Image Segmentation Using Hidden Markov Gauss Mixture Models

Kyungsuk (Peter) Pyun, *Member, IEEE*, Johan Lim, Chee Sun Won, *Member, IEEE*, and Robert M. Gray, *Fellow, IEEE*

***Abstract*—Image segmentation is an important tool in image processing and can serve as an efficient front end to sophisticated algorithms and thereby simplify subsequent processing. We develop a multiclass image segmentation method using hidden Markov Gauss mixture models (HMGMMs) and provide examples of segmentation of aerial images and textures. HMGMMs incorporate supervised learning, fitting the observation probability distribution given each class by a Gauss mixture estimated using vector quantization with a minimum discrimination information (MDI) distortion. We formulate the image segmentation problem using a maximum *a posteriori* criteria and find the hidden states that maximize the posterior density given the observation. We estimate both the hidden Markov parameter and hidden states using a stochastic expectation-maximization algorithm. Our results demonstrate that HMGMM provides better classification in terms of Bayes risk and spatial homogeneity of the classified objects than do several popular methods, including classification and regression trees, learning vector quantization, causal hidden Markov models (HMMs), and multiresolution HMMs. The computational load of HMGMM is similar to that of the causal HMM.**

***Index Terms*—Bond-percolation (BP) model, Gauss mixture models (GMMs), Gauss mixture vector quantizer (GMVQ), image classification, image segmentation, parameter estimation, 2-D hidden Markov models (HMMs).**

I. Introduction

IMAGE segmentation is used in multimedia services for extracting explicit information about content so that human observers can interpret images clearly by highlighting specific regions of interest. For example, if segmentation of important regions from the background areas can be automated, the subsequent quantizer can be optimized to allocate more resources in areas of interest. The main purpose of this paper is to devise an automatic context-dependent segmentation algorithm yielding a reasonable classification error or Bayes risk between the original and the automatically segmented image.

Manuscript received May 14, 2006; revised April 11, 2007. This work was partially supported by National Science Foundation under Grant CCR 0309701 and by the Hewlett Packard Laboratories. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Giovanni Poggi.

K. Pyun was with the Information Systems Laboratory (ISL), Department of Electrical Engineering, Stanford University, Stanford, CA 94305 USA. He is now with the Imaging and Printing Group (IPG), Hewlett-Packard Company, San Diego, CA 92127 USA (e-mail: kspyun@gmail.com).

J. Lim is with the Department of Applied Statistics, Yonsei University, Seoul, Korea (e-mail: johanlim@yonsei.ac.kr).

C. S. Won is with the Department of Electronic Engineering, Dongguk University, Seoul, Korea (e-mail: cswon@dongguk.edu).

R. M. Gray is with the Information Systems Laboratory (ISL), Department of Electrical Engineering, Stanford University, Stanford, CA 94305 USA (e-mail: rmgray@stanford.edu).

Digital Object Identifier 10.1109/TIP.2007.899612

Compression and classification are intimately related to each other. Classification can be considered as a form of compression since segmenting an image into disjoint regions with class labels can be considered as a mapping from inputs into labels using an encoder. On the other hand, compression can be viewed as classification since a codeword label is assigned to each input pixel group. A successful combination of compression and classification was achieved in the isolated speech recognizer [17] designed as a cascade of compression and classification. The lossy compression technique of vector quantization coupled with suitable distortion measures has been successfully applied to image classification and segmentation [1], [31], [34]. This paper approaches image segmentation and classification problem from the same compression point of view.

For classification of different regions of interest (ROI), we use Gauss mixture models (GMM). GMMs have been a popular tool for fitting smooth densities to data in statistics and statistical signal processing [20], [29]. GMM has the following advantages. First, the Gaussian pdf maximizes the differential entropy and has the largest Shannon distortion-rate function and the worst high-rate distortion-rate tradeoff given the mean and the covariance [37], [23], [12], [18]. GMM inherits the nice properties of Gaussian pdf mentioned above, with better modeling of multimodal source densities [1]. Second, GMM models lead to a robust quantizer minimizing the quantizer mismatch [18].

A Gauss mixture vector quantizer (GMVQ) is a vector quantizer where the codewords correspond to Gaussian models [16]. For image distortion measure for VQ, we use the MDI distortion [17] since MDI has been shown to be competitive with other distortion measures for image classification and segmentation [35]. The overall VQ structure is a classified VQ [15], which can be viewed as a mixture of GMMs. To design a GMM, we employ a clustering approach using the Lloyd algorithm instead of the popular EM algorithm [13] to design a Gauss mixture. Advantages of the Lloyd algorithm over the EM algorithm include more rapid convergence, typically within 20 iterations in our experiments, and roughly half of the computational complexity of EM algorithm.

Accurate classification of GMM does not necessarily coincide with clear segmentation since the latter requires smooth boundaries between classes and better clustering for the same classes. Thus, additional spatial modeling is required for image segmentation. Among many context-dependent image segmentation algorithms [4], [14], [25], [26] one popular method is based on 2-D hidden Markov models. A 2-D HMM assumes that 1) the hidden states (true segmentations) are from a Markovian model, and that, 2) given the states of blocks, the feature vectors

are drawn from conditionally independent distributions, usually assumed to be Gaussian whose parameters vary with the states.

Since GMVQ is block-based, our algorithm is based on a block-based segmentation paradigm. Here, interblock modeling is necessary for the spatial correlation between the blocks. For block-based causal modeling, Li *et al.* [25], [26] assume that both training and testing images share same underlying Markovian model up to transition probabilities. They use causal HMMs [26] and MHMMs [25] and estimate transition probabilities with path-constrained 2-D Viterbi algorithms in training step. Also, note that both [26] and [25] impose one dimensional causal dependency, which is not natural for 1-D noncausal image data.

To overcome this drawback, the hidden Markov random field (MRF) was used. MRF model-based techniques can show generality for modeling images and fit naturally into the Bayesian formulation. MRF techniques were applied to image restoration [40], mine segmentation [36], and texture segmentation [30]; [30] used Markov chain Monte Carlo (MCMC) method on Bayesian paradigm. Another image segmentation method is multilevel logistic (MLL) model [42], [43]. However, the MLL model parameter is assumed to known *a priori*, limiting its applications. More complex models include pairwise Markov field (PMF) [33] and triplet Markov field (TMF) [3], [6]. The problem was the added computational difficulty. MRF-based expectation-maximization (EM) algorithm is a powerful tool for image segmentation. However, the conditional expectation in the EM algorithm is difficult to calculate. To overcome the computational difficulty, [10] used mean-field like approximation and our paper uses the mode approximation method as in [4]. This paper suggests to use HMGMM with the simplest noncausal Markovian model, the bond-percolation model (BP) [19] and its multistate extension (MBP). We do not assume homogeneity in spatial coherence among images, and estimate the hidden Markov model (HMM) parameter in testing step using stochastic EM procedure. The BP and MBP methods we employ for HMGMM resolve computational difficulty with the computation time comparable to that using causal HMMs [26].

A major novelty of the paper is from what we assume and how we estimate the spatial coherence of images. Both Li *et al.* [26] and Won *et al.* [43] assume that both training and testing images share same underlying Markovian model up to transition probabilities. They estimate those probabilities in a training step and simply use them to test new images. However, this homogeneity assumption is arguable in many images. In our paper, we assume that the spatial coherence can vary from one image to the other image, and we estimate it (the HMM parameter) in testing step.

