A METHOD FOR MULTI-DIMENSIONAL IMAGE SEGMENTATION

Keith Hartt, Mark J. Carlotto, and Mark W. Brennan

The Analytic Sciences Corporation
55 Walkers Brook Drive
Reading, Massachusetts 01867
(617)942-2000

ABSTRACT

A multi-dimensional (multi-spectral) image segmentation technique is presented, based on a multivariate Gaussian mixture distribution which includes spatial dependence through the incorporation of a Markov random field that governs the grouping of pixel classifications into regions. Estimation of parameters for each component in the mixture, an unsupervised clustering problem, is performed using a specialization of the EM algorithm, an iterative scheme for maximum likelihood parameter estimation. The commonly used heuristic Isodata algorithm is cast in this theoretical framework and is shown to be an approximate method under a restricted model assumption; furthermore, a computational approach is presented which allows our more general procedure to be employed with comparable efficiency. As the EM algorithm is essentially a parameter-refinement procedure, which only locally maximizes the likelihood function, good initial estimates for parameters are crucial in obtaining satisfactory results. Two methods are presented in this regard which involve incremental building of model complexity: 1) a histogram analysis technique in which a smoothed 1D cluster histogram is parsed into a number of components based on inflection points; 2) a straightforward method in which a maximum-extent cluster is replaced by a two-component mixture. The segmentation estimate is optimal in the sense that, based on the statistical model, it is the most probable outcome for pixel classifications given the multi-dimensional imagery. Our segmentation technique is being used extensively in the building of a surface-material-class database from Landsat TM imagery; practical experience with the method is presented.

1. INTRODUCTION

Methods for multi-dimensional image segmentation abound; surveys of previous work can be found in [7], [9], [12]. The goal of image segmentation is to extract a set of regions in which certain important characteristics are homogeneous. Kanade [12] divides multispectral image segmentation techniques into three categories comprised of: 1) those which use local spatial information for region merging; 2) those which use spectral information (i.e., the spatially independent distribution of the multi-dimensional data) for region splitting; 3) those which use spectral and spatial information.

In this paper a technique is presented with which we have had practical success, and which has a computationally efficient implementation. It is based on computational techniques and a realistic image model with sound theoretical underpinnings. The method is a member of category 3), as it is based on a mixture distribution which incorporates spatial dependence. As is customary, each component, or cluster, in the mixture is modeled using a normal distribution. Here, an unsupervised clustering technique is used to estimate the number of clusters and the statistical characteristics of each cluster. One method presented for parsing, or sub-dividing, clusters is a scale-space histogram analysis technique [3] extended to a multi-dimensional setting. Cluster parameters are refined using an iterative technique. The spatial grouping of pixels into regions, i.e., the classification of each pixel as a member of a cluster, is governed by a categorically valued Markov random field which has a multi-level logistic distribution [6]. Image models using Markov random fields for segmentation and restoration in a single-image setting have been widely used [5], [6], [9], [13]. The model promotes the formation of compact regions by assigning a high probability to groupings of identically classified pixels. The optimization criterion is based on a Bayesian formulation; in particular, the segmentation is a MAP estimate, the most probable pixel classifications given the multi-dimensional data. In the computation of the optimal segmentation, a relaxation procedure is used which involves locally updating the segmentation estimate [5], [9].

The paper is organized as follows. In Section 2, the clustering procedure is described. The output of the clustering is essentially parameters used in the segmentation, which is described in Section 3. Experimental results are given in Section 4.

2. CLUSTERING

The model used for clustering is given by the following normal mixture distribution, which is the marginal distribution at all pixel sites of the multi-dimensional image data.

$$p(\mathbf{x}) = \sum_k \pi_k \left(\frac{1}{\prod_m \sqrt{2\pi}\,\sigma_{m,k}} \right) \exp\left(-\frac{1}{2} \sum_m \left(\frac{x_m - \mu_{m,k}}{\sigma_{m,k}} \right)^2 \right) \quad (1)$$

where m indexes data dimension, $\mathbf{x}$ denotes $\{x_m\text{: all } m\}$, k indexes the cluster or component; π_k is a relative frequency or prior weight, where $\sum \pi_k = 1$; $\mu_{m,k}$ is a mean, $\sigma_{m,k}$ a standard deviation. Note that the i.i.d. assumption relative to location in the lattice is relaxed in the segmentation technique presented later.

