Segmentation of Dynamic PET or fMRI Images Based on a Similarity Metric

Jovan G. Brankov, *Member, IEEE*, Nikolas P. Galatsanos, *Senior Member, IEEE*, Yongyi Yang, *Senior Member, IEEE*, and Miles N. Wernick, *Senior Member, IEEE*

***Abstract*—In this paper, we present a new approach for segmentation of image sequences by clustering the pixels according to their temporal behavior. The clustering metric we use is the normalized cross-correlation, also known as similarity. The main advantage of this metric is that, unlike the traditional Euclidean distance, it depends on the shape of the time signal rather than its amplitude. We model the intra-class variation among the time signals by a truncated exponential probability density distribution, and apply the expectation-maximization (EM) framework to derive two iterative clustering algorithms. Our numerical experiments using a simulated, dynamic PET brain study demonstrate that the proposed method achieves the best results when compared with several existing clustering methods.**

***Index Terms*—Clustering, dynamic PET, fMRI, image segmentation, similarity.**

I. Introduction

In time-sequence imaging modalities, such as dynamic PET and fMRI, an important problem is how to group the image pixels into spatial regions in which the pixels exhibit similar temporal behavior. This is useful, for example, in kinetic-modeling and functional neuroimaging applications.

The key issue for choosing an unsupervised clustering approach to this problem is to select a metric as a basis for determining class membership. In this paper, we propose to use the normalized cross-correlation coefficient (or *similarity*) between two signals as the metric.

The similarity metric is appropriate for our application because it compares the shapes of the time signals rather than their amplitudes. This is desirable in applications where the goal is to identify image regions consisting of pixels showing similar behavior, but not necessarily with uniform amplitude.

Many traditional clustering algorithms, such as the k-means algorithm [1] and Gaussian mixture approach [2], are based on Euclidean or Mahalanobis distance metrics. In the field of nuclear medicine, several algorithms for dynamic image segmentation have been proposed [3]–[5] based on these traditional clustering approaches. In recent work in [6], clustered component analysis (CCA) was developed for fMRI applications to partially remove the amplitude dependency of traditional clustering algorithms.

Manuscript received December 6, 2002; revised June 26, 2003. This work was supported in part by NIH/NINDS Grant NS35273 and NIH/NHLBI Grant HL65425.

The authors are with the Department of Electrical and Computer Engineering, Illinois Institute of Technology, Chicago, IL 60616 USA (e-mail: brankov@iit.edu; galatsanos@cs.uoi.gr; yy@ece.iit.edu; wernick@iit.edu).

Digital Object Identifier 10.1109/TNS.2003.817963

In the following sections, we introduce our data model and derive maximum-likelihood (ML) clustering algorithms using a generalized expectation-maximization (EM) framework [7]. We then describe evaluations of the proposed methods using simulated dynamic PET brain data.

II. Methods

A. Problem Formulation

Let us denote the time sequence at pixel n by a vector $\mathbf{Y}_n \in R^M$, $n = 1, 2, \cdots, N$, and let X_n denote the class label (i.e., region assignment) for $\mathbf{Y}_n$. Assuming the image is to be segmented into K classes (spatial regions), we describe the observed time sequences by the following model:

$$E[\mathbf{Y}_n | X_n = k] = \alpha_n \mathbf{e}_k \tag{1}$$

where α_n is the unknown amplitude of $\mathbf{Y}_n$, and $\mathbf{e}_k$ is a unit vector that defines the mean direction of class k within the space. The class labels X_n are assumed to be independent with unknown prior probabilities $P\{X_n = k\} = p_k$.

Given the data $\mathbf{Y}_n$, our objective is decide the class label X_n for each pixel. Since the parameters required to make this decision are unknown, we estimate them by a maximum-likelihood (ML) approach, according to the statistical model described in the following section.

