation Scheme," *Pattern Recognition,* vol. 37, no. 12, pp. 2323-2335, 2004.

[24] Z. Tu and S.C. Zhu, "Image Segmentation by Data-Driven Markov Chain Monte Carlo," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 24, no. 5, pp. 657-673, May 2002.

[25] P. Andrey and P. Tarroux, "Unsupervised Segmentation of Markov Random Field Modeled Textured Images Using Selectionist Relaxation," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 20, no. 3, pp. 252-262, Mar. 1998.

[26] L. Cheng and T. Caelli, "Unsupervised Image Segmentation: A Bayesian Approach," *Proc. 16th Int'l Conf. Vision Interface,* June 2003.

[27] J. Zhang, "The Mean Field Theory in EM Procedures for Markov Random Fields," *IEEE Trans. Signal Processing,* vol. 40, no. 10, pp. 2570-2583, 1992.

[28] C. Bouman and M. Shapiro, "A Multiscale Random Field Model for Bayesian Image Segmentation," *IEEE Trans. Image Processing,* vol. 3, no. 2, 1994.

[29] Z. Kato, M. Berthod, and J. Zerubia, "A Hierarchical Markov Random Field Model and Multitemperature Annealing for Parallel Image Classification," *Graphic Models and Image Processing,* vol. 58, no. 1, pp. 18-37, 1996.

[30] J. Laferte, P. Perez, and F. Heitz, "Discrete Markov Image Modeling and Inference on the Quadtree," *IEEE Trans. Image Processing,* vol. 9, no. 3, 2000.

[31] R. Wilson and C. Li, "A Class of Discrete Multiresolution Random Fields and Its Application to Image Segmentation," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 25, no. 1, pp. 42-56, Jan. 2002.

[32] A. Barbu and S.C. Zhu, "Generalizing Swendsen-Wang to Sampling Arbitrary Posterior Probabilities," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 27, no. 8, pp. 1239-1253, Aug. 2005.

[33] Y. Boykov, O. Veksler, and R. Zabih, "Fast Approximate Energy Minimization via Graph Cuts," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 23, no. 11, pp. 1222-1239, Nov. 2001.

[34] S. Wesolkowski and P. Fieguth, "Hierarchical Regions for Image Segmentation," *Proc. Int'l Conf. Image Analysis and Recognition,* 2004.

[35] A. Sarkar, M.K. Biswas, and K.M.S. Sharma, "A Simple Unsupervised MRF Model Based Image Segmentation Approach," *IEEE Trans. Image Processing,* vol. 9, no. 5, pp. 801-812, 2000.

[36] S.C. Zhu and A. Yuille, "Region Competition: Unifying Snakes, Region Growing, and Bayes/MDL for Multiband Image Segmentation," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 18, no. 9, pp. 884-900, Sept. 1996.

[37] N. Metropolis, A.W. Rosenbluth, M.N. Rosenbluth, A.H. Teller, and E. Teller, "Equations of State Calculations by Fast Computing Machines," *J. Chemical Physics,* vol. 21, pp. 1087-1092, 1953.

[38] A.P. Dempster, N.M. Laird, and D.B. Rubin, "Maximum Likelihood from Incomplete Data via the EM Algorithm," *J. Royal Statistical Soc. B,* vol. 39, no. 1, pp. 1-38, 1977.

[39] E. Rignot and R. Chellappa, "Segmentation of Polarimetric Synthetic Aperture Radar Data," *IEEE Trans. Image Processing,* vol. 1, no. 3, pp. 281-300, 1992.

[40] B. Yue, "SAR Sea Ice Recognition Using Texture Methods," master's thesis, Dept. of System Design Eng., Univ. of Waterloo, 2001.

[41] X. Descombes, R.D. Morris, J. Zerubia, and M. Berthod, "Estimation of Markov Random Field Prior Parameters Using Markov Chain Monte Carlo Maximum Likelihood," *IEEE Trans. Image Processing,* vol. 8, no. 7, pp. 954-963, 1999.

[42] M. Hansen and B. Yu, "Model Selection and Minimum Description Length Principle," *J. Am. Statistic Assoc.,* vol. 96, pp. 746-774, 2001.

[43] Y. Boykov, O. Veksler, and R. Zabih, "Markov Random Fields with Efficient Approximations," *Computer Vision and Pattern Recognition,* pp. 648-655, 1998.

[44] J.A. Feldman and Y. Yakimovsky, "Decision Theory and Artificial Intelligence I: Semantics-Based Region Analyzer," *Artificial Intelligence,* vol. 5, no. 4, pp. 349-371, 1974.

[45] M. Sonka, V. Hlavac, and R. Boyle, *Image Processing, Analysis, and Machine Vision.* Thomson Course Technology, 1998.

[46] Q. Yu and D.A. Clausi, "SAR Sea-Ice Image Analysis Based on Iterative Region Growing Using Semantics," *IEEE Trans. Geoscience and Remote Sensing,* vol. 45, no. 12, 2007.

[47] Z. Kato, T.C. Pong, and J.C.M. Lee, "Color Image Segmentation and Parameter Estimation in a Markovian Framework," *Pattern Recognition Letters,* vol. 22, no. 3-4, pp. 309-321, 2001.

[48] S.K. Alexander, "Image Sampling by Hierarchical Annealing," *Proc. Int'l Conf. Image Processing,* pp. 249-252, Sept. 2003.

[49] R. Samadani, "A Finite Mixture Algorithm for Finding Proportions in SAR Images," *IEEE Trans. Image Processing,* vol. 4, no. 8, pp. 1182-1186, 1995.

Qiyao Yu received the BASc degree from Tsinghua University, China, in 1997, the MASc degree from Memorial University of Newfoundland, Canada, in 2002, and the PhD degree in systems design engineering from the University of Waterloo, Canada, in 2006. His research interests include image and video processing, pattern recognition, and remote sensing.

David A. Clausi received the BASc, MASc, and PhD degrees in systems design engineering from the University of Waterloo, Canada, in 1990, 1992, and 1996, respectively. After completing his PhD, he worked in the medical imaging field at Mitra Imaging Inc., Waterloo. He started his academic career as an assistant professor in geomatics engineering at the University of Calgary, Canada, in 1997. He returned to his alma mater in 1999 and was awarded tenure and promotion to associate professor in 2003. He is an active interdisciplinary and multidisciplinary researcher. He has an extensive publication record, publishing refereed journal and conference papers on remote sensing, computer vision, algorithm design, and biomechanics. His primary research interest is the automated interpretation of synthetic aperture radar (SAR) sea ice imagery, in support of operational activities of the Canadian Ice Service. The research results have led to successfully commercial implementations. He has received numerous scholarships, conference paper awards, and two Teaching Excellence Awards. He is a senior member of the IEEE.

▷ **For more information on this or any other computing topic, please visit our Digital Library at** www.computer.org/publications/dlib.