This paper is generalizes the two-class segmentation approach proposed in [34] to multiclass segmentation problem by introducing an MBP model. We apply this extension to segmenting Brodatz texture images. Here, we should remark that the task of classifying Brodatz texture images is quite different from the task for aerial images in [34] in that training images do not provide any information on spatial coherence of true segments. Thus, [25] and [26] are not applicable to the problem.

The paper is organized as follows. In Section II, we briefly review the GMVQ. Section III introduces the HMGMM and proposes segmentation procedures using the HMGMM. In segmenting an image, we treat the single parameter of the noncausal dependency model as unknown, and jointly estimate it with the true segmentation. Section IV applies the suggested procedures to the segmentation of aerial images and Brodatz texture images. Section V concludes the paper.

II. Gauss Mixture Vector Quantization

We apply Gauss mixture models (GMMs) to estimate the density of the source. Real world signals have multiple modes in their probability density functions, which are often well approximated by a Gauss mixture. An important problem in Gauss mixture modeling is the estimation of the means and covariances of the Gaussian components. One approach is to estimate a partial covariance matrix of an autoregressive (AR) model from sample averages, which is nonsingular in most cases. To estimate the parameters of the Gauss mixture, we use the Lloyd clustering method. In GMVQ design by the Lloyd algorithm, an input vector X is mapped to the closest codeword (covariance and mean), i.e., the one minimizing the MDI distortion (see Section II-A for details). After applying this minimum distortion mapping to the entire training set, the codeword is replaced by the centroid of data assigned to the same class. These two steps are iterated until convergence (for details, refer to [15]). The resulting GMVQ is the collection of the probabilities of occurrence, means, and variances, $\{p_l, m_l, K_l\}$ which describes a single image. p_l, the underlying prior, is given by the relative frequency. However, the estimate $(\hat{m_x}, \hat{K_x})$ of the mean and covariance matrix given one observation of x is more challenging to compute. This is the same problem as that of the AR model for 100 milliseconds of speech [17]. AR models have been used for speech [17] and images [1], [11], [28], [35]. Often, the assumptions of AR models are stationarity and linearity, where the inverse covariance matrix has zeros in the positions corresponding to the large lags. If the assumption is that the random field is spatially stationary, the sample mean and sample covariance are good estimates and one only need compute the partial covariance K_x for lags smaller than two which is smaller than the size of x.

A. MDI Distortion Measure

The minimum discrimination information (MDI) distortion assigns a distortion resulting from modeling an input vector x as a pdf g as follows: If we had an estimate $\hat{f}_x$ of the density that produced x, then the relative entropy (or Kullback-Leibler divergence) $H(\hat{f}_x \| g)$ provides a measure of how dissimilar g and $\hat{f}_x$ are. In general, we cannot produce an accurate estimate of the joint pdf from a single vector x, but we can estimate the moments of the underlying distribution from sample averages. For example, we can form an estimate $\hat{K}$ of the covariance $E[(x - \hat{x})(x - \hat{x})^T]$ and an estimate $\hat{m}$ of the mean $E(x)$. The MDI distortion is defined as the minimum $H(\hat{f} \| g)$ over all pdfs $\hat{f}$ whose mean and variance is $\hat{m}$ and $\hat{K}$, respectively.

If g is Gaussian, then the estimate $\hat{f}_x$, minimizing the relative entropy given second-order constraints, is also Gaussian. Moreover, if $\hat{f}_x$ is Gaussian, the only parameters to be estimated

are the mean estimate $\hat{m}_x$ and the covariance estimate $\hat{K}_x$. The MDI distortion between two Gaussian pdf $\hat{f}_x$ and g is

$$H(\hat{f}_x||g) = \frac{1}{2}\left[(\hat{m}_x - m)^t K^{-1}(\hat{m}_x - m) + \mathrm{Tr}(\hat{K}_x K^{-1}) + \log\frac{|K|}{|\hat{K}_x|} - k\right] \quad (1)$$

where m and K is mean vector and covariance matrix of pdf g, and k is the dimension of x.

III. Testing With HMGMM

A. Intrablock and Interblock Processing

As with other block-based classifiers [1], [26], an image to be classified is divided into blocks and is modeled separately for intrablock and interblock information. Block processing is used for capturing global information and reducing the computation. Intrablock information plays a role in classifying blocks according to their own states, whereas interblock information works as a regularizing factor in segmentation. In this study, intrablock information is modeled by a GMM, while interblock information is modeled by a noncausal HMM. In every block, feature vectors are extracted as the statistics of the block. In prior work, DCT coefficients [26], wavelet coefficients [25], and the raw image itself [2] have been successfully used as feature vectors, depending on the application. As noted in [1], a GMM provides a useful model of these features.

The intrablock feature we use is the density of raw intensity of the block, evaluated from a GMM. It is estimated by a separate GMVQ for each class during the training phase in supervised learning. The parameters of the GMM are obtained from the Lloyd algorithm with an MDI distortion. Even for block processing, further reduction in computation can be obtained and the problem of the high dimensionality of image signal is solved by AR modeling. We consider 8×8 input vectors as 2-D blocks rather than 1-D vectors. For a 2-D AR model, we assume that the first-order dependence is enough to classify an image. A covariance estimator is the partial covariance of neighboring pixels within a 2-D AR window, where the sample averages are taken by sliding the window inside the block. As in [16], a full covariance matrix does not need to be calculated to find the minimum distortion codeword, which significantly reduces the number of computations required to design a GMVQ. We introduce the HMGMM below in detail and study the resulting segmentation procedure.

B. HMGMM

The following notation and assumptions are used here. An image space is divided into $N_x \times N_y$ nonoverlapping blocks. This block processing method is for global information and computational reduction. Let $\Omega = \{(i,j); 1 \le i \le N_x, 1 \le j \le N_y\}$ is the set of all image block indices, and (i,j) corresponds to the index of image blocks. A hidden random field for blocks is denoted as $\mathbf{Z} = \{Z_{ij}; (i,j) \in \Omega\}$, and the observed image random field as $\mathbf{X} = \{X_{ij}; (i,j) \in \Omega\}$.

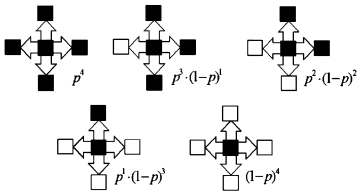

Fig. 1. BP model.

The proposed HMGMM is defined as two stages. First, given the hidden states $\mathbf{Z} = \{Z_{ij}\}$, the observed images $\mathbf{X} = \{X_{ij}\}$ are conditionally independent Gauss mixture distributions estimated by GMVQ as

$$P(\mathbf{X} = \mathbf{x}|\mathbf{Z} = \mathbf{z}) = \prod_{(i,j)\in\Omega} P(X_{ij} = x_{ij}|Z_{ij} = z_{ij}) \quad (2)$$

where $x_{ij} \in \Re$ and $P(X_{ij} = x_{ij}|Z_{ij} = z_{ij})$ is estimated by a GMM. Second, we model interblock dependency between the hidden states $\mathbf{Z}$ using the BP model or its extension to MBP model with single tuning parameter.