B. Statistical Model

We begin by defining the following similarity metric that will form the basis for the clustering procedure:

$$\rho_n = \frac{\mathbf{Y}_n^T \mathbf{e}_{X_n}}{||\mathbf{Y}_n||}. \tag{2}$$

Geometrically, ρ_n is the cosine of the angle between $\mathbf{e}_{X_n}$ and $\mathbf{Y}_n$; thus, ρ_n is invariant to the magnitude of $\mathbf{Y}_n$, and ρ_n is close to 1 when the intra-class variation is small.

To implement our approach, we require a likelihood function for $\mathbf{Y}_n$. We begin by selecting a probability model for ρ_n, which relates to the variations in $\mathbf{Y}_n$. The variation in $\mathbf{Y}_n$ and, therefore, in ρ_n, results principally from physiological variations, but also includes the effect of imaging noise. Since the probability density function (pdf) representing these variations cannot be expressed exactly, we instead select a reasonable pdf model, which is standard practice in developing unsupervised clustering approaches.

To represent variability in the similarity metric ρ_n, we adopt the following truncated exponential pdf model:

$$p(\rho_n|A_n, X_n) = \frac{1}{\gamma_n} \exp\left[-A_n\left(1 - |\rho_n|\right)\right], |\rho_n| \leq 1 \quad (3)$$

where A_n is known as the concentration parameter (describing the degree of intra-class variation), and γ_n is a normalizing constant.

The pdf in (3) has been shown numerically [8] to approximate the pdf of the similarity metric under a Gaussian noise model. This pdf is also similar to the von Mises distribution [9], which is the analog of the normal distribution for angular data.

When the magnitude of the observation vector $\mathbf{Y}_n$ is large compared with its variability about a given class direction (i.e., when the signal-to-noise ratio is high), ρ_n will tend be close to unity. The pdf model in (3) reflects this fact, becoming narrower as the concentration parameter increases. Later, this interpretation of the concentration parameter will be used in the estimation procedure.

From the relation between ρ_n and $\mathbf{Y}_n$ in (2), we can immediately express the pdf in terms of $\mathbf{Y}_n$ as follows:

$$p(\mathbf{Y}_n|A_n, \mathbf{e}_{X_n}) = \frac{1}{\gamma_n} \exp\left[-A_n\left(1 - \frac{\left|\mathbf{Y}_n^T \mathbf{e}_{X_n}\right|}{||\mathbf{Y}_n||}\right)\right]. \quad (4)$$

C. ML Estimation by a Generalized EM Approach

We aim to estimate the model parameters by ML estimation; however, direct maximization of the likelihood function is intractable. Therefore, we develop generalized EM algorithms [7], [10] to find the solution iteratively.

To put the problem in an EM framework, we define the complete data as a concatenation of the observations $\mathbf{Y} = [\mathbf{Y}_1, \cdots, \mathbf{Y}_N]$ and their class labeling $\mathbf{X} = [X_1, X_2 \cdots X_N]$. The likelihood function for the complete data $[\mathbf{Y}, \mathbf{X}]$ is

$$p(\mathbf{Y}, \mathbf{X}|\boldsymbol{\theta}) = \prod_{n=1}^{N} p(\mathbf{Y}_n|X_n, \boldsymbol{\theta}) P(X_n|\boldsymbol{\theta}) \quad (5)$$

where $\boldsymbol{\theta} = [\mathbf{E}, \mathbf{P}, \mathbf{A}]$, in which $\mathbf{E} = [\mathbf{e}_1, \cdots, \mathbf{e}_K]$, $\mathbf{P} = [p_1, \cdots, p_K]$ and $\mathbf{A} = [A_1, \cdots, A_N]$.

Each iteration of the EM algorithm consists of maximization of the conditional expectation $\log[p(\mathbf{Y}, \mathbf{X}|\boldsymbol{\theta})|\mathbf{Y}, \boldsymbol{\theta}^i]$, where $\boldsymbol{\theta}^i$ denotes the parameter estimate at the i^{th} iteration. In the E-step, one computes

$$\mathcal{Q}(\boldsymbol{\theta}; \boldsymbol{\theta}^i) = E\left[\log p(\mathbf{Y}, \mathbf{X}|\boldsymbol{\theta})|\mathbf{Y}, \boldsymbol{\theta}^i\right];$$

in the M-step, the parameter estimate is updated as follows:

$$\boldsymbol{\theta}^{i+1} = \arg\max_{\boldsymbol{\theta}} \mathcal{Q}(\boldsymbol{\theta}; \boldsymbol{\theta}^i).$$

The EM algorithm starts with an initial estimate $\boldsymbol{\theta}^o$, then repeats the E- and M-steps until convergence, i.e., until $||\boldsymbol{\theta}^i - \boldsymbol{\theta}^{i+1}|| < \varepsilon$ for some ε.