Some words justifying the clustering model are in order. The general aim is to represent the data using a probability distribution comprised of unimodal distributions, each of which corresponds to a cluster of data. The data are described compactly by cluster membership. A normal distribution provides a tractable, reasonable unimodal distribution. Note that if the number of clusters is not upper-bounded, the model is perfect in the sense that each multi-dimensional datum can be assigned an infinitesimally narrow unimodal distribution centered on itself, resulting in the data having a likelihood of one. This of course is not a useful model, so the number of components must be reasonably restricted. The model assumes a diagonal covariance matrix for component normal distributions for simplicity; the computational burden is relaxed considerably relative to a general covariance matrix. Furthermore, even if an independence assumption is not valid, the model is adequate if the marginal description affords inter-cluster statistical separability. In our view, based on experiment, a compute-time/generality tradeoff supports the model used here.

EM Algorithm

The EM algorithm is a general procedure for maximum likelihood estimation given incomplete data. The specialization of the EM algorithm applicable to the mixture estimation problem at hand is given by the following iteration [16]

$$\begin{aligned} \hat{\pi}_k^{(n)} &= \frac{1}{N^2} \sum_{i,j} p^{(n-1)}(k \mid \mathbf{x}_{ij}) \\ \hat{\mu}_{m,k}^{(n)} &= \frac{1}{\pi_k^{(n)} N^2} \sum_{i,j} x_{m,i,j}\, p^{(n-1)}(k \mid \mathbf{x}_{ij}) \\ \hat{\sigma}^{2\,(n)}_{m,k} &= \frac{1}{\pi_k^{(n)} N^2} \sum_{i,j} (x_{m,i,j} - \hat{\mu}_{m,k}^{(n)})^2\, p^{(n-1)}(k \mid \mathbf{x}_{ij}) \end{aligned} \quad (2)$$

where n indexes the iteration, (i,j) indexes location in the image lattice consisting of N^2 pixels, $\mathbf{x}_{ij}$ denotes $\{x_{m,i,j}: \text{all } m\}$. Parameter estimates for each cluster are in the form of sample statistics weighted by a cluster membership probability, given by

$$p^{(n-1)}(k \mid \mathbf{x}_{ij}) = \frac{p^{(n-1)}(\mathbf{x}_{ij} \mid k)\, \pi_k}{\sum_k p^{(n-1)}(\mathbf{x}_{ij} \mid k)\, \pi_k} \qquad (3)$$

where $p^{(n-1)}(\mathbf{x}_{ij} \mid k)$ is the normal distribution of the multi- dimensional data for component k, computed using the parameter set $\{\pi_k, \mu_{m,k}, \sigma_{m,k}\}^{(n-1)}$ obtained at iteration (n-1). Note k is interpreted as a random integer as well as an index to clarify the presentation; e.g., p(k) could be used to denote π_k. The convergence behavior of the EM algorithm is that at each iteration, the likelihood function is increased; more information in this regard can be found in [16]. The measure used to test for convergence of the EM algorithm is given by the discrimination information [14] between mixture distributions at successive iterations

$$I = \sum_m \sum_k \frac{1}{2} (\sigma^{2\,(n)}_{m,k} - \sigma^{2\,(n-1)}_{m,k})(\sigma^{-2\,(n)}_{m,k} - \sigma^{-2\,(n-1)}_{m,k}) + \frac{1}{2} (\sigma^{-2\,(n)}_{m,k} + \sigma^{-2\,(n-1)}_{m,k})(\mu^{(n)}_{m,k} - \mu^{(n-1)}_{m,k})^2 \qquad (4)$$

Discrimination information between the model at one iteration and the previous one falling below a specified threshold signals convergence.

