Gaussian-Mixture-Model-Based Spatial Neighborhood Relationships for Pixel Labeling Problem

Thanh Minh Nguyen and Q. M. Jonathan Wu, *Senior Member, IEEE*

***Abstract*—In this paper, we present a new algorithm for pixel labeling and image segmentation based on the standard Gaussian mixture model (GMM). Unlike the standard GMM where pixels themselves are considered independent of each other and the spatial relationship between neighboring pixels is not taken into account, the proposed method incorporates this spatial relationship into the standard GMM. Moreover, the proposed model requires fewer parameters compared with the models based on Markov random fields. In order to estimate model parameters from observations, instead of utilizing an expectation–maximization algorithm, we employ gradient method to minimize a higher bound on the data negative log-likelihood. The performance of the proposed model is compared with methods based on both standard GMM and Markov random fields, demonstrating the robustness, accuracy, and effectiveness of our method.**

***Index Terms*—Gaussian mixture models (GMMs), image segmentation, pixel labeling, spatial neighborhood relationships.**

I. Introduction

SEGMENTATION of images has found widespread applications in image processing and image recognition systems. A correct segmentation result provides more information for diagnosis. However, images corrupted with high levels of noise may result in inaccurate image segmentation. Up until now, many algorithms have been developed for image segmentation such as clustering methods [1], histogram-based methods [2], artificial neural networks [3], and multiscale segmentation methods [4]. Over the last two decades, there has been a growing research interest in model-based techniques [5]–[7]. In general, the algorithms based on model-based techniques can be divided into two groups.

The first group of model-based techniques consists of the standard Gaussian mixture model (GMM) [8]–[10]. Many researchers have used it to study a number of important problems in image segmentation [11], [12]. In general, this approach can be described as follows. The observation at the ith pixel of an image is denoted by Ω_j. Each pixel is considered a random variable whose possibility density function $p(x_i|\Omega_j)$ is a Gaussian function. The model assumes a common prior distribution π_j, which independently generates the pixel labels. In addition, this prior distribution has no dependence on the pixel index i. In order to estimate the model parameters, the expectation–maximization (EM) algorithm [13], [14] is used to maximize the log-likelihood of the given data set. It can be easily seen that the spatial relationship between neighboring pixels is not taken into account of the standard GMM [16]. For this reason, although the standard GMM is a well-known and simple method for image segmentation, its segmentation result is sensitive to noise, varying illumination and other environmental factors such as wind, rain, or camera shaking.

The second group of techniques based on the Markov random fields [17]–[22] has received great attention for modeling and processing image data, as it aims to reduce the sensitivity of the segmentation result with respect to noise. The major difference is that instead of using the common prior distribution π_j for all pixels, the prior distribution π_{ij} of the second group is different for each pixel and depends on the neighbors of the pixel of interest and the corresponding parameters [23]. Thus, the spatial relationship between neighboring pixels is taken into account. Due to the introduction of the prior distribution π_{ij}, the M step of the EM algorithm used in standard GMM cannot be applied to estimate the model parameters from the observations. For this reason, the Gradient Projection algorithm [22] was proposed to implement the M step. Another method [20] based on a closed-form update equation is used to implement the M step and estimate the parameters. Although this approach works well in noisy image segmentation, it is too complex and requires a great number of parameters [22] compared with the standard GMM.

Based on these considerations, in this paper, we introduce a new model that incorporates the spatial relationship between neighboring pixels based on the standard GMM. Our approach differs from those discussed previously by the following statements. First, since the proposed model is based on standard GMM, it is simple and can be easily implemented, as compared with the models based on Markov random fields. Second, compared with the standard GMM, the main difference in the proposed method is that the prior distribution is different for each pixel and depends on the neighbors of the pixel. Third, since the neighboring pixels in an image are similar in some

Manuscript received November 9, 2009; revised January 24, 2011 and June 3, 2011; accepted June 19, 2011. Date of publication August 15, 2011; date of current version December 7, 2011. This work was supported in part by the Canada Chair Research Program and in part by the Natural Sciences and Engineering Research Council of Canada. This paper was recommended by Associate Editor L. Wang.

The authors are with the Department of Electrical and Computer Engineering, University of Windsor, Windsor, ON N9B-3P4, Canada (e-mail: nguyen1j@uwindsor.ca; jwu@uwindsor.ca).

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/TSMCB.2011.2161284

sense, by taking the information of the local spatial interactions between neighboring pixels, the proposed system works well for noisy image. Moreover, the proposed model requires fewer parameters compared with the model based on Markov random fields. Finally, to estimate the unknown parameters of the proposed model, instead of using EM algorithm, we use the gradient method to minimize a higher bound on the data negative log-likelihood. The proposed method is applied to segment synthetic and real grayscale images. The performance of the proposed model is compared with methods based on both standard GMM and Markov random fields, demonstrating the robustness, accuracy, and effectiveness of our method.

The remainder of this paper is organized as follows: Section II will describe the related problems. The proposed approach is presented in Section III. Learning algorithms for our method are presented in Section IV. In Section V, we show the experimental results obtained by applying the proposed algorithm to various synthetic and real grayscale images. Conclusions are given in Section VI.

II. Model-Based Techniques for Image Segmentation

In this section, a discussion of the two groups of algorithms based on model-based techniques is presented. The first one is a standard mixture model in which each pixel is considered to be independent of its neighbors and, therefore, is not explicitly incorporated into the model. The second one considers the interactions and dependences between neighboring pixels into account.

We first introduce common notations used throughout this paper. We are dealing with grayscale images consisting of N pixels and K classes in the image denoted by $(\Omega_1, \Omega_2, \ldots, \Omega_K)$. Given a data set of observations $\mathbf{X} = (x_1, x_2, \ldots, x_N)$, we wish to model the data using a mixture of Gaussians. For each pixel x_i, given the posterior probability $p(\Omega_j|x_i)$ for all classes, a determination is made if the given pixel belongs to the class with the largest posterior probability.

A. Standard GMM

According to the standard GMM [8], [32], the prior distribution π_j of the pixel x_i belonging to the class Ω_j should satisfy the constraints $0 \leq \pi_j \leq 1$ and $\sum_{j=1}^{K} \pi_j = 1$. It assumes that the density function at an observation x_i is given by

$$p(x_i) = \sum_{j=1}^{K} \pi_j p(x_i|\Omega_j). \tag{1}$$

The Gaussian distribution $p(x_i|\Omega_j)$ is given by

$$p(x_i|\Omega_j) = \frac{1}{\sqrt{2\pi\sigma_j^2}} \exp\left(-\frac{(x_i - \mu_j)^2}{2\sigma_j^2}\right) \tag{2}$$

where μ_j and σ_j is the mean and the covariance of the Gaussian distribution $p(x_i|\Omega_j)$, respectively. From (1), the log of the likelihood function [24] is given by

$$L(\Theta) = \sum_{i=1}^{N} \log\left(\sum_{j=1}^{K} \pi_j p(x_i|\Omega_j)\right). \tag{3}$$

In order to maximize the likelihood function given in (3), we need to determine the parameters $\Theta = (\mu_j, \sigma_j, \pi_j)$ of the GMM. Various techniques have been previously developed to determine these parameters, based on maximizing their likelihood for a given data set in (3). In this context, a review of these techniques is provided in [10], [24], [25], and [33]. After optimizing the parameters of the GMM, we have to assign labels to the pixels

$$x_i \in \Omega_j : \text{IF } p(\Omega_j|x_i) \geq p(\Omega_c|x_i), \qquad j, c = 1, 2, \ldots, K \tag{4}$$

where the posterior probability $p(\Omega_j|x_i)$ can be expressed using Bayes' theorem of the form

$$p(\Omega_j|x_i) = \frac{\pi_j p(x_i|\Omega_j)}{\sum_{k=1}^{K} \pi_k p(x_i|\Omega_k)}. \tag{5}$$

It is well known that the neighboring pixels within an image are somewhat similar. One of the major disadvantages of the standard GMM is that the pixel observations x_i, as shown in (1)–(3), are considered to be independent samples, and thus, the information of the neighboring pixels is not taken into account. Another limitation of standard GMM [22] is that for each class Ω_j, the prior distribution π_j has the same weight values for each pixel in the image.