Assume a first-order Markov random field, such that the dependency comes from the up or down or left or right direction, as shown in Fig. 1, where the rectangle represents the block, the white and black rectangle represent the two different status of its spin, the block dependency is shown by the arrow, and p represents the probability that the neighboring block is in the same state in HMM. The probability measure is proportional to $p^{n_{\mathbf{con}}(\mathbf{Z})} \times (1-p)^{n_{\mathbf{dis}}(\mathbf{Z})}$, where $n_{\mathbf{con}}(\mathbf{Z})$ and $n_{\mathbf{dis}}(\mathbf{Z})$ are the numbers of the concordant and the discordant pairs, respectively. For a first-order MRF, there are four neighbor blocks around the image block of interest. Then there are five possible configurations for the model, as shown in Fig. 1. We assume that the model is isotropic and homogeneous (i.e., it is direction and location independent). If there are two neighbors in the same state, and two neighbors in a different state, which is the third case shown in upper right corner in Fig. 1, then the probability of this configuration is $p^2(1-p)^2$. The other remaining probabilities are calculated similarly. Thus, we represent the above BP model using symmetric Markov transition matrix

$$\mathbf{P} = \begin{pmatrix} p & 1-p \\ 1-p & p \end{pmatrix}$$

and the probability measure imposed is

$$P(\mathbf{Z}) = \frac{1}{\Psi(p)} p^{n_{\mathbf{con}}(\mathbf{Z})}(1-p)^{n_{\mathbf{dis}}(\mathbf{Z})} \quad (3)$$

where $n_{\mathbf{con}}(\mathbf{Z})$ is the number of concordant pairs, $n_{\mathbf{dis}}(\mathbf{Z})$ is the number of discordant pairs, and

$$\Psi(p) = \sum_{\mathbf{Z}} p^{n_{\mathbf{con}}(\mathbf{Z})}(1-p)^{n_{\mathbf{dis}}(\mathbf{Z})}.$$

The BP model assumes that the spins can be either parallel or anti-parallel which can be used for two-class segmentation

problems. As a generalization of this concept, we consider MBP, the model with q-state symmetric Markov transition matrix

$$\mathbf{P} = \begin{pmatrix} p & \frac{1-p}{q-1} & \frac{1-p}{q-1} & \cdots & \frac{1-p}{q-1} \\ \frac{1-p}{q-1} & p & \frac{1-p}{q-1} & & \\ \frac{1-p}{q-1} & \frac{1-p}{q-1} & p & & \vdots \\ \vdots & & & \ddots & \\ \frac{1-p}{q-1} & & & \cdots & p \end{pmatrix}$$

where p is the same as the hidden Markov parameter in the bond percolation model. The corresponding distribution of MBP is

$$P(\mathbf{Z}) = \frac{1}{\Psi(p)} p^{n_{\mathbf{con}}(\mathbf{Z})} \left(\frac{1-p}{q-1}\right)^{n_{\mathbf{dis}}(\mathbf{Z})} \tag{4}$$

where

$$\Psi(p) = \sum_{\mathbf{Z}} p^{n_{\mathbf{con}}(\mathbf{Z})} \left(\frac{1-p}{q-1}\right)^{n_{\mathbf{dis}}(\mathbf{Z})}.$$

In the model, the probability of a disconcordant pair is assumed to be the same as $(1-p)/(q-1)$ without regards to the direction of change. For $q = 2$ states, the above q-state MBP model reduces to the BP.

C. Segmentation With HMGMM

We assume the only parameter p in the MBP model is unknown and we estimate it jointly with the true segmentation. We use a maximum *a posteriori* (MAP) estimate for the true segmentation and compute the approximate maximum likelihood estimator of p often referred to as the hyper parameter for Gibbs prior [44]. Recall that $\mathbf{X}$ represents the observed image and $\mathbf{Z}$ is the unknown true segmentation. In the HMGMM, the MAP estimate of $\mathbf{Z}$, say $\mathbf{Z}^*$, becomes

$$\begin{aligned} \mathbf{Z}^* &= \operatorname{argmax}_{\mathbf{z}} \log P(\mathbf{Z} = \mathbf{z}|\mathbf{X}) \\ &= \operatorname{argmax}_{\mathbf{z}} \{\log P(\mathbf{X}|\mathbf{Z} = \mathbf{z}) + \log P(\mathbf{Z} = \mathbf{z})\}. \end{aligned} \tag{5}$$

The first term in the MAP (5), $\log P(\mathbf{X}|\mathbf{Z} = \mathbf{z})$ is modeled by the GMVQ. This step can be viewed as a training stage in supervised learning, whereas $P(\mathbf{Z} = \mathbf{z})$ (the probability of spatial correlations), which is the second term in the MAP (5), is evaluated in a testing stage. When a new image of interest comes for testing, the unknown parameter in MAP equation is evaluated each time, producing a different parameter for each run. The reason that this parameter is not estimated during training is that the clustering probability for the same class is different for each image.

Finding the MAP estimate in (5) is computationally very challenging and several Monte Carlo procedures [27] have been proposed in the literature. The simplest procedure, given the known parameter p, would be the iterative conditional mode (ICM) [4]. However, since the problem is not convex, the ICM strongly depends on the initial value and is easily captured at a local mode. Simulated annealing [14] is suggested to resolve such local mode difficulty by choosing an appropriate tempering schedule. The most naive procedure is based on samples from the posterior distribution $(\mathbf{Z}|\mathbf{X})$. In a high-dimensional problem, such as the image segmentation problem, drawing samples simultaneously for the whole image is not easy. We can indirectly solve this problem by generating a Markov chain and drawing samples from it (Markov chain Monte Carlo method (MCMC) [27]).

The estimation of p should be done jointly with the MAP estimation of $\mathbf{Z}$. The well known EM procedure shows that the maximum likelihood estimator (MLE), $\hat{p}$, maximizes the expected log-likelihood function

$$\begin{aligned} \mathrm{EL}(p) &= \mathrm{E}_{\mathbf{Z}}[\log p(\mathbf{X}, \mathbf{Z})|\mathbf{X}] \\ &= \mathrm{E}[\log p(\mathbf{X}|\mathbf{Z})|\mathbf{X}] + \mathrm{E}[\log p(\mathbf{Z})|\mathbf{X}] \\ &= \text{constant} - \log \Psi(p) \\ &\quad + \mathrm{E}\left[\sum_{\forall(i,j)\sim(k,l)} \mathrm{I}(Z_{ij} = Z_{kl}) \cdot \log p \right. \\ &\quad \left. + \sum_{\forall(i,j)\sim(k,l)} \mathrm{I}(Z_{ij} \neq Z_{kl}) \cdot \log(1-p) \middle| \mathbf{X}\right] \end{aligned}$$

where $\mathrm{E}[\log p(\mathbf{X}|\mathbf{Z})|\mathbf{X}]$ is a constant from training and $I()$ is an indicator function. By the derivation of $\mathrm{EL}(p)$, the MLE $\hat{p}$ is the solution to the equation

$$\frac{\Psi'(p)}{\Psi(p)} = \mathrm{E}\left(\frac{\sum_{(i,j)\sim(k,l)} \mathrm{I}(Z_{ij} = Z_{kl})}{p} + \frac{\sum_{(i,j)\sim(k,l)} \mathrm{I}(Z_{ij} \neq Z_{kl})}{1-p} \middle| \mathbf{X}\right).$$

