[5] A. Hashizume, P. S. Yeh, and A. Rosenfeld, "A method of detecting the orientation of aligned components," *Pattern Recognit. Lett.*, vol. 4, pp. 125–132, 1986.
[6] L. A. Fletchler and R. A. Kasturi,"A robust algorithm for text string separation from mixed text/graphics images," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 10, pp. 910–918, Oct. 1988.
[7] D. S. Le, G. R. Thoma, and H. Wechsler, "Automated page orientation and skew angle detection for binary document images,"*Pattern Recognit.*, pp. 1325–1344, Oct. 1994.
[8] S. Chen and R. M. Haralick, "An automatic algorithm for text skew estimation in document images using recursive morphological transforms," in *Proc. IEEE Int. Conf. Image Processing (ICIP'94)*, Austin, TX, pp 139–143.
[9] D. S. Bloomberg, G.E. Copec, and L. Dasari, "Measuring document image skew and orientation," in *Proc. SPIE*, San Jose, CA, Feb. 1995, vol. 2422, pp. 302–316.
[10] H. Yan, "Skew correction of document images using interline cross-correlation," *CVGIP Graphic Models Image Processing*, vol. 55, pp. 538–543, Nov. 1993.
[11] P. Rousseeuw and A. Leroy, *Robust Regression and Outlier Detection*. New York: Wiley, 1987.
[12] A. Siegel,"Robust regression using repeated medians," *Biometrika*, vol. 69, pp. 242–244, 1982.
[13] M. Peter, D. Mintz, A. Rozenfeld, and D. Y. Kim, "Robust regression methods for computer vision: A review," *Int. J. Comput. Vis.*, vol. 6, pp. 59–70, 1991.
[14] S. Chaudhuri, S. Sharma, and S. Chatterjee, "Recursive estimation of motion parameters," *Comput. Vis. Image Understand.*, vol. 64, Nov. 1996.
[15] S. S. Sinha and B. G. Schunck, "A two-stage algorithm for discontinuity-preserving surface reconstruction," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 14, pp. 36–55, 1992.

EM Algorithm for Image Segmentation Initialized by a Tree Structure Scheme

Jong-Kae Fwu and Petar M. Djurić

***Abstract*— In this correspondence, the objective is to segment vector images, which are modeled as multivariate finite mixtures. The underlying images are characterized by Markov random fields (MRF's), and the applied segmentation procedure is based on the expectation-maximization (EM) technique. We propose an initialization procedure that does not require any prior information and yet provides excellent initial estimates for the EM method. The performance of the overall segmentation is demonstrated by segmentation of simulated one-dimensional (1-D) and multidimensional magnetic resonance (MR) brain images.**

I. Introduction

Image segmentation is a process that groups the image pixels with homogeneous attributes together and assigns them adequate labels. Many image segmentation methods are iterative and require good initial conditions for reliable performance. In general, if the initial conditions are inappropriate, the performance of the methods degrades significantly. In most of the literature related to iterative methods, the class parameters are assumed known, or if they are unknown, it is supposed that their initial values can be estimated from available training data [2].

In this correspondence, we propose a new approach for getting initial estimates [4]. It is based on a tree structure (TS) scheme usually used in the context of vector quantization [6]. We combine it with the expectation-maximization (EM) algorithm, which is a popular iterative method for maximum likelihood parameter estimation and image segmentation [3]. As with other iterative schemes, it is also sensitive to initialization, which can cause significant problems in practical applications. In [5], the EM algorithm was used to segment medical images, where the initial parameters were chosen according to a heuristic rule. Here we show that the combination of the TS and EM algorithm (TS-EM) requires neither training data nor operator's assistance. This, however, does not sacrifice the EM's excellent performance in parameter estimation and image segmentation.

This correspondence is organized as follows: The problem statement is given in Section II, and the details of the TS-EM algorithm are provided in Section III. Simulation results are presented in Section IV, and a brief conclusion is drawn in Section V.