B. GMM Based on Markov Random Fields

To overcome these problems and reduce the sensitivity of the segmentation result to noise, several researchers [15]–[22], [26]–[28] have suggested modifications to incorporate the local spatial interactions between neighboring pixels. In [22], the authors proposed a Spatially Variant Finite Mixture Model (SVFMM) for pixel labeling. According to the SVFMM model, the prior distribution π_{ij} of the pixel x_i belonging to the class Ω_j should satisfy the following constraints: $0 \leq \pi_{ij} \leq 1$ and $\sum_{j=1}^{K} \pi_{ij} = 1$. The SVFMM model assumes that the density function at an observation x_i is given by

$$p(x_i) = \sum_{j=1}^{K} \pi_{ij} p(x_i|\Omega_j). \tag{6}$$

The Gaussian distribution $p(x_i|\Omega_j)$ is the same as (2). The main difference between the two groups of model-based techniques given by (1) and (6) is the prior distribution π_{ij}. Different from the first group, the prior distribution in the second group has different weight values for each pixel, for each class Ω_j in the image. The SVFMM method in [20] and [22] introduces a prior distribution π_{ij} for the parameter set $\Pi = \{\pi_{ij}\}$, where $i = 1, 2, \ldots, N$ and $j = 1, 2, \ldots, K$, that takes

into account the spatial information based on the following Gibbs function:

$$p(\Pi)=\frac{1}{Z}\exp\left(-U(\Pi)\right), \qquad U(\Pi)=\beta\sum_{i=1}^{N}\sum_{m\in\mathcal{N}_i} g(u_{im}) \quad (7)$$

where $g(u_{im})$ is a nonnegative and monotonically increasing function [15]. Z is a normalizing constant. β is a parameter, and $u_{im}=\sum_{j=1}^{K}(\pi_{ik}-\pi_{mk})^2$ specifies the distance between the neighboring pixels $\mathcal{N}_i$ of prior distribution.

From (6), the log of the likelihood function is given by

$$L(\Theta)=\sum_{i=1}^{N}\log\left(\sum_{j=1}^{K}\pi_{ij}p(x_i|\Omega_j)\right)+\log p(\Pi). \quad (8)$$

The EM algorithm can be utilized to maximize the log of the likelihood function in (8) with respect to the parameters $\Theta=(\mu_j,\sigma_j,\pi_{ij})$. However, the estimation of prior distribution π_{ij} in the M step requires solving a constrained optimization problem with the following constraints: $0\le\pi_{ij}\le 1$ and $\sum_{j=1}^{K}\pi_{ij}=1$. Therefore, in order to maximize the log of the likelihood function in (8), the algorithm becomes much more computationally complex. In [22], the Gradient Projection method is utilized to maximize this likelihood function. Another method based on a closed-form update equation [20] has been proposed to improve the results. More details can be found in [20]–[22]. After the optimization of the parameters, the posterior probability $p(\Omega_j|x_i)$ in (9) is used to assign labels to each pixel

$$p(\Omega_j|x_i)=\frac{\pi_{ij}p(x_i|\Omega_j)}{\sum_{k=1}^{K}\pi_{ik}p(x_i|\Omega_k)}. \quad (9)$$

Although this approach works well in noisy image segmentation problems, the evaluation of this approach is generally too complex. Moreover, if we want to segment a grayscale image consisting of N pixels into K classes, we have to deal with $K\times(2+N)$ parameters (K parameters of μ_j, K parameters of σ_j, and $N\times K$ parameters of π_{ij}). This implies that the larger the image, the more the number of parameters that we have to estimate [22]. In order to reduce the computational complexity, we have proposed a new algorithm detailed in the following section.

III. Proposed Algorithm

In this section, we suggest a new model based on the standard GMM that applies to the image classification problem. The main advantages of the proposed system are as follows. The proposed model is different from the standard GMM in that the prior distributions π_{ij} are different for each pixel and depend on the neighboring pixels. Compared with the model based on Markov random fields, the proposed model is easier to implement and requires fewer parameters.

Given a grayscale image of size $U\times V$ pixels, for convention of description, the ith pixel x_i of this image will be presented by $x^{(u,v)}$, where $i=U(u-1)+v$, $u=1,2,\ldots,U$, and $v=1,2,\ldots,V$.

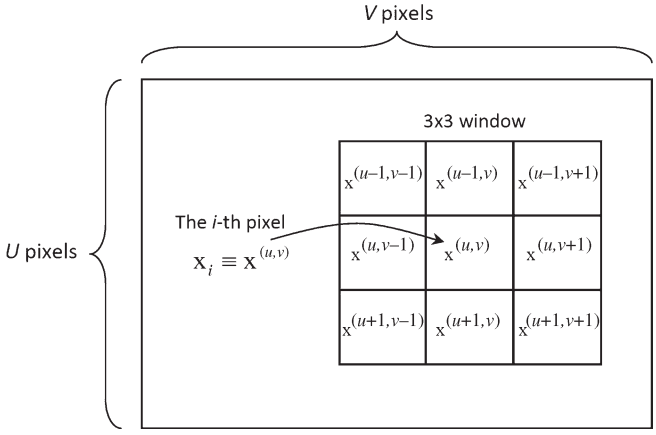

Fig. 1. 3×3 window within the neighborhood of the ith pixel.

To incorporate the local spatial interactions between the neighboring pixels, a 3×3 window, as shown in Fig. 1, is selected. For each ith window, a set of neighborhood pixels is denoted by $\mathbf{W}^{(u,v)}=(x^{(u-1,v-1)},x^{(u-1,v)},x^{(u-1,v+1)},x^{(u,v-1)},x^{(u,v)},x^{(u,v+1)},x^{(u+1,v-1)},x^{(u+1,v)},x^{(u+1,v+1)})$. Let $\overline{x^{(u,v)}}$ denote the mean value of all the pixels in this window

$$\overline{x}_i=\overline{x^{(u,v)}}=\frac{1}{9}\sum_{n=-1}^{1}\sum_{m=-1}^{1}x^{(u+n,v+m)}. \quad (10)$$

We define $\xi_j(\overline{x}_i)$ to represent the weight of the ith window for each class Ω_j

$$\xi_j(\overline{x}_i)=\xi_j\left(\overline{x^{(u,v)}}\right)=\exp\left(-\frac{\left(\overline{x^{(u,v)}}-c_j\right)^2}{2b_j^2}\right) \quad (11)$$

where c_j and b_j, $j=1,2,\ldots,K$, are parameters whose optimal values can be obtained by utilizing the methodology presented in the following section. For the ith window, the mean value $\overline{\xi}_j(\overline{x}_i)$ of the set of neighborhood weights $\xi_j(\overline{x}_i)$ can be expressed in the form

$$\begin{aligned}\overline{\xi}_j(\overline{x}_i)&=\overline{\xi}_j\left(\overline{x^{(u,v)}}\right)\\&=\frac{1}{9}\sum_{n=-1}^{1}\sum_{m=-1}^{1}\exp\left(-\frac{\left(\overline{x^{(u+n,v+m)}}-c_j\right)^2}{2b_j^2}\right).\end{aligned} \quad (12)$$