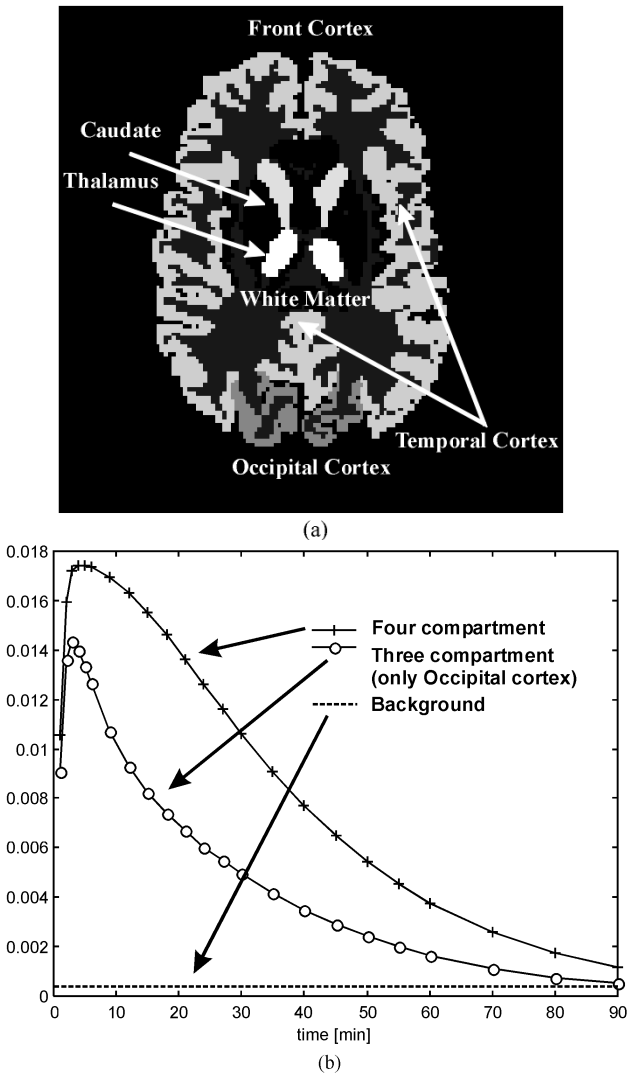

Fig. 1. Simulated data: (a) Zubal brain phantom and (b) time-activity curves.

Soft-Decision Similar Component Analysis: We develop two clustering algorithms, the first of which is called *soft-decision similar component analysis* (SCA). In the SCA algorithm, the E- and M-steps of the EM algorithm are as follows.

E-Step:

$$\begin{aligned}\mathcal{Q}(\boldsymbol{\theta}; \boldsymbol{\theta}^i) &= E\left[\log\left(p(\mathbf{Y}, \mathbf{X}|\boldsymbol{\theta})\right)|\mathbf{Y}, \boldsymbol{\theta}^i\right] \\ &= \sum_{n=1}^{N} E\left[\log p(\mathbf{Y}_n|X_n, \boldsymbol{\theta}) + \log P(X_n|\boldsymbol{\theta})|\mathbf{Y}_n, \boldsymbol{\theta}^i\right] \\ &= \sum_{n=1}^{N}\sum_{k=1}^{K} p(X_n = k|\mathbf{Y}_n, \boldsymbol{\theta}^i)\left[-\log(\gamma_n)\right. \\ &\quad \left. -A_n\left(1 - \left|\frac{\mathbf{Y}_n^{\mathbf{T}} \mathbf{e}_{X_n}}{||\mathbf{Y}_n||}\right|\right) + \log(p_k)\right] \end{aligned} \quad (6)$$