It is worthwhile to note the relationship of this algorithm to the commonly used Isodata algorithm [7]. The mixture EM algorithm is equivalent under the assumption that standard deviations and relative frequencies are uniform across clusters, and when the cluster membership probability is approximated by

$$p(k \mid \mathbf{x}_{ij}) = \begin{cases} 1 & \text{if } k = \arg\max_k p(k \mid \mathbf{x}_{ij}) \\ 0 & \text{otherwise} \end{cases} \qquad (5)$$

In essence, a hard decision about cluster membership is made at each iteration by using the approximation; under the assumption that relative frequencies and standard deviations are uniform, this decision at each pixel site is given by the cluster with its center closest in the Euclidean sense to the multi-dimensional datum. The members of each cluster then each contribute to a new estimate for their center to be used in the next iteration.

Cluster Parsing

The EM algorithm is a parameter refinement procedure. Initial estimates of parameter values are necessary, and the number of components must be specified. In this section, procedures for building up the model by subdividing clusters is presented. The methods involve first assuming a single cluster, and then recursively subdividing.

A simple heuristic for subdividing is to divide in two the cluster having the largest standard deviation, in the hope of obtaining a model consisting of compact clusters. The parameters of the subdivided cluster are related to the parent cluster by

$$\pi_0 = \pi_1 = \frac{\pi}{2}\ ;\ \mu_0 = \mu - \frac{\sigma}{\sqrt{2}}\ ;\ \mu_1 = \mu + \frac{\sigma}{\sqrt{2}}\ ;\ \sigma_0 = \sigma_1 = \frac{\sigma}{\sqrt{2}} \qquad (6)$$

where $\{\pi, \mu, \sigma\}$ is the parameter set associated with cluster k along dimension m, with

$$\{m,k\} = \arg\max_{\{m,k\}} \sigma_{m,k} \qquad (7)$$

All other parameters remain unchanged. The relationship (6) is such that the two-component mixture distribution is equivalent to the parent distribution up to the second moment. This sort of heuristic is commonly used in conjunction with Isodata [17]. It is more attractive here however, as standard deviations are an explicit part of the clustering model.

A second method for cluster sub-division is based on a multi-dimensional extension of scale-space histogram analysis [3]. The method (and implementation) is based on the following statistic, which is an empirical marginal distribution, a cluster histogram for one dimension of the multi-dimensional data.

$$h_k(x_m) = \frac{1}{N^2} \sum_{\{(i,j):\, x_{m,i,j} = x_m\}} p(k \mid \mathbf{x}_{ij}) \qquad (8)$$

Note that this is a sufficient statistic for calculations in (2). Compute time is reduced because once the statistic is computed, the large sums over all (i,j) are replaced by small sums over all values of x_m, for typical eight-bit images. Use of the statistic lessens the computation carried out at each pixel site, the computation which dominates overall compute time. Additionally, in parsing clusters, the histogram statistic allows the use of scale-space histogram analysis in a multi-dimensional setting.

Scale-space histogram analysis is described briefly as follows. The method performs a modal analysis on a 1D histogram using inflection points, zero-crossings of the second derivative, in a smoothed version of the histogram. Initial estimates for cluster parameters are related to the inflection points $\{x_n\}$ by

$$\hat{\pi}_n = \int_{x_n}^{x_{n+1}} h_s(x)\, dx$$
$$\hat{\mu}_n = \frac{1}{2}(x_n + x_{n+1})\ ;\ \hat{\sigma}_n = \frac{1}{2}(x_{n+1} - x_n) \qquad (9)$$

for n even, where $h_s()$ is the smoothed histogram, and relative frequencies are appropriately normalized. A Gaussian-shaped smoothing kernel is used, because it has the attractive property that the number of modes monotonically decreases with increasing kernel width [3].