We propose a new way to include the spatial relationship between neighboring pixels into the prior probability distribution π_{ij}. This prior distribution has different values for each pixel corresponding to each class Ω_j in the image, as given by

$$\pi_{ij}=\frac{\overline{\xi}_j(\overline{x}_i)}{\sum_{k=1}^{K}\overline{\xi}_k(\overline{x}_i)}. \quad (13)$$

The idea to incorporate the spatial constraints in our method is based on a fact that neighboring pixels in an image are similar in some sense. Based on this relationship, we replace each pixel value in an image with the average value of its

neighbors, including itself. The prior probability π_{ij} in (13) plays a role as a filter for smoothing images corrupted by noise. Note that the prior probability π_{ij} is calculated subject to the constraints $0 \leq \pi_{ij} \leq 1$ and $\sum_{j=1}^{K} \pi_{ij} = 1$. Then, the density function at an observation x_i, as shown in (6), is given by $p(x_i) = \sum_{j=1}^{K} \pi_{ij} p(x_i|\Omega_j)$, where the Gaussian distribution $p(x_i|\Omega_j)$ is the same as that given in (2).

We now compare the proposed method with the standard GMM [8], [32]. As we can see, the main difference arises from the prior distribution. In the standard GMM, the scalar value (linear weight) is used for the prior distribution. On the other hand, the exponential weight is used for the prior distributions of the proposed method, as illustrated in (11)–(13). Although the exponential weight is previously used to describe the prior distribution in [25], it does not take into account the spatial constraints in an image and the spatial information between the neighboring pixels does not influence the decision process. The main advantage of using the exponential weights in the proposed method is that the prior distribution is different for each pixel and depends on the neighbors of the pixel and the corresponding parameters.

Given the prior probability distribution π_{ij} in (13), the log of the likelihood function can be represented by

$$L(\Theta) = \sum_{i=1}^{N} \log \left(\sum_{j=1}^{K} \pi_{ij} p(x_i|\Omega_j) \right). \tag{14}$$

Since the logarithm is a monotonically increasing function, it is often more convenient to consider the negative logarithm of the likelihood function [24], [25]

$$E(\Theta) = -L(\Theta) = -\sum_{i=1}^{N} \log \left(\sum_{j=1}^{K} \pi_{ij} p(x_i|\Omega_j) \right). \tag{15}$$

Applying the complete data in [8], [25], minimization of the negative logarithm of the likelihood function in (15) corresponds to minimization of the following function:

$$\mathcal{E}(\Theta^{\text{old}}|\Theta^{\text{new}}) = -\sum_{i=1}^{N} \sum_{j=1}^{K} p^{\text{old}}(\Omega_j|x_i) \log\left(\pi_{ij}^{\text{new}} p^{\text{new}}(x_i|\Omega_j)\right). \tag{16}$$

The $\mathcal{E}$ in (16) can be regarded as an error function. Therefore, maximizing the likelihood L in (14) is then equivalent to minimizing $\mathcal{E}$ in (16). The minimization of the error function $\mathcal{E}$ with respect to the parameters $\Theta = (\mu_j, \sigma_j, c_j, b_j),\ i = 1, 2, \ldots, K$, will be discussed in the next section. After the optimization of the parameters, the posterior probability $p(\Omega_j|x_i)$, as shown in (9), is used to assign labels to each pixels.

IV. Parameter Estimation Algorithm

Thus far, the discussion has focused on probability estimation to determine the class Ω_j to which the pixel x_i should be assigned. To generalize the posterior probability $p(\Omega_j|x_i)$, we need to adjust the parameters $\Theta = (\mu_j, \sigma_j, c_j, b_j),\ i = 1, 2, \ldots, K$, to minimize the error function $\mathcal{E}$ in (16) [or maximize the likelihood function L in (14)]. In this section, instead of utilizing EM algorithm, we employ the gradient method for adjusting these parameters to minimize the error function $\mathcal{E}$. The proposed algorithm can be summarized as follows:

Given a grayscale image consisting of N pixels $x_i,\ i = 1, 2, \ldots, K$, the objective is to segment it into K classes. From (16), we can rewrite the error function as

$$\mathcal{E}(\Theta^{\text{old}}|\Theta^{\text{new}}) = -\sum_{i=1}^{N} \sum_{j=1}^{K} p^{\text{old}}(\Omega_j|x_i) \times \left(\log(\pi_{ij}) - \frac{1}{2}\log(2\pi) - \frac{1}{2}\log(\sigma_j^2) - \frac{(x_i - \mu_j)^2}{2\sigma_j^2} \right). \tag{17}$$

To minimize the error function $\mathcal{E}(\Theta^{\text{old}}|\Theta^{\text{new}})$ with respect to $\Theta = (\mu_j, \sigma_j, c_j, b_j)$, we apply the gradient descent technique [24], [25], [36]–[38]. The process is outlined in the following steps.

Step 1: Initialize the parameters μ_j and σ_j by using the K-mean method [30]. Then, by selecting $c_j = \mu_j$ and $b_j = \sigma_j$, we obtain the initial parameters of Θ^{old}.

Step 2: Evaluate $p^{\text{old}}(\Omega_j|x_i)$, as given by (9):

$$p^{\text{old}}(\Omega_j|x_i) = \frac{\pi_{ij}^{\text{old}} p^{\text{old}}(x_i|\Omega_j)}{\sum_{k=1}^{K} \pi_{ik}^{\text{old}} p^{\text{old}}(x_i|\Omega_k)}$$

where $p^{\text{old}}(x_i|\Omega_j)$ and π_{ij}^{old} are calculated from (2) and (13), respectively.

Step 3: Update parameters $\Theta = (\mu_j, \sigma_j, c_j, b_j)$ to obtain the new parameters Θ^{new}. The parameters are calculated and updated by using the gradient method [24], [25], [29], [36]–[38]

$$\Theta^{\text{new}} = \Theta^{\text{old}} - \eta \nabla \mathcal{E}(\Theta^{\text{old}}) \tag{18}$$

where $\nabla \mathcal{E} = (\partial\mathcal{E}/\partial\mu_j, \partial\mathcal{E}/\partial\sigma_j, \partial\mathcal{E}/\partial c_j, \partial\mathcal{E}/\partial b_j)$ and η is the rate parameter. In this paper, we choose $\eta = 10^{-7}$. The derivative of $\mathcal{E}$ with respect to μ_j is given by

$$\frac{\partial \mathcal{E}}{\partial \mu_j} = -\sum_{i=1}^{N} p^{\text{old}}(\Omega_j|x_i) \frac{x_i - \mu_j}{\sigma_j^2}. \tag{19}$$

Similarly, the derivative of $\mathcal{E}$ with respect to σ_j is given by

$$\frac{\partial \mathcal{E}}{\partial \sigma_j} = -\sum_{i=1}^{N} p^{\text{old}}(\Omega_j|x_i) \left(-\frac{1}{\sigma_j} + \frac{(x_i - \mu_j)^2}{\sigma_j^3} \right). \tag{20}$$