By the way, from the definition of $\Psi(p)$

$$\begin{aligned} \frac{\Psi'(p)}{\Psi(p)} &= \mathrm{E}\left(\frac{\sum_{(i,j)\sim(k,l)} \mathrm{I}(Z_{ij} = Z_{kl})}{p} + \frac{\sum_{(i,j)\sim(k,l)} \mathrm{I}(Z_{ij} \neq Z_{kl})}{1-p}\right) \\ &= \frac{1}{p} \sum_{(i,j)\sim(k,l)} \mathrm{E}[\mathrm{I}(Z_{ij} = Z_{kl}] \\ &\quad + \frac{1}{1-p} \sum_{(i,j)\sim(k,l)} \mathrm{E}[\mathrm{I}(Z_{ij} \neq Z_{kl}] \\ &= \frac{1}{p} \sum_{(i,j)\sim(k,l)} p + \frac{1}{1-p} \sum_{(i,j)\sim(k,l)} (1-p) \\ &= 2 \times n_{\mathrm{tot}} \end{aligned}$$

where n_{tot} is the number of all edges between blocks in the image (that is, $n_{\mathrm{tot}} = \sum_{\forall(i,j)\sim(l,m)} 1$). Thus, $\hat{p}$ should satisfy

$$2 \times n_{\mathrm{tot}} = \mathrm{E}\left[\frac{\sum_{\forall(i,j)\sim(k,l)} \mathrm{I}(Z_{ij} = Z_{kl})}{p} + \frac{\sum_{\forall(i,j)\sim(k,l)} \mathrm{I}(Z_{ij} \neq Z_{kl})}{1-p} \middle| \mathbf{X}\right]. \tag{6}$$

By solving this for both in the cases of BP model and MBP model

$$\hat{p} = \frac{1}{n_{\text{tot}}} \sum_{\forall (i,j) \sim (l,k)} P(Z_{ij} = Z_{lk} | \mathbf{X}). \tag{7}$$

The main difficulty in estimating p comes from the conditional expectation in (7). Several suggestions have been made to approximate the expectation in the literature [9], [39]. We use the EM procedures based on samples from the posterior distribution with the current estimate. However, the E-step with Gibbs sampling or Metropolis-Hastings algorithms [41] can be computationally intensive. The mode approximation and the mean field approximations [44] suggest updating the estimate based on the current MAP estimate. Reference [4] proposes approximating the expected likelihood using the conditional mode of $\mathbf{Z}$ given the estimate $\hat{\mathbf{Z}}$, so that

$$\begin{aligned} \text{EL}(p) \approx \text{constant} &- \log \Psi(p) \\ &+ \sum_{\forall (i,j) \sim (k,l)} \text{I}(\hat{Z}_{ij} = \hat{Z}_{kl}) \cdot \log p \\ &+ \sum_{\forall (i,j) \sim (k,l)} \text{I}(\hat{Z}_{ij} \neq \hat{Z}_{kl}) \cdot \log(1-p) \end{aligned} \tag{8}$$

and

$$\hat{p} = \frac{\sum_{\forall (i,j) \sim (l,m)} \text{I}(\hat{Z}_{ij} = \hat{Z}_{lm})}{n_{\text{tot}}}. \tag{9}$$

On the other hand, the mean field method approximates

$$\begin{aligned} \text{EL}(p) \approx \text{constant} &- \log \Psi(p) \\ &+ \sum_{\forall (i,j) \sim (k,l)} \text{E}[\text{I}(Z_{ij} = Z_{kl}) | \mathbf{X}, \hat{\mathbf{Z}} \\ &- \{\hat{Z}_{ij}, \hat{Z}_{kl}\}] \cdot \log p \\ &+ \sum_{\forall (i,j) \sim (k,l)} \text{E}[\text{I}(Z_{ij} \neq Z_{kl}) | \mathbf{X}, \hat{\mathbf{Z}} \\ &- \{\hat{Z}_{ij}, \hat{Z}_{kl}\}] \cdot \log(1-p), \end{aligned}$$

and

$$\hat{p} \approx \frac{\sum_{\forall (i,j) \sim (l,m)} P(Z_{ij} = Z_{lm} | \mathbf{X}, \hat{\mathbf{Z}} - \{\hat{Z}_{ij}, \hat{Z}_{lm}\})}{n_{\text{tot}}}. \tag{10}$$

There are other estimation methods based on the current MAP estimates [24], [38].

The overall segmentation procedure becomes iterations of two steps, in general. One is how to find the MAP estimate of $\mathbf{Z}$ which becomes the segmentation of the testing image. The other is how to estimate the parameters p. In this paper, we update the MAP estimates according to the Gibbs sampler, and, to estimate p, we use the mode approximation by Besag [4] using the current MAP estimate.

IV. Results

The experiments presented in the following sections show that noncausal model results are considerably more precise than those produced by other well-known algorithms with similar computational loads, such as those based on causal HMMs. The HMGMM classifier was applied to aerial images in order to classify them according to whether they are images of man-made or natural areas, and to composite textures in order to segment the particular texture of interest. We have used BP for the segmentation of two classes, and have used MBP for the segmentation of more than two classes.

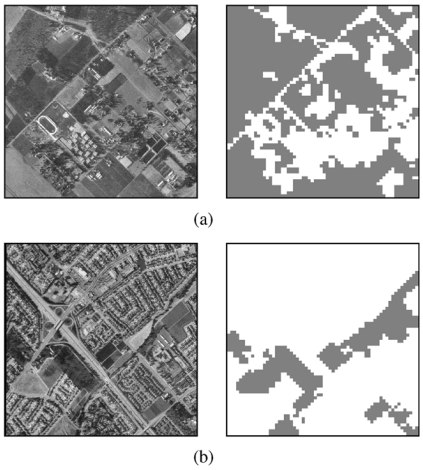

Fig. 2. Aerial images. (a) Image 1. (b) Image 2. Left: original, 8-bpp grayscale image; right: hand-labeled classified image. White: man-made; gray: natural.

During training, we divided the images into 8×8 blocks. Each block was processed independently by sliding a 2×2 AR window in a traditional scanning pattern. This first-order AR operation reduced the feature dimension of inputs from 64 dimensions to 4 dimensions. Estimates of the 4×1 mean vector m_i and the 4×4 covariance matrix K_i were used as features of the 8×8 block X. The number of clusters L was chosen to minimize the classification error based on six-fold cross-validation for all input images by varying the number of clusters in GMM from 1 to 10, as seen in Fig. 6. With this fixed cluster number L, the Lloyd algorithm with an MDI distortion was run to convergence to produce a GMM. After training, the test image was divided into 8×8 blocks and an initial classification using MDI clustering was performed independently for each block. After this initial classification result, the test image was divided again into 2×2 blocks for the fine classification. We invoked a BP model for spatial homogeneity using the interblock dependence between 2×2 blocks. Stochastic EM was employed to find the MAP hidden states.