where

$$p(X_n = k|\mathbf{Y}_n, \boldsymbol{\theta}^i) = \frac{p(\mathbf{Y}_n|X_n = k, \boldsymbol{\theta}^i) P(X_n = k|\boldsymbol{\theta}^i)}{\sum_{k=1}^{K} p(\mathbf{Y}_n|X_n = k, \boldsymbol{\theta}^i) P(X_n = k|\boldsymbol{\theta}^i)}. \quad (7)$$

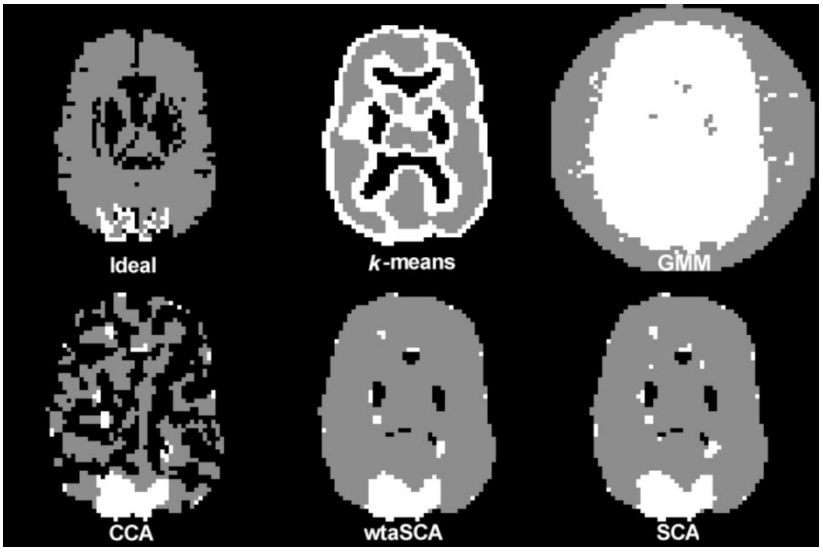

Fig. 2. Segmentation results by various methods. The number of assumed classes for all methods was three. The proposed wtaSCA and SCA methods produced the most accurate segmentations.

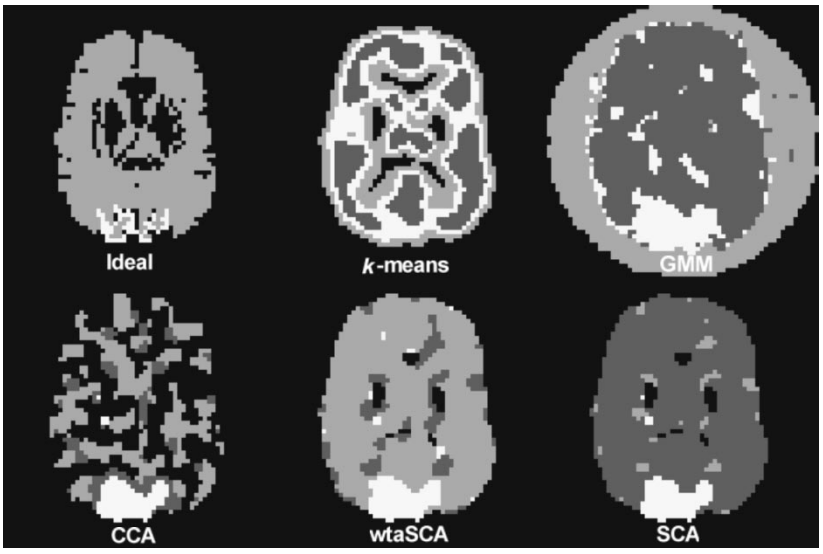

Fig. 3. Segmentation results by various methods. The wrong number of classes (four) was assumed for all the algorithms. Compared to other algorithms, the wtaSCA and SCA algorithms still performed reasonably well.