The derivative of $\mathcal{E}$ with respect to c_j can be expressed as

$$\frac{\partial \mathcal{E}}{\partial c_j} = -\sum_{i=1}^{N} \left(p^{\text{old}}(\Omega_j|x_i) \frac{\delta_{ij}}{\upsilon_{ij}} \right) + \sum_{i=1}^{N} \sum_{k=1}^{K} \left(p^{\text{old}}(\Omega_k|x_i) \frac{\delta_{ik}}{\sum_{l=1}^{K} \upsilon_{il}} \right) \tag{21}$$

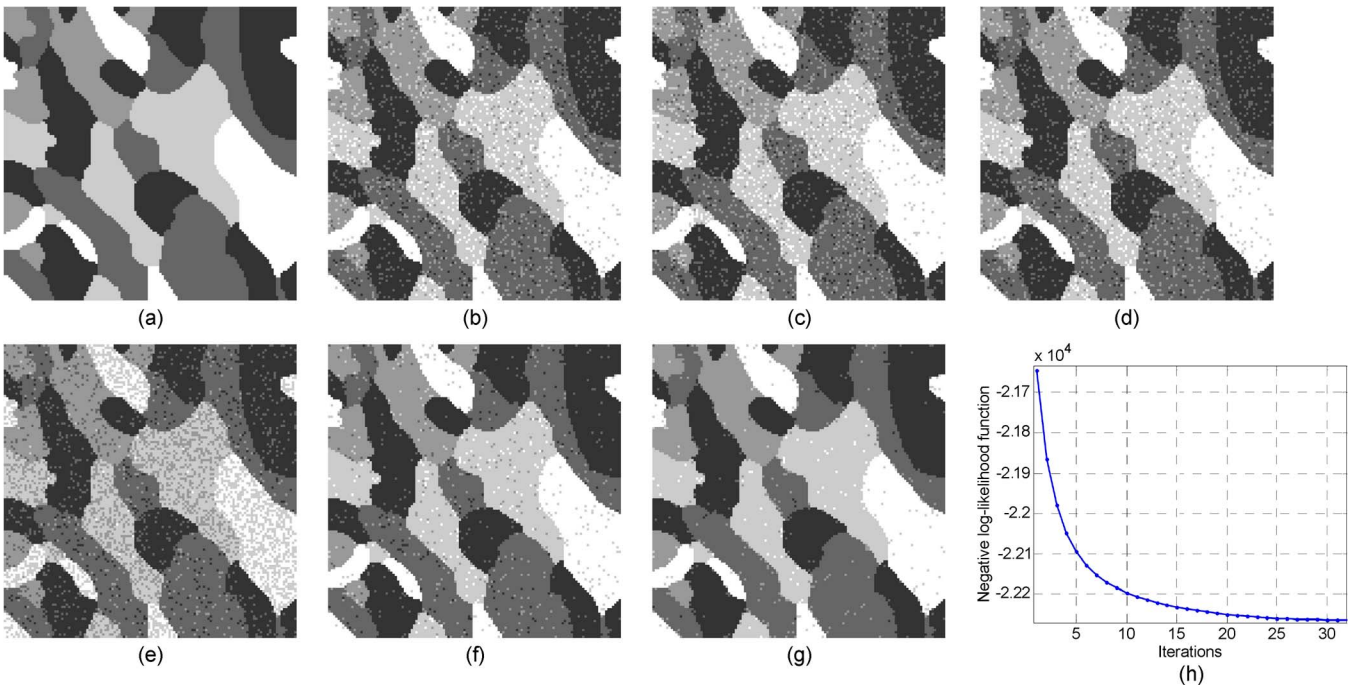

Fig. 2. First experiment (128 × 128 image resolution). (a) Original image, (b) corrupted original image with Gaussian noise (0 mean and 0.005 variance), (c) K-mean (MCR = 10.12%), (d) FCM (MCR = 10.29%), (e) standard GMM (MCR = 14.31%), (f) SVFMM (MCR = 4.13%), (g) the proposed method (MCR = 2.08%), and (h) minimization progress of the negative logarithm of the likelihood function of the proposed algorithm.

where

$$\delta_{ij} = \sum_{n=-1}^{1}\sum_{m=-1}^{1}\left\{\frac{\left(\overline{x^{(u+n,v+m)}}-c_j\right)}{2b_j^2} \times \exp\left(-\frac{\left(\overline{x^{(u+n,v+m)}}-c_j\right)^2}{2b_j^2}\right)\right\} \quad (22)$$

$$\upsilon_{ij} = \sum_{n=-1}^{1}\sum_{m=-1}^{1}\exp\left(-\frac{\left(\overline{x^{(u+n,v+m)}}-c_j\right)^2}{2b_j^2}\right). \quad (23)$$

Now, considering the derivative of the term $\mathcal{E}$ with respect to b_j, we have

$$\frac{\partial \mathcal{E}}{\partial b_j} = -\sum_{i=1}^{N}\left(p^{\text{old}}(\Omega_j|x_i)\frac{\tau_{ij}}{\upsilon_{ij}}\right) + \sum_{i=1}^{N}\sum_{k=1}^{K}\left(p^{\text{old}}(\Omega_k|x_i)\frac{\tau_{ik}}{\sum_{l=1}^{K}\upsilon_{il}}\right) \quad (24)$$

where

$$\tau_{ij} = \sum_{n=-1}^{1}\sum_{m=-1}^{1}\left\{\frac{\left(\overline{x^{(u+n,v+m)}}-c_j\right)^2}{2b_j^3} \times \exp\left(-\frac{\left(\overline{x^{(u+n,v+m)}}-c_j\right)^2}{2b_j^2}\right)\right\}. \quad (25)$$

TABLE I
COMPARISON OF THE PROPOSED METHOD TO OTHER METHODS FOR THE FIRST EXPERIMENT

Methods	Gaussian Noise (0 mean, var)		
	var=0.002	var=0.005	var=0.007
K-mean	1.56%	10.12%	17.89%
FCM	1.57%	10.29%	18.63%
Standard GMM	7.76%	14.31%	16.24%
SVFMM	0.56%	4.13%	6.32%
Proposed method	0.24%	2.08%	4.58%

TABLE II
FINAL PARAMETERS OF THE PROPOSED METHOD FOR THE FIRST EXPERIMENT

Parameters Θ	Class Ω_j; j=1,2,3,4,5				
	j=1	j=2	j=3	j=4	j=5
μ_j	0.9721	0.1929	0.5835	0.7954	0.3897
σ_j	0.0397	0.0707	0.0725	0.0701	0.0681
c_j	0.9641	0.1892	0.5789	0.7882	0.3857
b_j	0.0550	0.0689	0.0618	0.0604	0.0585

Step 4: Check for convergence of either the negative logarithm of the likelihood function or the parameter values. If the convergence criterion is not satisfied, then set $\Theta^{\text{old}} = \Theta^{\text{new}}$ and return to Step 2.

V. EXPERIMENTAL RESULTS

In this section, the performance of the proposed algorithm is compared with those of the K-mean [30], fuzzy c-means (FCM) [31], standard GMM [32], and the model based on Markov random fields—SVFMM method [20]. The experiments were performed using Matlab on Windows platform. All methods were run until convergence. The misclassification ratio (MCR) [16], which is the number of misclassified pixels divided by the total number of pixels, is used to compare the results obtained.