II. Problem Statement

Let $\mathcal{S} = \{s = (i,j)\colon 1 \le i < M_1, 1 \le j < M_2\}$ denote an $M_1 \times M_2$ lattice, and $\mathcal{N}_s$ a set that contains the neighboring sites of s. Suppose also that $\boldsymbol{X}$ is a one-dimensional (1-D) Markov random field (MRF) defined on $\mathcal{S}$, or $\boldsymbol{X} = \{\boldsymbol{x}_s\colon s \in \mathcal{S}\}$. The joint probability density function of $\boldsymbol{X}$ is a Gibbs density whose form is

$$f(X|\beta) = \frac{1}{Z}\exp(U(X|\beta)) \tag{1}$$

where Z is a normalizing constant, $U(X|\beta)$ is an energy function, and β is the hyperparameter of the MRF. A realization of $\boldsymbol{X}$ is denoted by X, and a realization of the random vector $\boldsymbol{x}_s$ by x_s.

Let $\underline{\boldsymbol{Y}} = \{\underline{\boldsymbol{y}}_s\colon s \in \mathcal{S}\}$ be a vector random field that represents the observed image, where $\underline{\boldsymbol{y}}_s$ is obtained when the noise $\underline{\boldsymbol{w}}_s$ is superimposed to the signal $\underline{g}(\boldsymbol{x}_s)$. That is

$$\underline{\boldsymbol{y}}_s = \underline{g}(\boldsymbol{x}_s) + \underline{\boldsymbol{w}}_s \tag{2}$$

with $\underline{g}(\boldsymbol{x}_s)$ being a vector function that maps the underlying label $\boldsymbol{x}_s$ to its associated attribute vector $\underline{\boldsymbol{\mu}}_{x_s}$. Note that this model is, in general, valid for images with homogeneous attributes within the same region, such as X-ray and MRI images, but may be invalid for other images, such as positron emission tomography (PET) and single photon emission computed tomography (SPECT). We assume that the noise samples $\underline{\boldsymbol{w}}_s$ of $\underline{\boldsymbol{W}}$ are independent and distributed according to a multivariate Gaussian distribution with zero mean and unknown covariance matrix.

Each pixel of X belongs to one of m different classes, where m is known. Given the above assumptions, the probability density function of the observed vector at the site s can be expressed as

$$f(\underline{y}_s|x_s,\Theta_{\underline{Y}}) = \frac{1}{(2\pi)^{p/2}|\boldsymbol{\Sigma}_{x_s}|^{1/2}} \cdot \exp\left(-\frac{1}{2}(\underline{y}_s - \underline{\mu}_{x_s})^T\boldsymbol{\Sigma}_{x_s}^{-1}(\underline{y}_s - \underline{\mu}_{x_s})\right) \tag{3}$$

where $\Theta_{\underline{Y}} = \{\underline{\mu}_{x_s}, \boldsymbol{\Sigma}_{x_s}\colon x_s = 1,2,\cdots,m\}$, $\boldsymbol{\Sigma}_{x_s}$ is the covariance matrix and p is the dimension of the vector $\underline{y}_s$. Note that, in general,

Manuscript received August 3, 1995; revised April 11, 1996. This work was supported by the National Science Foundation under Award MIP-9506743. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. William E. Higgins.

The authors are with the Department of Electrical Engineering, State University of New York, Stony Brook, NY 11794-2350 USA (e-mail: djuric@sbee.sunysb.edu).

Publisher Item Identifier S 1057-7149(97)01138-X.