The method is applied to the multi-dimensional data by applying the technique recursively to the 1D cluster histograms. The multi-dimensional data are weighted by the membership probability for a cluster, and the weighted data are then histogrammed along a dimension of the multi-dimensional data, resulting in a cluster histogram to which histogram analysis is applied. If multiple modes are found, the cluster is subdivided according to (9) for that dimension; all other parameters remain unchanged. The procedure is repeated, each time followed by EM iteration, for all clusters along all dimensions until all cluster histograms are unimodal by the histogram analysis criterion. The recursive process results in a powerful yet manageable method for the parsing of multi- dimensional clusters. An alternate interpretation of the formation of the cluster histograms is as follows. The multi-dimensional data histogram bins are weighted by cluster membership probabilities. The result is an empirical distribution of the cluster of multi-dimensional data. The weighted bins are projected onto coordinate axes, forming empirical marginal distributions.

Computational Considerations

The computational impact of using the cluster histogram statistic has been described. In addition, a look-up table approach is taken for the calculation of cluster-membership probabilities, which are given by

$$p(k \mid \mathbf{x}_{ij}) = \frac{a_k}{\sum_k a_k} \qquad (10)$$

$$a_k = \exp\left(\log(\pi_k) - \sum_m \left(\log(\sigma_{m,k}) + \frac{1}{2}\left(\frac{x_{m,i,j} - \mu_{m,k}}{\sigma_{m,k}} \right)^2 \right) \right)$$

The elements of the sum in the exponent are precomputed before sweeping the image, for each possible value of $x_{m,i,j}$. For typical eight-bit images, the computations comprise a small set, which allows storage in a look-up table. With the remaining logarithm in a_k precomputed, the computation of the cluster-membership probability at each pixel site involves a number of sums, an exponentiation, and a division. The result is that the total amount of computation is comparable to that in Isodata, which makes our algorithm an attractive, more general alternative.

3. SEGMENTATION

The result of clustering is a set of parameters for each cluster in the mixture model. In this section the segmentation, the final hard decision about the cluster membership of each multi-dimensional datum on the image lattice, is presented. The model used for segmentation is identical to that used for clustering with the exception that the independence assumption associated with marginal description of the cluster membership of each pixel is relaxed. Toward this end, the enhanced model for cluster membership which incorporates spatial dependence is described. The optimization criterion is then derived based on the model, and computational aspects of the segmentation algorithm are presented.

Multi-Level Logistic Model

The multi-level logistic distribution was proposed by Derin and Elliott in their work on single-image segmentation [6]. The distribution is conveniently presented in the context of Markov random fields, described through Gibbs distributions. The main utility of the Gibbs distribution lies in the fact that a random field can be described in a tractable manner in terms of local interactions which are in accord with desired behavior. After some preliminaries on general Gibbs models, the specifics of the multi-level logistic model are provided.

The Markov property on a lattice is defined with respect to a *neighborhood system* $\{\eta_{ij}\}$, where η_{ij} denotes the neighborhood of pixel site (i,j). The Markov property is given by

$$p(w_{ij} \mid \{w_{kl}: (k,l)\neq(i,j)\}) = p(w_{ij} \mid \{w_{kl}: (k,l)\in\eta_{ij}\}) \qquad (11)$$

where $\{w_{ij}\}$ is a realization of a general 2D field. A Markov random field can always be expressed in the form of a Gibbs distribution [1]

$$p(\{w_{ij}\}) = \frac{1}{Z}\exp\left(-\sum_{c\in C} V_c(\{w_{ij}\})\right) \qquad (12)$$

where c is a set of pixels that are neighbors of each other, called a *clique*; C is the set of all cliques; $V_c(\{w_{ij}\})$ is a *clique potential* which describes the interaction among members of clique c; Z is a normalizing constant. The relationship between the Gibbs distribution and the Markov property is demonstrated by the following. Consider

$$p(w_{ij} \mid \{w_{kl}: (k,l)\neq(i,j)\}) = \frac{p(\{w_{ij}\})}{\int p(\{w_{ij}\})\, dw_{ij}} \qquad (13)$$

where $p(\{w_{ij}\}))$ is given by (12). Note that the normalizing constant and all clique potentials cancel with the exception of the potentials involving cliques of which (i,j) is a member. This gives rise to the Markov property (11).