A. Segmentation of Aerial Images

For segmentation of aerial images, we applied six-fold cross validation to evaluate the performance. Five images were used for training and the remaining one was used for testing for each iteration. All 512×512 images were gray-level images with 8 bits per pixel. The six images and their corresponding manually classified images appear in Figs. 2–4. Six aerial photographs of different locations in the San Francisco Bay Area are shown. These photographs have also been studied by Oehler [31], Li [26], [25], and Pyun *et al.* [34]. Each photograph is paired with its hand-labeled classified image. White represents the man-made class, whereas black represents the natural class. Certain images, such as the image 1 in Fig. 2, reveal a complex mixture

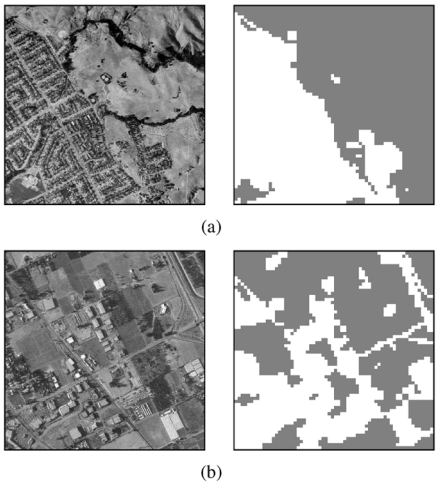

Fig. 3. Aerial images. (a) Image 3. (b) Image 4. Left: original 8-bpp grayscale images; right: hand-labeled classified images. White: man-made; gray: natural.

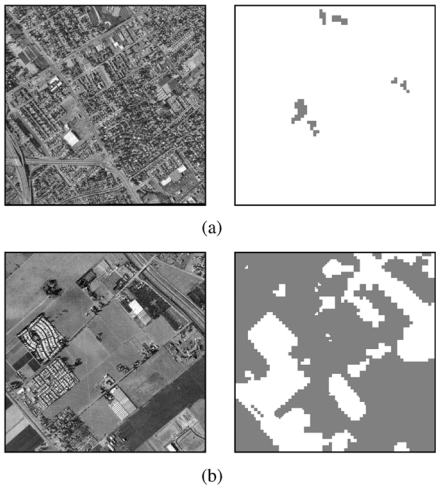

Fig. 4. Aerial images. (a) Image 5. (b) Image 6. Left: original 8-bpp grayscale images; right: hand-labeled classified images. White: man-made; gray: natural.

of man-made and natural areas, while other images, like image 3 in Fig. 3, show a mixture of mountainous areas and man-made structures such as houses.

For the stochastic EM algorithm, we experimented with a maximum of 5000 iterations. The log-likelihood plot after the maximization of each iteration of the stochastic EM algorithm is shown in Fig. 5. The log-likelihood converges within 1000 iterations. At around 400 iterations, the stochastic EM exits at a local maximum, which is typical for this algorithm owing to its stochastic nature. The plot is noisy for the same reason, but the fluctuations are proven to converge to the equilibrium status [5].

The Gibbs sampler normally converged to the stationary distribution in 1000 iterations, which took 165 s on a Pentium II 450-MHz computer with a Linux operating system. Although we needed to simulate Markov chains through lengthy iterations, the convergence speed for the stationary posterior distribution was comparable to that of other algorithms such as causal HMM [26], which took 200 s on a Pentium Pro 230-MHz computer.

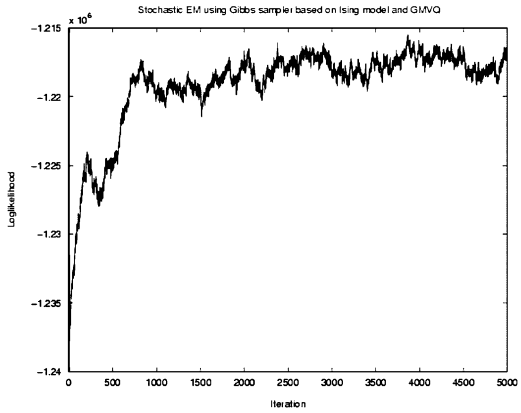

Fig. 5. Log-likelihood plot of the stochastic EM algorithm.

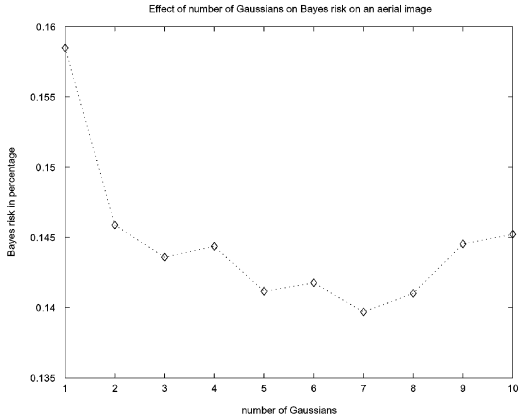

Fig. 6. Heuristic solution for the number of clusters.

We did the manual classification in 8×8 blocks. To do so, we took a majority vote among 2×2 blocks to classify 8×8 blocks, which is the smallest block size for which humans can easily classify image blocks. We applied HMGMM to segment the aerial images, and the composite textures. Each image was divided by 8×8 blocks to get a 4×1 mean vector and a 4×4 covariance matrix using a 2×2 AR window. The mean and covariance were used for features of the 8×8 block and input to the Lloyd algorithm. Initial classification was obtained by encoding each block independently by MDI distortion.

The results in Fig. 6 clearly show the advantage of using multiple clusters. A sharp decline in classification error was noted when clusters were increased from a single cluster to multiple clusters. Seven clusters with $P_e = 13.97\%$ yielded the minimum error, whereas a single cluster with $P_e = 15.85\%$ gave the worst result. Given a fixed training set, however, assigning too many clusters to each class resulted in less accurate classification performance owing to the imprecise estimate of the large number of GMM parameters. The classification performance degraded as the number of clusters was increased to more than seven. This degradation was caused by fitting the training data too well to the model and the bias-variance tradeoff [21]. Li *et al.* [25] obtained their best results by using five clusters for the natural class and nine clusters for the man-made class to fit wavelet coefficient features. Our results were similar, but we used the raw intensity images, with more accurate classification results.

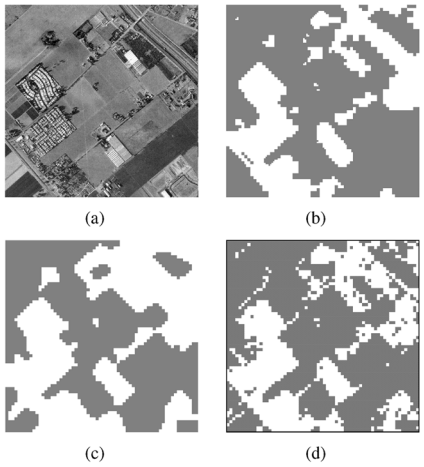

Fig. 7. Classification results of HMGMM and MHMM for an aerial image outside the training set: (a) original image, (b) manually classified image, (c) noncausal HMGMM with probability of error 14.72%, (d) MHMM with probability of error 11.57%. White: man-made; gray: natural.