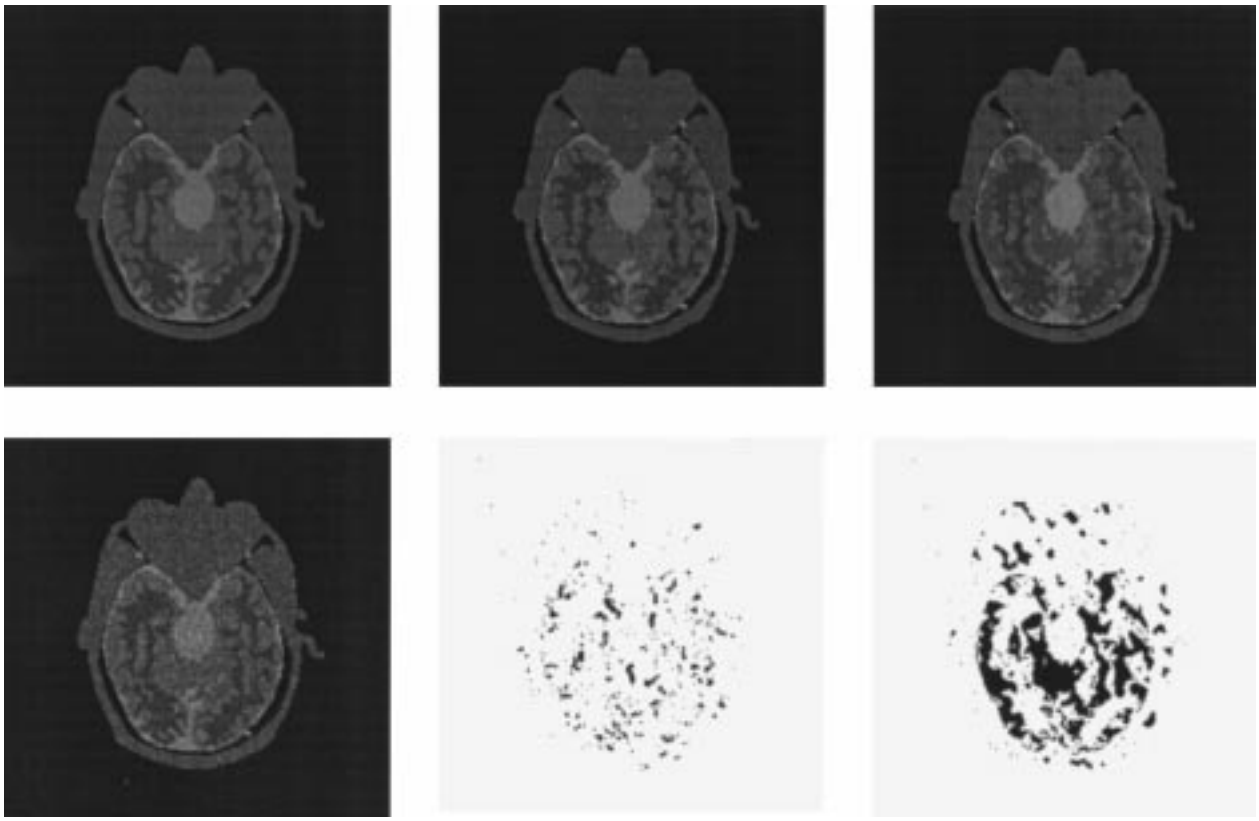

Fig. 1. Left: noiseless and noisy MR images. Middle: segmented image initialized by TS-EM and its error map. Right: segmented image initialized by the K-means algorithm and its error map.

the covariance matrices associated with each class of pixels are different.

Given these assumptions and $\underline{Y}$, the main objective is to segment $\underline{Y}$ into m classes and estimate the unknown parameters $\Theta_{\underline{Y}}$.

III. Algorithm

The segmentation of $\underline{Y}$ and the estimation of $\Theta_{\underline{Y}}$ are entangled tasks and are usually carried out simultaneously. To resolve the stated problem, we want to apply the EM algorithm. First we write

$$f(\underline{Y}, X|\Theta) = f(\underline{Y}|X, \Theta) f(X|\Theta) \tag{4}$$

$$\simeq \prod_{s\in\mathcal{S}} f(\underline{y}_s|x_s, \Theta_{\underline{Y}}) \prod_{s\in\mathcal{S}} f(x_s|\mathcal{N}_s, \beta) \tag{5}$$

where Θ includes both the class parameters $\Theta_{\underline{Y}}$ and the hyperparameter β of the MRF. Note that in writing (5) from (4), we adopted the "pseudolikelihood" (PL) principle [1]. Next, define $\Phi = \{\underline{\phi}_s : s \in \mathcal{S}\}$, where $\underline{\phi}_s^T = [\phi_s^1 \phi_s^2 \cdots \phi_s^m]$ is an indicator vector of length m whose components are ideally all zero except for the lth one, where l is the label of the sth pixel. In practice, as well as in this work, the elements of $\underline{\phi}_s$ represent the conditional probability that the vector $\underline{y}_s$ belongs to the lth class. From the EM algorithm, after the nth iteration, the lth element of $\underline{\phi}_s$ is [7]

$$\phi_s^{l(n)} = \frac{f(\underline{y}_s|x_s = l, \Theta^{(n)}) f(x_s = l|\mathcal{N}_s, \beta)}{\sum_{l=1}^{m} f(\underline{y}_s|x_s = l, \Theta^{(n)}) f(x_s = l|\mathcal{N}_s, \beta)} \quad l = 1, 2, \cdots, m. \tag{6}$$