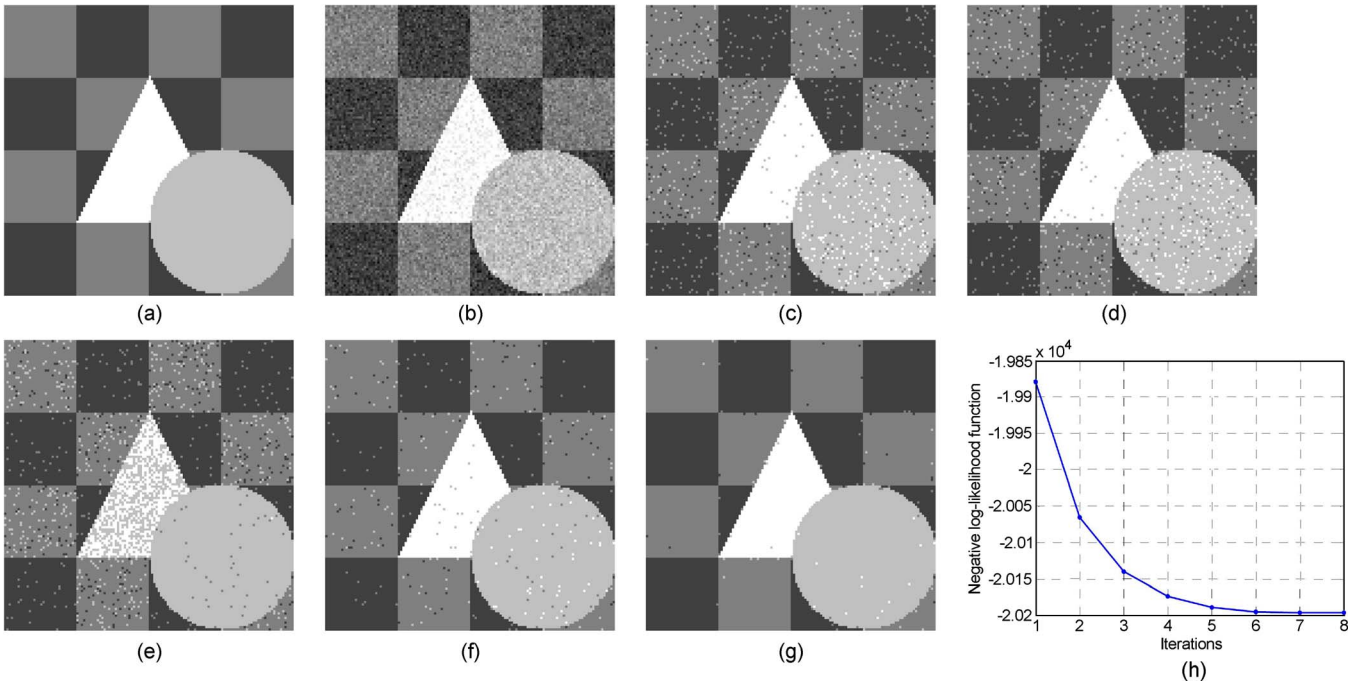

Fig. 3. Second experiment (128 × 128 image resolution). (a) Original image, (b) corrupted original image with Gaussian noise (0 mean and 0.005 variance), (c) K-mean (MCR = 4.48%), (d) FCM (MCR = 4.51%), (e) standard GMM (MCR = 6.71%), (f) SVFMM (MCR = 1.31%), (g) proposed method (MCR = 0.39%), and (h) minimization progress of the negative logarithm of the likelihood function of the proposed algorithm.

We first apply the algorithm to synthetic images and then to real-world images.

A. Synthetic Images

In the first set of experiments, synthetic images were generated and tested. A similar synthetic image (128 × 128 image resolution) to the one used in [16] and [20], shown in Fig. 2(a), was used to test the effectiveness of the algorithm. This image has five classes $(K = 5)$ with luminance values [0.2, 0.4, 0.6, 0.8, 1]. The image shown in Fig. 2(b) is made from the original image by corrupting with Gaussian noise (0 mean and 0.005 variance). The resulting images shown in Fig. 2(c) and (d) are obtained by applying K-mean and FCM to the synthetic image. As can be easily seen, the accuracy of these methods is quite poor with MCRs of 10.12% and 10.29%, respectively. In Fig. 2(e) and (f), we present the segmentation results obtained by employing standard GMM and SVFMM methods, respectively. The result obtained by employing the proposed method is shown in Fig. 2(g), converging after 32 iterations, as shown in Fig. 2(h). It can be seen that the image is classified accurately with a lower MCR of 2.08%. As shown in Table I, the effect of noise on the performance of the proposed detector is much less as compared with those of other methods. Moreover, for the proposed method, we only need to estimate with 20 parameters, as shown in Table II, compared with $K \times (2 + N) = 81930$ parameters for the SVFMM method.

In the second experiment, another synthetic image, as shown in Fig. 3(a), was used. It has four classes $(K = 4)$ with luminance values [0.25, 0.5, 0.75, 1]. This image was corrupted with Gaussian noise (0 mean and 0.005 variance), as shown in Fig. 3(b). Fig. 3(c)–(g) shows the segmentation results obtained by employing K-mean, FCM, standard GMM, and SVFMM methods and the proposed method, respectively. As can be seen from Fig. 3(h), the proposed method converges after eight iterations and can classify the image with the lowest MCR of 0.39% compared with other methods. Table III contains the results of these methods for the given synthetic image for varying amounts of noise. The final parameters obtained after optimization are given in Table IV.

TABLE III
COMPARISON OFTHE PROPOSED METHOD TO OTHER METHODS FOR THE SECOND EXPERIMENT

Methods	Gaussian Noise (0 mean, var)			
	var=0.002	var=0.005	var=0.007	var=0.01
K-mean	0.34%	4.48%	8.58%	15.26%
FCM	0.34%	4.51%	8.73%	15.31%
Standard GMM	0.41%	6.71%	9.20%	14.40%
SVFMM	0.09%	1.31%	2.82%	4.64%
Proposed method	0.03%	0.39%	1.59%	4.13%

TABLE IV
FINAL PARAMETERS OF THE PROPOSED METHOD FOR THE SECOND EXPERIMENT

Parameters Θ	Class Ω_j; j=1,2,3,4			
	j=1	j=2	j=3	j=4
μ_j	0.4960	0.7370	0.2469	0.9620
σ_j	0.0714	0.0728	0.0720	0.0521
c_j	0.4934	0.7295	0.2451	0.9567
b_j	0.0631	0.0669	0.0714	0.0605

To illustrate the computational cost of the proposed algorithm and others, an image with three classes $(K = 3)$, as shown in Fig. 4(a), is used in this simulation. This 128 × 128 pixel image can be easily generated by the computer. Each square box in this image has a size of 32 × 64 pixels, and the 2048 pixels contained within each box have the same luminance value [1/3, 2/3, 1]. The image shown in Fig. 4(b) is made

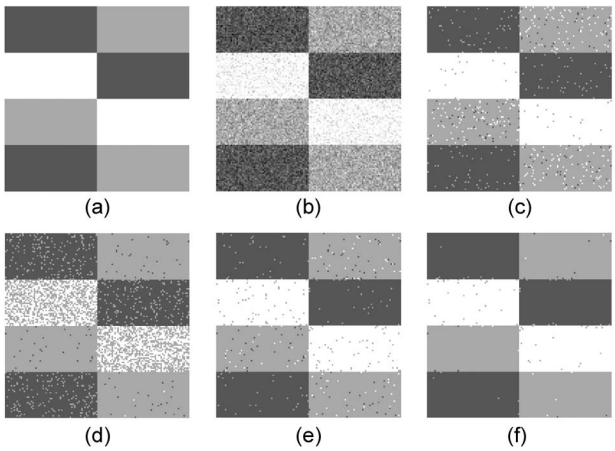

Fig. 4. Third experiment (128 × 128 image resolution). (a) Original image, (b) corrupted original image with Gaussian noise (0 mean and 0.01 variance), (c) FCM (time = 0.1 s, MCR = 4.17%), (d) standard GMM (time = 0.6 s, MCR = 11.38%), (e) SVFMM (time = 49.5 s, MCR = 1.13%), and (f) proposed method (time = 13.4 s, MCR = 0.31%).