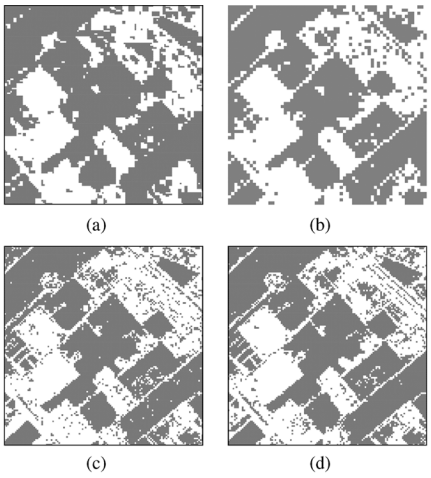

Fig. 8. Classification results of ARM, GM, and EM for an aerial image outside the training set: (a) causal HMM with probability of error 13.39%, (b) ARM with probability of error 18.46%, (c) CART with probability of error 20.29%, (d) LVQ with probability of error 18.13%. White: man-made; gray: natural.

We compared the results of classification by Bayes risk and segmentation by spatial homogeneity, with those of other popular methods such as classification and regression tree (CART) [7] for decision tree design, learning vector quantization (LVQ) [22], a causal HMM [26], multiresolution HMM [25], an autoregressive model (ARM) [2], a Gauss mixture (GM) based on log-likelihood [2], and Bayes VQ [31]. Bayes VQ was studied by Oehler [31] and Perlmutter [32] for segmentation of aerial images and medical images, respectively. For a fair comparison with HMGMM, we tried CART and LVQ with both interblock and intrablock features. However, ARM was originally assumed to be block-independent. A sample image, the manually classified image, and the algorithm classified images are shown in Figs. 7 and 8. Classification errors by Bayes risk for various design methods are given in Table I. The Bayes risk is the total number of mismatches divided by the total number of blocks. A binary risk of one and zero is used in aerial image segmentation, whereas a weighted unequal risk of misclassification is used in the medical image application. All results in Table I were based on six-fold cross validation.

TABLE I
COMPARISON OF CLASSIFICATION PERFORMANCE BETWEEN HMGMM AND OTHER ALGORITHMS

Algorithm	HMGMM	MHMM	causal HMM	GM
Bayes risk	13.97	16.02	18.80	14.42
Algorithm	ARM	CART	LVQ	Bayes VQ
Bayes risk	17.81	24.08	21.83	21.50

TABLE II
CLASSIFICATION PERFORMANCE FOR AERIAL IMAGES OUTSIDE THE TRAINING SET, BETWEEN HMGMM AND OTHER ALGORITHMS

Image index → Algorithm ↓	1	2	3	4	5	6
HMGMM	0.1501	0.1326	0.2261	0.1519	0.0302	0.1472
ARM	0.1972	0.1286	0.2516	0.2287	0.072	0.1905
GM	0.1563	0.1167	0.1790	0.1861	0.1231	0.1039
EM	0.2366	0.2825	0.2275	0.2556	0.2295	0.1652
HMM	0.1904	0.1765	0.2034	0.2405	0.1834	0.1339
CART	0.2263	0.1803	0.2899	0.2529	0.1425	0.2029
LVQ	0.2161	0.1918	0.2846	0.2492	0.1868	0.1813

The HMGMM result was obtained with seven clusters. Causal HMM [26] and MHMM [25] assumed the path to follow a 45° angle, which agrees with these particular aerial images. In HMGMM, there is no path dependency, so the orientations of classes do not matter for segmentation. The HMGMM algorithm yielded spatially smoother results than the MHMM and performed better than any other algorithm. Even the worst case with $P_e = 15.85\%$ using one cluster is more precise than MHMM, as shown in Table I. We also compared the performance for each iteration of six-fold cross validation. The detailed performance for each iteration is given in Table II. Classification performance is measured by the probability of error—the fraction of the number of incorrectly classified pixels to the total number of pixels in the image. Note that HMGMM gives good results for most of the iterations when compared with other algorithms. Finally, Figs. 7 and 8 show that we achieved better segmentation with less fragmented and more spatially homogeneous regions.

B. Segmentation of Composite Textures

As pointed out in Chapter I, we applied the HMGMM classifier to the problem of multiclass texture segmentation. We have chosen arbitrary compositions of textures from Brodatz database [8] to form composite textures, which were to be segmented. First, the successful segmentation of two class composite texture using HMGMM is shown in Fig. 9. In Figs. 10(a) and 11(a), the two 8-bpp images were mixtures of arbitrary three textures out of Brodatz textures. Their corresponding segmentation results for ARM algorithm and HMGMM algorithm are shown in Figs. 10(b) and (c) and 11(b) and (c), respectively. For each image, we trained each texture separately by applying supervised learning and obtained GMVQ. Here, we empirically chose the number of states in a testing image as six clusters [35], and the training was done using only textures appeared in the testing image.

The Brodatz texture database inherently do not contain any information on spatial coherence of true segments in composite textures. The composite textures were artificially generated

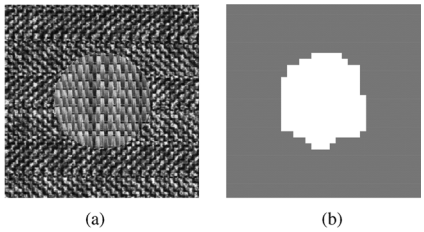

Fig. 9. Composite texture images: (a) original 8-bpp grayscale images, (b) HMGMM segmentation results, a composite texture of D17 and D55.

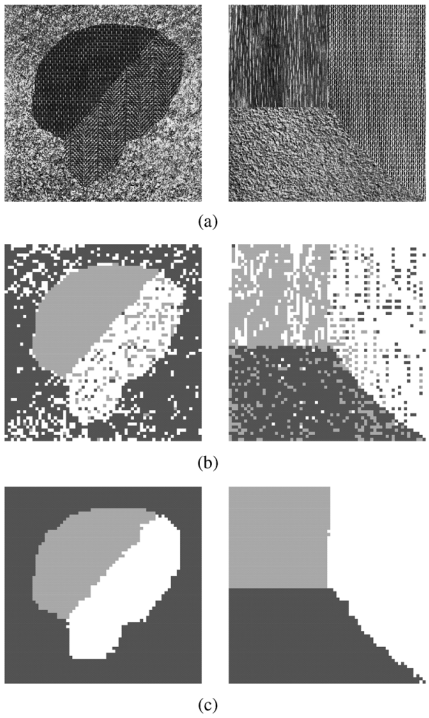

Fig. 10. Composite texture images: (a) original 8-bpp grayscale images, (b) ARM segmentation results, (c) HMGMM segmentation results. Left: Composite texture of D6, D9, and D17. Right: Composite texture of D4, D68, and D53.

and, thus, neither causal HMMs [26] nor MHMMs [25] are applicable to segmenting them. In this section, we compare HMGMM to ARM, since ARM performed as good as CART and LVQ. Also, their comparison could show the gain in misclassification error rate by considering spatial dependency in true segments.

In segmenting the mixture of three textures Figs. 10(a) and 11(a), the 2×2 symmetric transition matrix of BP model was modified to cover three states as explained in Section III-B. The segmentation results of HMGMM were superior to ARM results, as can be seen in Figs. 10(b) and (c) and 11(b) and (c). HMGMM results were homogeneous inside the class, while ARM results were very noisy. These results were obtained without any postprocessing procedure. The class boundaries of the HMGMM algorithm were relatively accurate even though the decision was based on 8×8 blocks. There was no special boundary modeling for HMGMM.