The remaining equations yield the $(n+1)$st iteration, and they are given [3], [7] by

$$\underline{\mu}_l^{(n+1)} = \frac{\sum_{s\in\mathcal{S}} \phi_s^{l(n)} \underline{y}_s}{\sum_{s\in\mathcal{S}} \phi_s^{l(n)}}, \qquad l = 1, 2, \cdots, m \tag{7}$$

$$\boldsymbol{\Sigma}_l^{(n+1)} = \frac{\sum_{s\in\mathcal{S}} \phi_s^{l(n)} (\underline{y}_s - \underline{\mu}_l)(\underline{y}_s - \underline{\mu}_l)^T}{\sum_{s\in\mathcal{S}} \phi_s^{l(n)}} \quad l = 1, 2, \cdots, m. \tag{8}$$

As mentioned before, the EM algorithm requires accurate initialization if it is to perform reliably.

Our objective is to develop a procedure for providing initial conditions, which is completely data driven. To achieve this goal, we propose a TS algorithm whose scheme is identical to the TS vector quantization (TSVQ) method [6]. The scheme represents a sequence of binary searches that usually leads to suboptimal solutions. It is implemented in stages, where at each stage a fixed number of classes are considered, and at the kth stage there are k classes assumed. In stage one $k = 1$, that is, we assume the data belong to one class only, and we easily find the class parameters. Note that for this stage we do not need initial values for the parameters. For the next stage, $k = 2$, the data are assumed to come from two classes. To apply the EM algorithm, we need the initial values of the class parameters. To obtain them, we perturb the estimated mean of the data from stage one in opposite directions along the eigenvector associated with the largest eigenvalue of the estimated covariance matrix. The initial values are determined by applying a simple iterative scheme that converges in

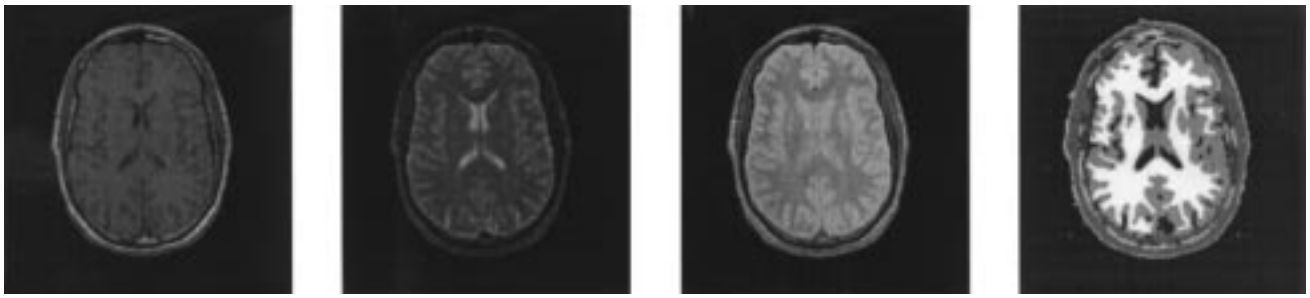

Fig. 2. From left to right: T1 image, T2 image, Pd image, and segmented image using TS-EM.

a few steps. Next, the EM algorithm is implemented, which further improves the estimates of the class parameters.

Once the estimation in the second stage is completed, we go on with setting initial values for the third stage $(k = 3)$. Here we have two possibilities: to perturb the estimated mean vector of the first or the second class from the second stage. We perturb them one at a time and repeat the procedure as in stage two. Since the two possibilities lead to two different solutions, we choose the better one according to the density function $f(X_q|\underline{Y})$, where q represents the first or second possibility and X_q maximizes $f(X_q|\underline{Y})$. With the completion of the third stage, we are ready to move on to the fourth one. The steps are similar to those already described. We continue in this manner until the number of classes equals m. Note that in the last stage, we have $m-1$ different initial conditions for the EM algorithm that operates on m classes.