The multi-level logistic distribution used here describes the cluster-membership field, consisting of a set of cluster memberships, one for each pixel site, denoted by $\{k_{ij}\}$; i.e., it is a 2D field of random integers which label clusters. Clique potentials are given by

$$V_c(\{k_{ij}\}) = \begin{cases} -\omega_c & \text{if all } k_{ij} \text{ in c are equal} \\ \omega_c & \text{otherwise} \end{cases} \qquad (14)$$

Only certain pair clique potentials are non-zero, which results in a field Markov with respect to the eight nearest neighbors. Cliques and associated potentials are depicted in Fig. 1.

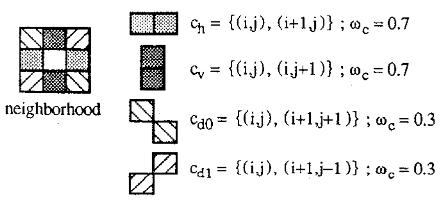

Fig. 1. Clique potentials.

Optimization Criterion

The segmentation is based on the MAP estimate, given by

$$\{\hat{k}_{ij}\} = \arg\max_{\{k_{ij}\}} p(\{k_{ij}\} \mid \{\mathbf{x}_{ij}\}) \qquad (15)$$

which is the most probable set of cluster memberships given the set of multi-dimensional data. The posterior distribution is given by Bayes' rule

$$p(\{k_{ij}\} \mid \{\mathbf{x}_{ij}\}) = \frac{\left[\prod_{ij} p(\mathbf{x}_{ij} \mid k_{ij})\right] p(\{k_{ij}\})}{p(\{\mathbf{x}_{ij}\})} \qquad (16)$$

Note that as in the clustering model, the multi-dimensional data are assumed independent given cluster membership, so that the joint conditional distribution is expressed as a product of marginals. However, the cluster-membership field incorporates dependence through the multi-level logistic model of spatial interaction. Spatial as well as spectral (in the case of multi-spectral data) clustering is incorporated in the segmentation model. The choice of clique potentials above associates a high probability with spatially compact regions consisting of identically classified pixel sites. Note that the denominator in (16) has no dependence on cluster membership, so is simply a normalizing constant which does not effect the maximization (15).

Relaxation Method

Direct computation of the maximization (15) is not practically feasible, as it involves a search over all possible values of the cluster-membership field, which grows exponentially with N^2. A relaxation method is therefore presented which involves iteratively updating the cluster membership at each site. The update involves a local computation given by the following conditional distribution derived from (16)

$$p(k_{ij} \mid \{k_{\kappa l}: (\kappa,l)\neq(i,j)\},\{\mathbf{x}_{ij}\}) = \frac{p(\mathbf{x}_{ij} \mid k_{ij})\, p(\{k_{ij}\})}{\sum_{k_{ij}} p(\mathbf{x}_{ij} \mid k_{ij})\, p(\{k_{ij}\})} \qquad (17)$$

where the sum in the denominator is over all values of k_{ij}. Note that the posterior distribution is Markov, so that the computation involves only the multi-dimensional datum and the cluster membership at site (i,j), and its neighborhood of cluster memberships. The calculation for the particular multi-level logistic model used here involving only pair cliques is given by

$$p(k_{ij} \mid \{k_{ij}: (i,j)\in\eta_{ij}\}, \mathbf{x}_{ij}) = \frac{a[k_{ij}]}{\sum_{k_{ij}} a[k_{ij}]} \qquad (18)$$

$$a[k_{ij}] = \exp\left(\sum_{\{c:\,(i,j)\in c\}} 2\omega_c\, \#\{(\kappa,l):\ (\kappa,l)\neq(i,j),\ (\kappa,l)\in c,\ k_{\kappa l}=k_{ij}\} - \sum_m \left(\log(\sigma_{m,k_{ij}}) + \frac{1}{2}\left(\frac{x_{m,i,j}-\mu_{m,k_{ij}}}{\sigma_{m,k_{ij}}}\right)^2\right)\right)$$

The elements in the first summation in the exponent are either $2\omega_c$ or zero, depending on whether the membership neighboring (i,j) in the clique is equal to k_{ij} or not, respectively. This results in a conditional probability of an outcome k_{ij} which increases with the number of neighbors having membership k_{ij}. It is instructive to compare this expression with (10), the clustering-model membership probability distribution. It is identical with the exception of the first term in the exponent, associated with the distribution of cluster memberships which, because of spatial dependence, are indexed relative to the image lattice. The look-up table method of computation is employed here as well.