M-Step:

$$\left(p_k^{i+1}, \mathbf{e}_k^{i+1}\right) = \arg\max_{(p_k, \mathbf{e}_k)} \mathcal{Q}(\boldsymbol{\theta}; \boldsymbol{\theta}^i) \text{ subject to } \sum_{k=1}^{K} p_k = 1,\ p_k \geq 0,\ \text{and } \|\mathbf{e}_k\| = 1.$$

The maximization problems in the M-step can be solved by using Lagrange multipliers as follows:

$$\frac{\partial}{\partial p_k}\left[\mathcal{Q}(\boldsymbol{\theta}; \boldsymbol{\theta}^i) + \lambda\left(\sum_{k=1}^{K} p_k - 1\right)\right] = 0 \tag{8}$$

which leads to

$$p_k^{i+1} = \frac{\sum_{n=1}^{N} p(X_n = k|\mathbf{Y}_n, \boldsymbol{\theta}^i)}{N}. \tag{9}$$

Similarly

$$\mathbf{e}_k^{i+1} = \frac{1}{W}\sum_{n=1}^{N} p(X_n = k|\mathbf{Y}_n, \boldsymbol{\theta}^i)\frac{A_n^i}{\|\mathbf{Y}_n\|}\mathbf{Y}_n \tag{10}$$

where W is the normalization constant that makes $\|\mathbf{e}_k\| = 1$.

The calculation of the concentration parameter A_n is analytically intractable. As explained earlier, A_n can be viewed as a signal-to-noise ratio; therefore, we compute it as follows:

$$A_n^{i+1} = \frac{\sum_{k=1}^{K} p(X_n = k|\mathbf{Y}_n, \boldsymbol{\theta}^i)\left(\mathbf{Y}_n^{\mathbf{T}}\mathbf{e}_k\right)^2}{\frac{1}{M}\sum_{k=1}^{K} p(X_n = k|\mathbf{Y}_n, \boldsymbol{\theta}^i)\left\|\mathbf{Y}_n - \left(\mathbf{Y}_n^T\mathbf{e}_k\right)\mathbf{e}_k\right\|^2}. \tag{11}$$

The E- and M-steps in (7)–(11) are performed for a fixed number of iterations to obtain estimates of the parameters. In our experiments, the objective function $\mathcal{Q}(\boldsymbol{\theta}; \boldsymbol{\theta}^i)$ was found to be monotonically increasing at every iteration.

TABLE I
RELATIVE ACTIVITY LEVELS BY BRAIN REGION

Brain region	Three compartment	Four compartment
Thalamus	-	1
Caudate	-	0.87
Frontal Cortex	-	0.805
Ant. Temporal + Sup. Temporal Cortex	-	0.805
White matter	-	0.1
Occipital Cortex	1	-

Once the parameters have been estimated, we cluster the pixels according to a maximum *a posteriori* (Bayesian) decision strategy, i.e., we choose the class label for pixel n to be X_n^*, where

$$X_n^* = \arg\max_k p(X_n = k|\mathbf{Y}_n, \hat{\boldsymbol{\theta}}) \tag{12}$$

and $\hat{\boldsymbol{\theta}}$ is the final estimate of $\boldsymbol{\theta}$.

Winner-Take-All Similar Component Analysis: The second clustering algorithm, which we refer to as *winner-take-all similar component analysis* (wtaSCA), is similar to the SCA above. It is derived by letting $A_n \to \infty$. In this algorithm, we replace (7) by the following:

$$p(X_n = k|\mathbf{Y}_n, \boldsymbol{\theta}^i) = \begin{cases} 1, & k = \arg\max_j \frac{|\mathbf{Y}_n^T \mathbf{e}_j|}{\|\mathbf{Y}_n\|} \\ 0, & \text{otherwise} \end{cases}. \tag{13}$$

In summary, the wtaSCA algorithm consists of iterative computations of (13), (9), and (10). Note that a similar "winner-take-all" approach to forming the decision is used to derive the k-means method.