The selection of the class that is being perturbed in the kth stage is obtained as follows. From Bayes' theorem

$$f(X_q|\underline{Y}) = \frac{f(\underline{Y}|X_q)f(X_q)}{f(\underline{Y})}. \tag{9}$$

Clearly, the maximization of $f(X_q|\underline{Y})$ does not depend on $f(\underline{Y})$, and therefore $f(\underline{Y})$ can be ignored. To avoid difficulties in dealing with all the possible configurations of X as well as the intractable partition function Z, which appears in $f(X_q)$, we adopt the pseudolikelihood. The maximum *a posteriori* (MAP) criterion then becomes

$$\hat{q} = \arg \max_{q \in \{1,2,\cdots,k-1\}} f(\underline{Y}|X_q)f(X_q) \tag{10}$$

$$= \arg \min_{q \in \{1,2,\cdots,k-1\}} \{F_d(\underline{y}_s, \hat{\Theta}) + F_c(x_s, \hat{\Theta})\} \tag{11}$$

where

$$F_d(\cdot) = \sum_{s \in \mathcal{S}} \{\tfrac{1}{2} \ln |\boldsymbol{\Sigma}_{x_s}| + \tfrac{1}{2} (\underline{y}_s - \underline{\mu}_{x_s})^T \boldsymbol{\Sigma}_{x_s}^{-1} (\underline{y}_s - \underline{\mu}_{x_s})\} \tag{12}$$

represents the fitting error (data term) and

$$F_c(\cdot) = \sum_{s \in \mathcal{S}} \{-\ln f(x_s|\mathcal{N}_s)\} \tag{13}$$

provides the smoothness (continuity) constraint.

The sequence of binary searches greatly reduces the computational load and provides very good results. The procedure can further be accelerated if the computations of each stage are implemented in parallel. This will allow linear increase in computation time with the increase of the number of classes. The performance of the overall TS-EM scheme is excellent because, for the pixel classification, we not only use the statistical properties of the pixels but also exploit their spatial interrelationships that are quantified by the MRF.

TABLE I
TRUE AND ESTIMATED PARAMETERS BY THE EM ALGORITHM INITIALIZED BY THE TS-EM AND K-MEANS ALGORITHMS

	Parameters				
true parameters values	20.00	80.00	100.00	120.00	160.00
TS-EM	20.04	78.29	100.08	121.22	162.17
K-means	20.05	81.22	105.68	127.73	163.01

The algorithm is implemented according to the following scheme. In the first stage, we start with one class, $x_s = 1, \forall s \in \mathcal{S}$. We evaluate $\hat{\underline{\mu}}_1$, which is the sample mean vector of all the pixels as well as the covariance matrix

$$\hat{\boldsymbol{\Sigma}}_1 = \frac{1}{M1 \times M2} \sum_{s \in S} (\underline{y}_s - \underline{\mu}_1)(\underline{y}_s - \underline{\mu}_1)^T.$$

In the kth stage, where $k = 2, 3, \cdots, m-1$, we perform the following steps:

Step 1: Set the initial mean vectors of the k classes as $(\hat{\underline{\mu}}_1 + \underline{\epsilon}, \hat{\underline{\mu}}_1 - \underline{\epsilon}, \hat{\underline{\mu}}_2, \cdots, \hat{\underline{\mu}}_{k-1})$, where $\underline{\epsilon}$ is a perturbation vector that splits the first class. Its direction is identical to the direction of the eigenvector associated with the largest eigenvalue of $\hat{\boldsymbol{\Sigma}}_1$.

Step 2: Classify all the pixels into one of the k classes by

$$x_s = l \quad \text{if} \quad d(\underline{y}_s, \hat{\underline{\mu}}_l) \le d(\underline{y}_s, \hat{\underline{\mu}}_{l'}), \quad l \ne l' \quad \text{and} \quad l, l' \in \{1, 2, \cdots, k\} \tag{14}$$

where $d(\cdot)$ is the p-dimensional Euclidean norm, and l and l' are the different possible labels of the pixels at site s. Then update $\hat{\underline{\mu}}_l$, and $\hat{\boldsymbol{\Sigma}}_l$ for $l = 1, 2, \cdots, k$.

Step 3: Repeat Step 2 until convergence is achieved, i.e., until the number of changing underlying pixels is zero or below certain predefined number.

Step 4: Use $\hat{\underline{\mu}}_l$, and $\hat{\boldsymbol{\Sigma}}_l$ for $l = 1, 2, \cdots, k$ as initial estimates for the EM algorithm to segment the image. The hyperparameter β is estimated from the pseudolikelihood and updated after each iteration. Repeat the EM procedures until the estimates converge.