TABLE V
COMPUTATIONAL COST (IN SECONDS) OF THE PROPOSED METHOD TO OTHER METHODS FOR THE THIRD EXPERIMENT

Methods	FCM	Standard GMM	SVFMM SVFMM	Proposed method
Time (seconds)	0.1	0.6	49.5	13.4
MCR (%)	4.17	11.38	1.13	0.31

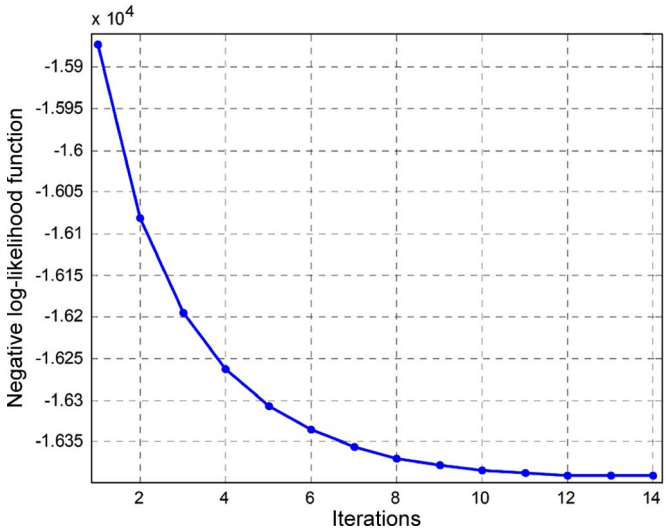

Fig. 5. Minimization progress of the negative logarithm of the likelihood function E in (15) of the proposed algorithm for the third experiment.

from the original image by corrupting with Gaussian noise (0 mean and 0.01 variance). All methods are initialized with the same initial condition and are performed on a PC (Core i3 with 4-GB RAM) until convergence by using Matlab in the Windows environment. As shown in Table V, although the proposed method is still low in terms of the speed, it has the lowest MCR and demonstrates a higher degree of robustness with respect to noise. Fig. 4(c) and (d) shows the segmented images by employing FCM and standard GMM methods, respectively. The result obtained from the SVFMM method in Fig. 4(e) shows a good performance with an MCR of 1.13%. However, compared with these methods, the proposed method in Fig. 4(f)

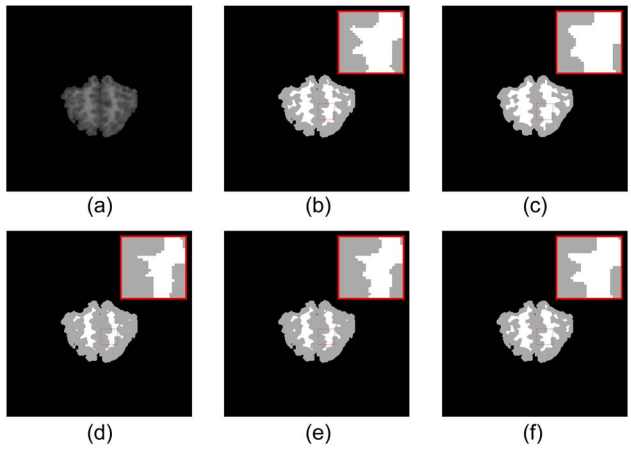

Fig. 6. First real image experiment. (a) Original image, (b) ground-truth segmented image, (c) FCM (MCR = 1.01%), (d) standard GMM (MCR = 1.04%), (e) SVFMM (MCR = 0.82%), and (f) proposed method (MCR = 0.26%).

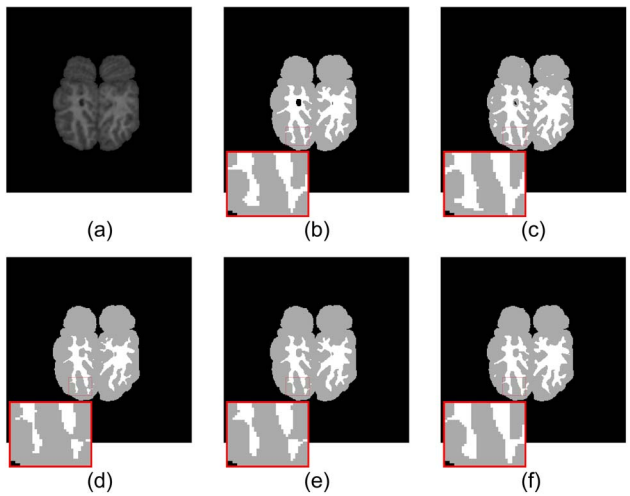

Fig. 7. Second real image experiment. (a) Original image, (b) ground-truth segmented image, (c) FCM (MCR = 1.20%), (d) standard GMM (MCR = 1.01%), (e) SVFMM (MCR = 0.72%), and (f) proposed method (MCR = 0.59%).

achieves a lower MCR of 0.31%. The convergence rate of our proposed method is shown in Fig. 5.

B. Real Images

In the first experiment in this section, a real-world image from the Center for Morphometric Analysis at Massachusetts General Hospital [34], as shown in Fig. 6(a), is used to test the accuracy and determine the efficiency of the proposed algorithm, as compared with other algorithms. The objective is to segment the image into three classes ($K = 3$): gray matter, white matter, and background. In this experiment, a low contrast between the gray matter and the white matter increases the complexity of the image. The ground truth of this image is shown in Fig. 6(b). As can be seen from the result shown in Fig. 6(c) and (d), the segmentation accuracies of FCM and standard GMM methods are quite poor. The SVFMM method in Fig. 6(e) demonstrates a better performance with an MCR of 0.82%. However, many details and fine tissue structures are