The relaxation method is a deterministic version of the Gibbs sampler [9]. The image is repeatedly swept, with replacement at each pixel site given by

$$\hat{k}_{ij} = \arg\max_{k_{ij}} p(k_{ij} \mid \{k_{ij}: (i,j)\in\eta_{ij}\}, \mathbf{x}_{ij}) \qquad (19)$$

the estimate which is the locally most probable value. Following the relaxation procedure, the state of the cluster membership field converges to a local maximum of the *global* posterior distribution (16). To avoid imposing the time causality on the lattice, the site visit ordering is chosen according to the coding method [1]. The initial configuration is the maximum likelihood estimate under an independence assumption, which is given by (19) with the conditional probability (18) modified by setting the first term in the exponent to zero.

4. EXPERIMENTAL RESULTS

Practical experience with the method is demonstrated with a surface-material-classification application to six-band Landsat Thematic Mapper (TM) multi-spectral data. Our method is currently being used in this way to produce a database over a large geographical area.

Here, a (1383x1125)-pixel area was extracted from TM Path 14 Row 30 which covers a study area surrounding Glens Falls, NY. The tasseled-cap transformation [4] was applied to the image resulting in brightness, greenness, and wetness images. This transformation is a linear rotation which allows reducing the dimensionality of the data in a manner appropriate for the sensor and the discrimination of natural surface materials. Figure 2 shows the brightness image, which is essentially panchromatic. A hybrid approach to image classification combining supervised and unsupervised classification strategies was taken. The approach involves the manual selection of training sets for each surface material category. However, the distribution of each category is not assumed unimodal; i.e., the clustering algorithm is applied to each category. The clusters for each category are combined, and the segmentation algorithm is applied. Figure 3 shows the segmentation result, where the grey-level mapping is given by the cluster mean in brightness. A total of 51 clusters are represented, each of which corresponds to one of nine surface material classes of interest: agriculture, urban, bare soil, open water, grassland, brush/scrub, deciduous forest, coniferous forest, and mixed forest.

Accuracy of the surface material classification obtained using the segmentation was measured based on ground truth obtained from 1:15,840 scale SCS orthophotos, field notes and terrestrial photographs. A 95% confidence interval was constructed for the overall classification accuracy according to [15]. The overall classification accuracy with respect to the nine surface material classes was 86.8%, which is good relative to other research using similar classification categories [11], [18]. The agriculture, open water, grassland, coniferous forest and mixed forest classes yielded the best results (82.9 - 95.6 %). The lowest classification agreement was for the urban (79.5 %) and brush/scrub (58.5 %) classes, which is attributable to their consisting of relatively complex mixtures. More detail on this experiment may be found in [2].

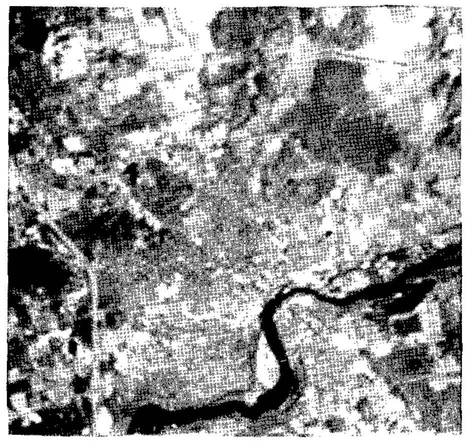

Fig. 2. Brightness image.

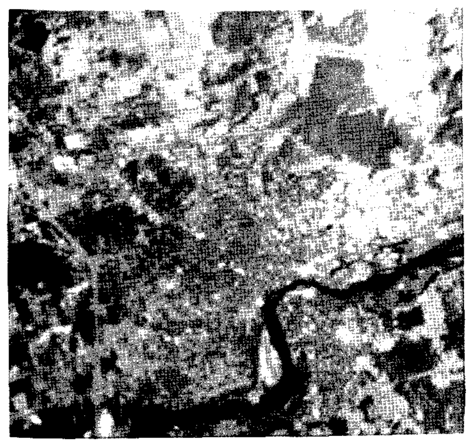

Fig. 3. Segmentation.