Step 5: As in Steps 1 to 4, split each class $(2, 3, \cdots, k-1)$ one at a time to form additional classes. Choose the result from all the $k-1$ candidates using the MAP criterion (11) as the result for the kth stage.

IV. SIMULATION RESULTS

In this section, we present the simulation results of two experiments. In the first, we applied the TS-EM procedure to one-dimensional $(p = 1)$ synthesized MR brain image, and in the second, to a three-dimensional $(p = 3)$ real MR brain image. The sizes of all these images are 256 × 256.

TABLE II
THE ROBUSTNESS AGAINST NOISE WITH VARIOUS CNR'S

CNR	20/10	20/15	20/20	20/25
TS-EM	99.4 %	97.5 %	93.6 %	77.2%
EM initialized by K-means	89.7%	86.3 %	75.9 %	72.1%

In Fig. 1, the left two images are the noiseless and noisy MR images. The middle two images are the segmented image initialized by TS-EM and its error map. The right images are the segmented image initialized by the K-means algorithm and its error map. The contrast-to-noise ratio (CNR) defined by

$$\text{CNR} = \min_{l,k} \left\{ \frac{|\mu_l - \mu_k|}{\sigma} \right\}, \qquad l \neq k \quad \text{and} \quad l, k \in \{1, 2, \cdots, m\} \tag{15}$$

was equal to $\frac{20}{10}$. In (15), μ_l and μ_k are the intensity levels of the pixels in two different *adjacent* regions, respectively, and σ is the noise standard deviation. There were five classes (tissues) with true mean values $\mu_1 = 20, \mu_2 = 80, \mu_3 = 100, \mu_4 = 120,$ and $\mu_5 = 160$. The noise deviations associated with the various class were $\sigma_1 = 3.5, \sigma_2 = \sigma_3 = \sigma_4 = \sigma_5 = 20$. The results clearly show that the method was able to segment the image successfully even in areas where the details of the patterns are fine. In Table I we also provide the true mean values of the various classes and compare them to the estimated values obtained by the EM method initialized by the TS and K-means algorithms.

We also present the results of the average percentage of correct classifications (PCC) for the different CNR's out of 50 trials. They are given in Table II, and they show that the TS-EM outperforms the K-means algorithm significantly.

The results of the second experiment are displayed in Fig. 2. The first three images are real images, and the remaining one is the segmented image.

V. CONCLUSIONS

A TS scheme for initialization of an EM algorithm for image segmentation and parameter estimation has been proposed. The scheme comprises a sequence of binary searches. It starts with the assumption that the pixels come from one class and, subsequently, it increases the number of classes one at a time. In each stage, the results obtained from the previous stage are used to construct initial estimates for the current stage. The algorithm stops when the image is segmented into a predefined number of classes. The performance of the algorithm was examined by computer simulations. They showed excellent results.

REFERENCES

[1] J. Besag, "Statistical analysis of non-lattice data," *Statistician*, vol. 24, pp 179–195, 1975.
[2] ——, "On the statistical analysis of dirty pictures," *J. Roy. Stat. Soc. B*, vol. 48, pp. 259–302, 1986.
[3] A. P. Dempster, N. M. Laird, and D. B. Rubin, " Maximum likelihood from incomplete data via the EM algorithm," *J. Roy. Stat. Soc. B* no. 1, pp. 1–38, 1977.
[4] J. K. Fwu and P. M. Djurić, "Unsupervised vector image segmentation by a tree structure—ICM algorithm," *IEEE Trans. Med. Imag.*, vol. 15, pp. 871–880, 1996.
[5] Z. Liang, R. J. Jaszczak, and R. E. Coleman, "Parameter estimation of finite mixtures using EM algorithm and information criteria with application to medical image processing," *IEEE Trans. Nucl. Sci.*, vol. 39, pp. 1126–1133, 1992.
[6] N. Phamdo, N. Farvardin, and T. Moriya, "A unified approach to tree-structure and multistage vector quantization for noisy channels," *IEEE Trans. Inform. Theory*, vol. 39, pp. 835–850, 1993.
[7] J. Zhang, J. Modestino, and D. Langan, "Maximum-likelihood parameter estimation for unsupervised stochastic model-based image segmentation," *IEEE Trans. Image Processing*, vol. 3, pp. 404–419, 1994.