III. NUMERICAL EXPERIMENTS

A. Evaluation Data

In this study, to evaluate the performance of the proposed method, a single slice (no. 70) of the Zubal brain phantom [11] (see Fig. 1(a)) was used to simulate a dynamic PET study of [^{11}C] carfetanil binding to μ-selective opiate receptors. A four-compartment and a three-compartment tracer kinetic model were used to produce time-activity curves (TACs) for each brain region (Fig. 1(b)). The model parameters were derived from the data in [12], and an input plasma-concentration function obtained in an actual PET study conducted by the Department of Radiology at the University of Chicago. We simulated 23 image frames with a total Poisson mean of four million counts. The pixel size was 4.36 mm/pixel and the intrinsic blur was 8 mm in full width at half maximum (FWHM).

The reconstructed images contain regions characterized by three different types of TACs. These regions are indicated in Fig. 2 and 3 (labeled "Ideal"). The regions of interest are: (a) background and ventricles, represented by black; (b) areas having TACs that are similar in shape, but differing significantly in amplitude (represented by gray, including thalamus, caudate, frontal cortex, temporal cortex, and white matter; see Table I for amplitudes); and (c) occipital cortex (represented by white). More details about the model can be found in [13].

Prior to clustering, in order to reduce the noise level, a low-pass filter with cutoff frequency of 0.5 cycles/pixel was applied.

TABLE II
PERCENTAGE OF PIXELS CORRECTLY CLASSIFIED (THREE CLASSES ASSUMED)

	Occipital Cortex	Other brain regions	Background	Average
k-means	39.9%	55.6%	92.4%	62.6%
GMM	0.0%	100.0%	37.8%	45.9%
CCA	100.0%	44.9%	93.4%	79.4%
wtaSCA	100.0%	93.6%	87.6%	93.7%
SCA	100.0%	94.3%	87.6%	94.0%

TABLE III
PERCENTAGE OF PIXELS CORRECTLY CLASSIFIED (FOUR CLASSES ASSUMED)

	Occipital Cortex	Other brain regions	Background	Average
k-means	61.80%	42.60%	89.89%	64.76%
GMM	100.0%	75.72%	40.78%	72.19%
CCA	96.40%	34.80%	91.25%	74.15%
wtaSCA	100.0%	77.02%	87.6%	88.22%
SCA	100.0%	87.02%	87.6%	91.55%

B. Other Methods Considered

In addition to the two proposed clustering algorithms, we also considered three well-known clustering procedures for comparison: a) Gaussian mixture model (GMM) parameter estimation [10]; b) k-means (also called the Linde-Buzo-Gray (LBG) algorithm [14] or generalized Lloyd algorithm), which is a winner-take-all equivalent of GMM; and c) clustered component analysis (CCA) proposed by Bouman *et al.* [6]. The CCA can be viewed as a special case of a probabilistic principal component analysis (PPCA) mixture model [15].

C. Results

Two experiments were performed to test the proposed clustering algorithms. In the first experiment, the number of classes was correctly assumed to be three. The clustering results obtained by different algorithms are shown in Fig. 2. In the second experiment, the number of classes was assumed incorrectly to be four, the purpose being to test the robustness of the proposed methods. The clustering results are shown in Fig. 3.

In addition, we show in Tables II and III the rates of correct classification obtained by the various algorithms for both experiments. As shown, the proposed wtaSCA and SCA significantly outperform the other methods.

One can see that in the second experiment the k-means algorithm failed to correctly recognize any of the regions. The GMM and CCA algorithms correctly classified the occipital cortex area, but failed to discriminate the other regions correctly.

It is also worth noting that the wtaSCA algorithm achieves good performance with very low computational complexity. The SCA algorithm is computationally more complex than wtaSCA,

TABLE IV
COMPUTATION TIME (RELATIVE TO wtaSCA)

	One iteration	Full convergence
k-means	1.77	1.8
GMM	187.0	266.0
CCA	8.5	5.4
wtaSCA	1	1
SCA	1.06	1.16

but less complex than the CCA and GMM algorithms. The CCA approach requires the computation of eigenvectors of an $M \times M$ data correlation matrix, while the GMM approach requires computation of the inverse of the covariance matrix. The computation times (on Pentium IV, 2.3 GHz) are shown in Table IV. All times are given relative to wtaSCA method.