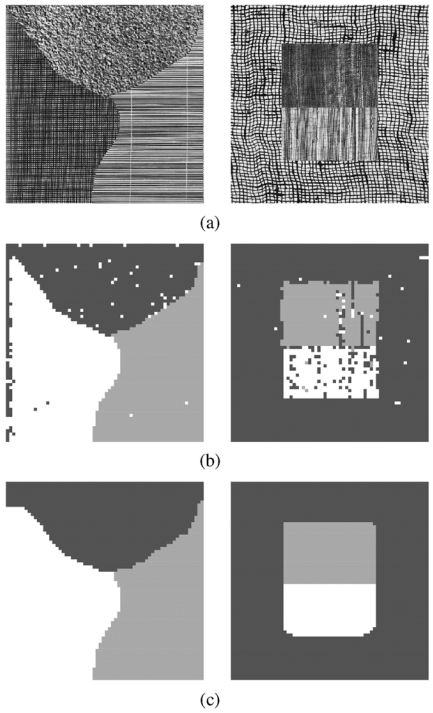

Fig. 11. Composite texture images: (a) original 8-bpp grayscale images, (b) ARM segmentation results, (c) HMGMM segmentation results. Left: Composite texture of D4, D49, and D21. Right: Composite texture of D104, D78, and D106.

V. Conclusion

We proposed a segmentation method combining GMVQ and 2-D hidden Markov modeling. The problem was formulated on the basis of Bayesian statistics. The block-wise context information using a 2-D hidden Markov model was incorporated to cluster the homogeneous classes. In order to improve segmentation by context, we devised an algorithm that models images by combining GMVQ and the MBP model to produce a 2-D noncausal HMGMM. A Bayesian prior was assumed to follow the MBP model. The stochastic EM algorithm was applied to optimize the MAP hidden states of the HMGMM of the image. This approach was used to identify man-made regions in aerial images and to segment multiclass composite textures. The resulting MAP states were more accurate and smoother than those produced by the other well-known algorithms, including the causal HMM, the multiresolution HMM, the CART, and the LVQ in terms of Bayes risk and spatial homogeneity of the segmented objects, with a computational load similar to that of the causal HMM.

Now we conclude the paper with some discussion and future works not addressed in the main body.

First, though we have shown several merits of the HMGMM, the HMGMM is still to be improved or investigated to resolve difficulties in practice. For example, we could have an images with damaged area or images with additional noise. We would like to see how does the proposed HMGMM to segment those images, and to modify the suggested procedure to resolve such difficulties.

Second, the HMM is growing fast. We find several more extended and more complex HMMs in recent literature. An example of such extended HMM is the triplet Markov field model (TMF) which allows correlated conditional errors. In texture segmentation, this correlated errors are understood to model the context-based information. The proposed HMGMM models the same information by taking sub-blocks which follows (multivariate) Gauss mixture model, but assumes those sub-blocks are independent to each other. It would be interesting to compare the HMGMM to those extended models numerically in time and misclassification rate.

ACKNOWLEDGMENT

The authors would like to thank the Associate Editor and the three referees for their many helpful comments and suggestions.

REFERENCES

[1] A. Aiyer, K. Pyun, Y. Huang, D. B. O'Brien, and R. M. Gray, "Lloyd clustering of Gauss mixture models for image compression and classification," *Signal Process.: Image Commun.*, vol. 20, pp. 459–485, 2005.

[2] A. K. Aiyer, "Robust image compression using gauss mixture models," Ph.D. dissertation, Elect. Eng. Dept., Stanford Univ., Stanford, CA, Aug. 2001.

[3] D. Benboudjema and W. Pieczynski, "Unsupervised image segmentation using triplet Markov fields," *Comput. Vis. Image Understand.*, vol. 99, no. 3, pp. 476–498, 2005.

[4] J. Besag, "On the statistical analysis of dirty pictures," *J. Roy. Statist. Soc. B*, vol. 48, pp. 259–302, 1986.

[5] J. Besag and P. J. Green, "Spatial statistics and Baysian computation," *J. Roy. Statist. Soc. B.*, vol. 55, pp. 25–37, 1993.

[6] J. Blanchet, F. Forbes, and C. Schmid, "Triplet Markov field designed for supervised classification of texture images," presented at the COMPSTAT, Rome, Italy, Aug. 2006.

[7] L. Breiman, J. H. Friedman, R. A. Olshen, and C. J. Stone, *Classification and Regression Trees*. Belmont, CA: Chapman & Hall, 1984.

[8] P. Brodatz, *Textures: Photographic Album for Artists & Designers*. New York: Dover, 1966.

[9] G. Celeux and J. Diebolt, "The SEM algorithm: A probabilisitic teacher algorithm derived from the EM algorithm for the mixture," *Comput. Statist. Quart.*, vol. 2, pp. 73–82, 1985.

[10] G. Celeux, G. Forbes, and N. Peyrard, "EM procedures using mean field-like approximations for Markov model-based image segmentation," *Pattern Recognit.*, vol. 36, no. 1, pp. 131–144, 2003.

[11] M. L. Comer and E.-J. Delp, "Segmentation of textured images using a multiresolution Gaussian autoregressive model," *IEEE Trans. Image Process.*, vol. 8, no. 3, pp. 408–420, Mar. 1999.

[12] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York: Wiley, 1991.

[13] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *J. Roy. Statist. Soc.*, vol. 39, no. 1, pp. 1–21, 1977.

[14] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distribution, and the Bayesian restoration of images," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 11, no. 6, pp. 689–691, Jun. 1984.

[15] A. Gersho and R. M. Gray, *Vector Quantization and Signal Compression*. Boston, MA: Kluwer, 1992.

[16] R. M. Gray, "Gauss mixture vector quantization," in *Proc. Int. Conf. Acoust., Speech, Signal Processing*, 2001, pp. 1769–1772.

[17] R. M. Gray, A. H. Gray, Jr., G. Rebolledo, and J. E. Shore, "Rate distortion speech coding with a minimum discrimination information distortion measure," *IEEE Trans. Inf. Theory*, vol. 27, no. 6, pp. 708–721, Nov. 1981.

[18] R. M. Gray and T. Linder, "Mismatch in high rate entropy constrained vector quantization," *IEEE Trans. Inf. Theory*, vol. 49, no. 5, pp. 1204–1217, May 2003.

[19] G. R. Grimmett, *Percolation*. New York: Springer, 1999.

[20] T. Hastie and R. Tibshirani, "Discriminant analysis by Gaussian mixtures," *J. Roy. Statist. Soc. B*, vol. 58, pp. 155–176, 1996.

[21] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning*. Berlin, Germany: Springer-Verlag, 2001.

[22] T. Kohonen, *Self-Organization and Associative Memory*. Berlin, Germany: Springer-Verlag, 1989.

[23] A. Lapidoth, "On the role of mismatch in rate distortion theory," *IEEE Trans. Inf. Theory*, vol. 43, no. 1, pp. 38–47, Jan. 1997.

[24] M. Lavielle and E. Moulines, "A simulated annealing version of the EM algorithm for non-Gaussian deconvolution," *Statist. Comput.*, vol. 7, no. 4, pp. 229–236, 1997.