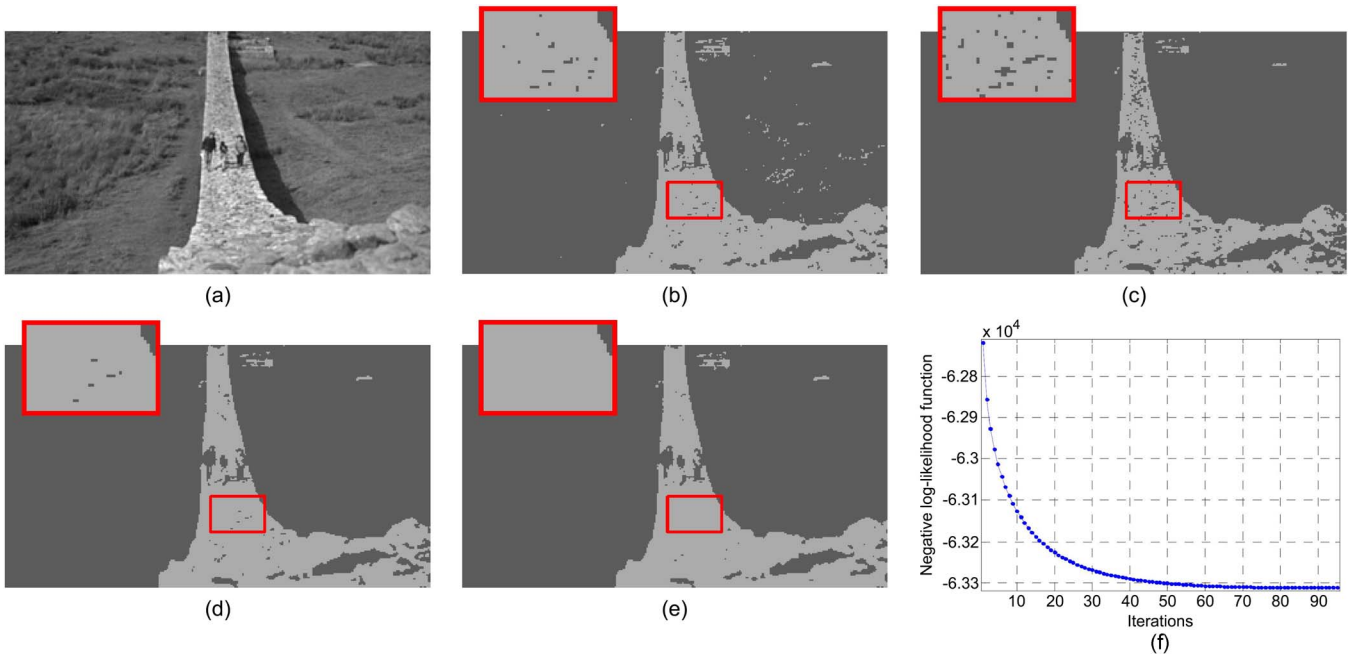

Fig. 8. Third real image experiment. (a) Original image, (b) FCM, (c) standard GMM, (d) SVFMM, (e) proposed method, and (f) the negative logarithm of the likelihood function of the proposed method.

removed (look into the red box). Compared with these methods, the proposed method in Fig. 6(f) reduces the effect of noise on the final segment image lesser.

Another real brain image [34], as shown in Fig. 7(a), is used in this experiment. The image ground truth with three classes is shown in Fig. 7(b). The SVFMM method shown in Fig. 7(e) produces a better result than FCM in Fig. 7(c) and standard GMM in Fig. 7(d). However, the effect of noise on the final segmented image remains high. We can see that many details are lost. Again, compared with other methods, the proposed method in Fig. 7(f) has a lower MCR of 0.59%.

In the next experiment, a real-world grayscale image [35], as shown in Fig. 8(a), was used to compare the proposed algorithm with other algorithms. We want to segment this image into two classes $(K = 2)$: road and background. A visual inspection of the results indicates that the effect of noise on the final segmented image of the FCM method in Fig. 8(b) is highly noticeable. In the standard GMM method in Fig. 8(c), the effect of noise on the final segmented image is still high. As seen from these results, SVFMM demonstrates a good performance compared with FCM and standard GMM. However, several pixels in the red box of Fig. 8(d) are misclassified. In the proposed method, as shown in Fig. 8(e), the effect of noise on the final segmented image is significantly small. The negative logarithm of the likelihood function of the proposed method is shown in Fig. 8(f).

In the final experiment, another grayscale image is used. The image shown in Fig. 9(b) is obtained by corrupting the original image in Fig. 9(a) with Gaussian noise (0 mean and 0.002 variance). The objective is to segment this image into four classes $(K = 4)$: sky, far mountain, near mountain, and stone. Fig. 9(c)–(f) shows the segmentation results obtained by employing FCM, standard GMM, SVFMM, and the proposed method, respectively. As shown from Fig. 9, FCM and standard

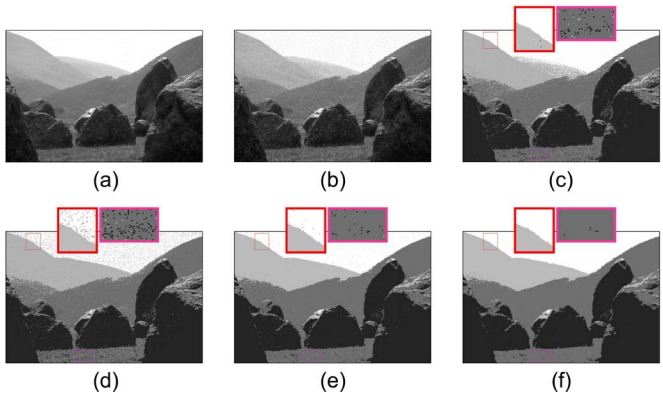

Fig. 9. Final real image experiment. (a) Original image, (b) corrupted original image with Gaussian noise (0 mean and 0.002 variance), (c) FCM, (d) standard GMM, (e) SVFMM, and (f) the proposed method.

GMM methods are unable to segment the image successfully. SVFMM demonstrates a better performance compared with those of the FCM and standard GMM methods. However, if we look closely in the red and pink boxes, we can see that the effect of noise on the final segmented image is still quite high. Again, the proposed method, as shown in Fig. 9(f), obtains a better result.

VI. Conclusion

In this paper, we have presented a new mixture model based on the standard GMM to segment the grayscale images. A new way to incorporate the spatial relationship between neighboring pixels into the GMM has been proposed. In the proposed method, the prior distribution π_{ij} is different for each pixel and depends on the neighbors of the pixel and the corresponding parameters. It is different from the standard GMM where the

spatial relationship between neighboring pixels is not taken into account. Compared with the models based on Markov random fields, which is based on the pixel labels to incorporate the local spatial constrains, our method takes into account the spatial correlation by the pixel intensities in an image. It is based on a well-known fact that the intensities of neighboring pixels in an image are similar in some sense. Moreover, the proposed model requires fewer parameters compared with the models based on Markov random fields. Finally, to estimate the unknown parameters of the proposed model, instead of utilizing EM algorithm, we employ gradient method to minimize a higher bound on the negative log-likelihood.

The proposed method has been tested with both synthetic and real images and has been shown to be robust with respect to noise, efficient with respect to the number of parameters used, and sufficiently accurate with respect to the classification results. Finally, in this paper, the number of classes (K) is manually selected. In the future, we would like to investigate the ways to automatically optimize this parameter.

ACKNOWLEDGMENT

The authors would like to thank the referees and the associate editor for suggesting improvements in the presentation.

REFERENCES

[1] C. H. Chen, L. F. Pau, and P. S. P. Wang, *The Handbook of Pattern Recognition and Computer Vision*, 2nd ed. Singapore: World Scientific Publ. Co., 1998.

[2] G. S. Linda and C. S. George, *Computer Vision*. Englewood Cliffs, NJ: Prentice-Hall, 2001, pp. 279–325.

[3] A. G. Bors and I. Pitas, "Optical flow estimation and moving object segmentation based on median radial basis function network," *IEEE Trans. Image Process.*, vol. 7, no. 5, pp. 693–702, May 1998.

[4] M. Tabb and N. Ahuja, "Unsupervised multiscale image segmentation by integrated edge and region detection," *IEEE Trans. Image Process.*, vol. 6, no. 5, pp. 642–655, May 1997.

[5] J. Besag, "On the statistical analysis of dirty pictures," *J. Roy. Statist. Soc. Ser. B*, vol. 48, no. 3, pp. 259–302, 1986.

[6] J. Liu and Y. H. Yang, "Multiresolution color image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 16, no. 7, pp. 689–700, Jul. 1994.

[7] D. A. Langan, J. W. Modestino, and Z. Jun, "Cluster validation for unsupervised stochastic model-based image segmentation," *IEEE Trans. Image Process.*, vol. 7, no. 2, pp. 180–195, Feb. 1998.