IV. CONCLUSION AND FUTURE WORK

Our experimental results demonstrate the feasibility of the proposed similarity-based clustering methods. The results indicate the ability of the proposed algorithms to correctly identify regions having distinct TACs. Among the methods tested, the proposed algorithms produced the best accuracy with the lowest computational complexity in our experiments. We anticipate that the proposed algorithms will be useful for automated kinetic-parameter estimation. In the future, we plan to augment this work by testing criteria that will allow automatic determination of the number of classes such as minimum description length (MDL). Furthermore, the incorporation of priors that constrain pixels of the same class to be spatially adjacent also will be investigated [16].

REFERENCES

[1] R. A. Johnson and D. W. Wichern, *Applied Multivariate Statistical Analysis*. Englewood Cliffs, NJ: Prentice-Hall, 1992.
[2] G. J. McLachlan and T. Krishnan, *The EM Algorithm and Extensions*. New York: Wiley, 1997.
[3] K. P. Wong, D. Feng, S. R. Meikle, and M. J. Fulham, "Segmentation of dynamic PET images using cluster analysis," *IEEE Trans. Nucl. Sci.*, vol. 49, no. 1, pp. 200–207, Feb. 2002.
[4] Y. Hode, A. Deruyver, B. Bendriem, and N. Volkow, "Temporal image fusion," in *Proc. IEEE Int. Conf. Image Processing*, vol. 2, 1995, pp. 472–475.
[5] F. O'Sullivan, "Imaging radiotracer model parameters in PET: a mixture analysis approach," *IEEE Trans. Med. Imaging*, vol. 12, pp. 399–412, Sept. 1993.
[6] C. A. Bouman, S. Chen, and M. J. Lowe, "Clustered component analysis for fMRI signal estimation and classification," in *Proc. IEEE Int. Conf. Image Processing*, vol. 1, 2000, pp. 609–612.
[7] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *J. Roy. Statist. Sect.*, vol. 39, pp. 1–38, 1977.
[8] G. S. Watson, *Statistics on Spheres*. New York: Wiley, 1983, pp. 109–119.
[9] K. V. Mardia and P. E. Jupp, *Directional Statistics*. New York: Wiley, 1999.
[10] D. M. Titterington, A. F. M. Smith, and U. E. Makov, *Statistical Analysis of Finite Mixture Distributions*. New York: Wiley, 1985.
[11] I. G. Zubal, C. R. Harrell, E. O. Smith, Z. Rattner, G. R. Gindi, and P. B. Hoffer, "Computerized three-dimensional segmented human anatomy," *Med. Phys.*, vol. 21, pp. 299–302, 1994.
[12] J. J. Frost, K. H. Douglass, H. S. Dannals, J. M. Links, A. A. Wilson, H. T. Ravert, W. C. Crozier, and H. N. Wagner Jr., "Multicompartmental analysis of $[^{11}C]$-carfentanil binding to opiate receptors in humans measured by positron emission tomography," *J. Cereb. Blood Flow Metab.*, vol. 9, pp. 398–409, 1989.
[13] J. G. Brankov, "Tomographic Image Reconstruction for Partially-Known Systems and Image Sequences," M.S. thesis, Illinois Inst. Technol., Chicago, IL, 1999.
[14] Y. Linde, A. Buzo, and R. M. Gray, "An algorithm for vector quantiser design," *IEEE Trans. Commun.*, vol. COM-28, pp. 84–95, Jan. 1980.
[15] M. E. Tipping and C. M. Bishop, "Mixture of probabilistic principal component analysis," *Neural Computing*, vol. 11, no. 2, pp. 443–482, 1998.
[16] Y. Zhang, M. Brady, and S. Smith, "Segmentation of brain MR images through hidden Markov random field model and the expectation maximization algorithm," *IEEE Trans. Med. Imaging*, vol. 20, pp. 45–57, Jan. 2001.