[25] J. Li and R. M. Gray, "Image classification based on a multi-resolution two dimensional hidden Markov model," in *Proc. Int. Image Processing*, 1999, pp. 348–352.

[26] J. Li, A. Najmi, and R. M. Gray, "Image classification by a two-dimensional hidden Markov model," *IEEE Trans. Signal Process.*, vol. 48, no. 2, pp. 517–533, Feb. 2000.

[27] J. S. Liu, *Monte Carlo Strategies in Scientific Computing*. Berlin, Germany: Springer-Verlag, 2001.

[28] J. Mao and A. K. Jain, "Texture classification and segmentation using multiresolution simultaneous autoregressive models," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 2, pp. 173–188, Feb. 1992.

[29] G. J. McLachlan and K. E. Basford, *Mixture Models: Inference and Applications to Clustering*. New York: Marcel Dekker, 1988.

[30] D. Melas and S. P. Wilson, "Double Markov random fields and bayesian image segmentation," *IEEE Trans. Signal Process.*, vol. 50, no. 2, pp. 357–365, Feb. 2002.

[31] K. L. Oehler and R. M. Gray, J. A. Storer and M. Cohn, Eds., "Combining image classification and image compression using vector quantization," in *Proc. IEEE Data Compression Conf.*, Snowbird, UT, Mar. 1993, pp. 2–11.

[32] K. O. Perlmutter, "Compression and classification of images using vector quantization and decision trees," Ph.D. dissertation, Elect. Eng. Dept., Stanford Univ., Stanford, CA, 1995.

[33] W. Pieczynski and A.-N. Tebbache, "Pairwise Markov random fields and segmentation of textured images," *Mach. Graph. Vis.*, vol. 9, no. 3, pp. 705–718, 2000.

[34] K. Pyun, C. S. Won, J. Lim, and R. M. Gray, "Robust image classification based on a non-causal hidden Markov Gauss mixture model," presented at the Int. Conf. Image Processing, Rochester, NY, Sep. 2002.

[35] K. Pyun, C. S. Won, J. Lim, and R. M. Gray, "Texture classification based on multiple Gauss mixture vector quantizers," in *Proc. Int. Conf. Multimedia and Expo*, Aug. 2002, vol. 2, pp. 501–504.

[36] S. Reed, Y. Petillot, and J. Bell, "An automatic approach to the detection and extraction of mine features in sidescan sonar," *IEEE J. Ocean. Eng.*, vol. 28, no. 1, pp. 90–105, Jan. 2003.

[37] D. J. Sakrison, "Worst sources and robust codes for difference distortion measures," *IEEE Trans. Inf. Theory*, vol. 21, no. 5, pp. 301–309, May 1975.

[38] N. Ueda and R. Nakano, "Deterministic annealing variant of the EM algorithm," in *Proc. Neural Information Processing Systems 7*, 1995, pp. 545–552.

[39] G. C. G. Wei and M. A. Tanner, "Applications of multiple imputation to the analysis of censored regression data," *Biometrics*, vol. 47, no. 4, pp. 1297–1309, Dec. 1991.

[40] G. Winkler, *Image Analysis, Random Fields and Markov Chain Monte Carlo Methods: A Mathematical Introduction*. Berlin, Germany: Springer, 2003.

[41] W. K. Hastings, "Monte Carlo sampling methods using Markov chains and their applications," *Biometrika*, vol. 57, pp. 97–109, 1970.

[42] C. S. Won and R. M. Gray, *Stochastic Image Processing*. Norwell, MA: Kluwer, 2004.

[43] C. S. Won, K. Pyun, and R. M. Gray, "Automatic object segmentation in images with low depth of field," presented at the Int. Conf. Image Processing, Rochester, NY, Sep. 2002.

[44] Z. Zhou, R. N. Leahy, and J. Qi, "Approximate maximum likelihood hyperparameter estimation for Gibbs priors," *IEEE Trans. Image Process.*, vol. 6, no. 6, pp. 844–861, Jun. 1997.

Kyungsuk (Peter) Pyun (S'97–M'03) received the B.S. degree with honors from Seoul National University, Seoul, Korea, in 1996, and the M.S. and Ph.D. degrees from Stanford University, Stanford, CA, in 1999 and 2003, respectively, all in electrical engineering.

Since 2003, he has been with Hewlett-Packard Company (HP), where he is currently a member of technical staff at the HP San Diego R&D site, CA. He was a Research Staff Member at HP Labs, Palo Alto, CA, and Xerox, Palo Alto. He was an Adjunct Assistant Professor at Boise State University, Boise, ID, in 2005. His research interests include image segmentation, statistical classification and learning, and signal/image processing.

Johan Lim received the Ph.D. degree in statistics from Stanford University, Stanford, CA, in 2003.

He is currently an Assistant Professor in the Department of Applied Statistics, Yonsei University, Seoul, Korea, on leave from the Department of Statistics, Texas A&M University, College Station.

Chee Sun Won (M'88) received the B.S. degree in electronic engineering from Korea University, Seoul, in 1982, and the M.S. and Ph.D. degrees in electrical and computer engineering from the University of Massachusetts, Amherst, in 1986 and 1990, respectively.

From 1989 to 1992, he was a Senior Engineer with GoldStar Co., Ltd. (LG electronics), Seoul. In 1992, he joined Dongguk University, Seoul, where he is currently a Professor in the Electronic Engineering Department. He was a Visiting Associate Professor at Stanford University, Stanford, CA, from July 2001 to August 2002. His research interests include MRF image modeling, image segmentation, content-based image retrieval, and image watermarking.

Robert M. Gray (S'68–M'69–SM'77–F'80) was born in San Diego, CA, on November 1, 1943. He received the B.S. and M.S. degrees from the Massachusetts Institute of Technology, Cambridge, in 1966, and the Ph.D. degree from the University of Southern California, Los Angeles, in 1969, all in electrical engineering.

Since 1969, he has been with Stanford University, where he is currently the Lucent Technologies Professor of Engineering and Professor and Vice Chair of Electrical Engineering. His research interests are quantization, lossy compression, and statistical classification.

Dr. Gray was a member of the Board of Governors of the IEEE Information Theory Group (1974–1980 and 1985–1988), Associate Editor (1980–1983) of the IEEE TRANSACTIONS ON INFORMATION THEORY, and Co-Chair of the 1993 IEEE International Symposium on Information Theory. He was corecipient, with L. D. Davisson, of the 1976 IEEE Information Theory Group Paper Award and corecipient, with A. Buzo, A. H. Gray, and J. D. Markel, of the 1983 IEEE ASSP Senior. He received the IEEE Signal Processing Society 2005 Meritorious Service Award, the 1997 Technical Achievement Award, and the 1993 Society Award. He received IEEE Centennial and Third Millennium Medals. In 1998, he received a Golden Jubilee Award for Technological Innovation from the IEEE Information Theory Society. He is a Fellow of the Institute of Mathematical Statistics and was a Fellow of the John Simon Guggenheim Foundation in 1982. He received a 2002 Presidential Award for Excellence in Science, Mathematics, and Engineering Mentoring (PAESMEM) and the 2003 Distinguished Alumni in Academia Award from the University of Southern California. He was elected to the National Academy of Engineering in 2007.