[8] D. M. Titterington, A. F. M. Smith, and U. E. Makov, *Statistical Analysis of Finite Mixture Distributions*. Hoboken, NJ: Wiley, 1985.

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

[10] R. A. Redner and H. F. Walker, "Mixture densities, maximum likelihood and the EM algorithm," *SIAM Rev.*, vol. 26, no. 2, pp. 195–239, Apr. 1984.

[11] H. Permuter, J. Francos, and I. Jermyn, "A study of Gaussian mixture models of color and texture features for image classification and segmentation," *Pattern Recognit.*, vol. 39, no. 4, pp. 695–706, Apr. 2006.

[12] C. Carson, S. Belongie, H. Greenspan, and J. Malik, "Blobworld: Image segmentation using expectation–maximization and its application to image querying," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 8, pp. 1026–1038, Aug. 2002.

[13] G. J. McLachlan and T. Krishnan, "The EM algorithm and extensions," in *Wiley Series in Probability and Statistics*. New York: Wiley, 1997.

[14] M. A. T. Figueiredo and A. K. Jain, "Unsupervised learning of finite mixture models," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 3, pp. 381–396, Mar. 2002.

[15] P. J. Green, "Bayesian reconstructions from emission tomography data using a modified EM algorithm," *IEEE Trans. Med. Imag.*, vol. 9, no. 1, pp. 84–93, Mar. 1990.

[16] Y. Zhang, M. Brady, and S. Smith, "Segmentation of brain MR images through a hidden Markov random field model and the expectation–maximization algorithm," *IEEE Trans. Med. Imag.*, vol. 20, no. 1, pp. 45–57, Jan. 2001.

[17] H. S. Choi, D. R. Haynor, and Y. Kim, "Partial volume tissue classification of multichannel magnetic resonance images—A mixel model," *IEEE Trans. Med. Imag.*, vol. 10, no. 3, pp. 395–407, Sep. 1991.

[18] P. Santago and H. D. Gage, "Quantification of MR brain images by mixture density and partial volume modeling," *IEEE Trans. Med. Imag.*, vol. 12, no. 3, pp. 566–574, Sep. 1993.

[19] P. Santago and H. D. Gage, "Statistical models of partial volume effect," *IEEE Trans. Image Process.*, vol. 4, no. 11, pp. 1531–1540, Nov. 1995.

[20] K. Blekas, A. Likas, N. P. Galatsanos, and I. E. Lagaris, "A spatially constrained mixture model for image segmentation," *IEEE Trans. Neural Netw.*, vol. 16, no. 2, pp. 494–498, Mar. 2005.

[21] A. Diplaros, N. Vlassis, and T. Gevers, "A spatially constrained generative model and an EM algorithm for image segmentation," *IEEE Trans. Neural Netw.*, vol. 18, no. 3, pp. 798–808, May 2007.

[22] G. S. Sanjay and T. J. Hebert, "Bayesian pixel classification using spatially variant finite mixtures and the generalized EM algorithm," *IEEE Trans. Image Process.*, vol. 7, no. 7, pp. 1014–1028, Jul. 1998.

[23] W. Qian and D. M. Titterington, "Estimation of parameters in hidden Markov models," *Philos. Trans. Roy. Soc. Lond. A Math. Phys. Sci.*, vol. 337, no. 1647, pp. 407–428, Dec. 1991.

[24] C. M. Bishop, *Pattern Recognition and Machine Learning*. New York: Springer-Verlag, 2006.

[25] C. M. Bishop, *Neural Networks for Pattern Recognition*. Oxford, U.K.: Oxford Univ. Press, 1995.

[26] S. L. Sclove, "Application of the conditional population-mixture model to image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-5, no. 4, pp. 428–433, Jul. 1983.

[27] Z. Liang, J. R. MacFall, and D. P. Harrington, "Parameter estimation and tissue segmentation from multispectral MR images," *IEEE Trans. Med. Imag.*, vol. 13, no. 3, pp. 441–449, Sep. 1994.

[28] Z. Liang, R. J. Jaszczak, and R. E. Coleman, "On reconstruction and segmentation of piecewise continuous images," in *Proc. Int. Conf. Inf. Process. Med. Imag.*, 1991, vol. 12, pp. 94–104.

[29] J. R. Jang, C. T. Sun, and E. Mizutani, *Neuro-Fuzzy and Soft Computing: A Computational Approach to Learning and Machine Intelligence*. Englewood Cliffs, NJ: Prentice-Hall, 1997, pp. 133–134.

[30] A. K. Jain and R. C. Dubes, *Algorithms for Clustering Data*. Englewood Cliffs, NJ: Prentice-Hall, 1988.

[31] J. C. Bezdek, *Pattern Recognition With Fuzzy Objective Function Algorithms*. New York: Plenum, 1981.

[32] G. McLachlan and D. Peel, *Finite Mixture Models*. New York: Wiley, 2000.

[33] N. Nasios and A. G. Bors, "Variational learning for Gaussian mixture models," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 36, no. 4, pp. 849–862, Aug. 2006.

[34] [Online]. Available: http://www.cma.mgh.harvard.edu/ibsr/data.html

[35] D. Martin, C. Fowlkes, D. Tal, and J. Malik, "A data base of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics," in *Proc. Int. Conf. Comput. Vision*, Jul. 2001, vol. 2, pp. 416–423.

[36] K. Lange and J. A. Fessler, "Globally convergent algorithms for maximum a posteriori transmission tomography," *IEEE Trans. Image Process.*, vol. 4, no. 10, pp. 1430–1438, Oct. 1995.

[37] P. Baldi and Y. Chauvin, "Smooth on-line learning algorithms for hidden Markov models," *Neural Comput.*, vol. 6, no. 2, pp. 307–318, Mar. 1994.

[38] J. Nocedal and S. J. Wright, *Numerical Optimization*. New York: Springer-Verlag, 1999.

Thanh Minh Nguyen received the B.E. degree in electrical and electronic engineering from the Ho Chi Minh City University of Technology, Ho Chi Minh City, Vietnam, in 2004 and M.A.Sc. degree from the Department of Electrical Engineering, Dayeh University, Changhua, Taiwan, in 2006. He is currently working toward the Ph.D. degree in the Department of Electrical and Computer Engineering, University of Windsor, Windsor, ON, Canada.

His research interests are computer vision and pattern recognition.

Q. M. Jonathan Wu (M'92–SM'09) received the Ph.D. degree in electrical engineering from the University of Wales, Cardiff, U.K., in 1990.

From 1995, he was with the National Research Council of Canada for ten years, where he became a Senior Research Officer and Group Leader. He is currently a Full Professor with the Department of Electrical and Computer Engineering, University of Windsor, Windsor, ON, Canada. He is the Tier 1 Canada Research Chair in Automotive Sensors and Sensing Systems. He has published more than 200 peer-reviewed papers in areas of computer vision, image processing, intelligent systems, robotics, microsensors and actuators, and integrated microsystems. He is on the editorial board of the *International Journal of Robotics and Automation*. His current research interests include 3-D computer vision, active video object tracking and extraction, interactive multimedia, sensor analysis and fusion, and visual sensor networks.

Dr. Wu is an Associate Editor for IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS—PART A: SYSTEMS AND HUMANS. He is a member of the IEEE Technical Committee on Robotics and Intelligent Sensing. He has served on the technical program committees and international advisory committees for many prestigious international conferences.