Summary

A new segmentation algorithm is described using a rigorous statistical framework. The presentation includes the extension of a histogram analysis technique to a multi-dimensional setting, as well as the extension of results of single-image segmentation studies for application to multi-dimensional data.

REFERENCES

[1] J. E. Besag, "Spatial interaction and the statistical analysis of lattice systems," *J. Roy. Stat. Soc.*, Ser. B, vol. 36, 1974, pp. 192-236.

[2] M. W. Brennan, T. J. Moore, and B. L. Spence, "Ten meter database development using Landsat Thematic Mapper and digitized map data: preliminary results," *Tech. Papers 1989 ASPRS-ACSM Ann. Conv.*, vol. 3, pp. 165-172.

[3] M. J. Carlotto, "Histogram analysis using a scale-space approach," *IEEE Trans. Pat. An. Mach. Intel.*, vol. PAMI-9, no. 1, Jan. 1987, pp. 121-129.

[4] E. P. Crist and R. C. Cicone, "A physically-based transformation of Thematic Mapper data -- the TM tasseled cap," *IEEE Trans. Geo. Rem. Sens.*, vol. GE-22, no. 3, May 1984, pp. 256-263.

[5] F. S. Cohen and D. B. Cooper, "Real time textured image segmentation based on noncausal Markovian random field models," *Proc. SPIE Conf. Intel. Robots*, Cambridge, MA, Nov. 1983.

[6] H. Derin and H. Elliott, "Modeling and segmentation of noisy and textured images using Gibbs random fields," *IEEE Trans. Pat. An. Mach. Intel.*, vol. PAMI-9, no. 1, Jan. 1987, pp. 39-55.

[7] R. O. Duda and P. E. Hart, *Pattern Classification and Scene Analysis*. New York: Wiley, 1973.

[8] K. S. Fu and J. K. Mui, "A survey on image segmentation," *Patt. Recog.*, vol. 13, pp. 3-16.

[9] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images," *IEEE Trans. Pat. An. Mach. Intel.*, vol. PAMI-6, Nov. 1984, pp. 721-741.

[10] R. M. Haralick and L. G. Shapiro, "Image segmentation techniques," *Comp. Vis., Gr., Img. Proc.*, vol. 29, 1985, pp. 100-132.

[11] M. Hashim, "Crop identification using merged Landsat multispectral scanner and Thematic Mapper data: preliminary analysis," *Tech. Papers 1988 ASPRS-ACSM Ann. Conv.*, vol. 4, pp. 11-20.

[12] T. Kanade, "Region segmentation: signal vs semantics," *Computer Graphics and Image Processing*, vol. 19, 1980, pp. 279-297.

[13] P. A. Kelly, H. Derin and K. D. Hartt, "Adaptive segmentation of speckled images using a hierarchical random field model," *IEEE Trans. Acoust., Speech, Sig. Proc.*, vol. ASSP-36, Oct. 1988, pp. 1628-1641.

[14] S. Kullback, *Information Theory and Statistics*. New York: Dover, 1968.

[15] J. Neter, W. Wasserman, and G. A. Whitmore, " *Applied Statistics*. Boston: Allyn and Bacon, Inc., 1982.

[16] R. Redner and H. Walker, "Mixture densities, maximum likelihood and the EM algorithm," *SIAM Review*, vol. 26, no. 2, Apr. 1984, pp. 195-239.

[17] J. T. Tou and R. C. Gonzalez, *Pattern Recognition Principles*. Reading, MA: Addison Wesley, 1974.

[18] J. Vogel, "An evaluation of the use of TM digital data for updating the land cover component of the SCS 1987 multiresource inventory of New Jersey," *Tech. Papers 1987 ASPRS-ACSM Ann. Conv.*, vol. 1, pp. 113-119.