International Journal on Artificial Intelligence Tools

A regression mixture model with spatial constraints for clustering spatiotemporal data

K. Blekas, C. Nikou, N. Galatsanos
Department of Computer Science, University of Ioannina
45110 Ioannina, Greece
{kblekas,cnikou,galatsanos}@cs.uoi.gr

N. V. Tsekos
School of Medicine, Washington University in St. Louis
Whashington, USA
tsekosn@mir.wustl.edu

Received (Day Month Year)
Revised (Day Month Year)
Accepted (Day Month Year)

We present a new approach for curve clustering designed for analysis of spatiotemporal data. Such data contains both spatial and temporal patterns that we desire to capture. The proposed methodology is based on regression and Gaussian mixture modeling. The novelty of the herein work is the incorporation of spatial smoothness constraints in the form of a prior for the data labels. This allows to take into account the property of spatiotemporal data according to which spatially adjacent data points have higher probability to belong to the same cluster. The proposed model can be formulated as a Maximum a Posteriori (MAP) problem, where the Expectation Maximization (EM) algorithm is used to estimate the model parameters. Several numerical experiments with both simulated data and real cardiac perfusion MRI data are used for evaluating the methodology. The results are promising and demonstrate the value of the proposed approach.

Keywords: Regression mixture model; curve clustering; Expectation Maximization; Markov random field; smoothness prior.

1. Introduction

Clustering is a very interesting and challenging research problem and a wide spectrum of methodologies has been used to address it. Probabilistic mixture modeling is a well established model-based approach for clustering that offers many advantages. One such advantage is that it provides a natural platform to evaluate the quality of the clustering solution [1], [2]. Curve clustering is a special case of clustering in which the available data have one or both of the following two features: first they are of very large dimension and thus conventional clustering methods are computationally prohibitive, and second they are not of equal length and thus conventional clustering methods cannot straightforwardly be applied. In such cases it is natural

initially to fit the available data with a parametric model and then to cluster based on that model. Different types of functional models have been used to for such data. Among them polynomial and spline regression are the most commonly used models [3] and have been successfully applied to a number of diverse applications, ranging from gene clustering in bioinformatics to clustering of cyclone trajectories, see for example [4] [5] and [6].

Another important application of curve clustering is in the analysis of spatiotemporal data where it is desired to capture both the spatial and temporal patterns of the data. For example, in medical imaging modalities, such as dynamic PET and functional MRI, an important problem is how to group image pixels into spatial regions in which the pixels exhibit similar temporal behavior. This is very useful, for example, in kinetic-modeling and functional imaging applications, see for example [7], [8], and the references therein. In such studies it is important to measure both the temporal characteristics of the grouped pixels and simultaneously to accurately classify the pixels into groups of similar temporal behavior.

Thus, to determine class membership in this type of data, apart from the distance between the coefficients of the model, it is also beneficial to use spatial information. These constraints capture the prior knowledge that adjacent pixels most likely belong to the same class. The idea of combining Gaussian mixture models with spatial smoothness prior has been used previously with success for segmentation of natural images [9] and [10].

In this paper we extend this idea to the problem of time-sequence analysis via regression based curve clustering. In other words, we fit the curve data with a regression mixture model and set a Markov random field (MRF) smoothness prior over the class labels of the data in order to achieve data belong to the same neighborhood to have the same labels. Then, a maximum a posteriori expectation maximization algorithm (MAP-EM) [11], [2] is applied to learn this model and cluster the data. We also estimate the number of components of the mixture model, and therefore the number of clusters, using the Bayesian information criterion (BIC) [12]. Finally, the performance of the proposed methodology is evaluated using a series of experiments with simulated data generated with different polynomial or splines functional in noisy environments. Since the ground truth is known in such data, we compared the simple regression mixture model without the smoothness prior with the one that uses it. As performance metrics we used the percentage of misclassified data and the error in the fit of the original curves. We also applied this method to in vivo dynamic cardiac MRI data with promising results as evaluated by human experts.

More specifically, in section 2 we present the simple regression model, the proposed method based on the smoothness prior and the EM methodology used for estimating model's parameters. To assess the performance of the proposed methodology we present in section 3 numerical experiments with both artificial data where the ground truth is known and real cardiac perfusion MRI data. Finally, in section 4 we give our conclusions and suggestions for future research.

2. Model and prior specification

2.1. *Regression mixture models*

Suppose the spatiotemporal data $Y = \{y_{il}\}_{i=1,\ldots,N}^{l=1,\ldots,T}$, where i denotes the spatial index and l the temporal index that corresponds to time locations t_l. This kind of data consists of T images each with N pixels. Thus, at each pixel location i we have the temporal sequence y_i of length T. It must be noted that although during the present description of the regression model it is assumed that all y_i sequences are of equal length, this can be easily changed. In such case, each y_i for $i = 1, \ldots, N$ is of variable length T_i. This corresponds to the general case of the model that will be demonstrated later in our numerical experiments section.

To model curves y_i we use p-order polynomial regression on the time range $t = (t_1, \ldots, t_T)$ with an additive noise term given by

$$y_i = X\beta + e_i \ , \tag{1}$$

where X is the Vandermonde matrix, i.e.

$$X = \begin{pmatrix} 1 & t_1 & \ldots & t_1^p \\ \vdots & \vdots & \ldots & \vdots \\ 1 & t_T & \ldots & t_T^p \end{pmatrix}$$

and β is the $p+1$-vector of regression coefficients. Finally, the error term e_i is a T-dimensional vector that is assumed to be Gaussian and independent over time, i.e. $e_i \sim \mathcal{N}(0, \Sigma)$ with a diagonal covariance matrix $\Sigma = diag(\sigma_1^2, \ldots, \sigma_T^2)$. Thus, by assuming $X\beta$ deterministic, we can model the joint probability density of the curve y with the normal distribution $\mathcal{N}(X\beta, \Sigma)$.

In this study we consider the problem of curve clustering, i.e. the division of the set of curves y_i with $i = 1, \ldots N$ into K clusters, where each cluster will contain curves of the same generation mechanism (polynomial regression model). To this direction, the regression mixture model is a useful generative model that can be used to capture heterogeneous sources of curves. This can be described by the following probability density function:

$$f(y_i|\Theta) = \sum_{j=1}^{K} \pi_j p(y_i|\theta_j) \ , \tag{2}$$

which has a generic and powerful meaning in model-based clustering. Following this scheme, each curve is generated by first selecting a source j (cluster) according to probabilities π_j and then by performing sampling based on the corresponding regression relationship with parameters $\theta_j = \{\beta_j, \Sigma_j\}$ as described by the normal density function $p(y_i|\theta_j) = \mathcal{N}(X\beta_j, \Sigma_j)$. Moreover, the unknown mixture probabilities satisfy the constraints: $\pi_j \geq 0$ and $\sum_{j=1}^{K} \pi_j = 1$.

Based on the above formulation, the clustering problem becomes a maximum likelihood (ML) estimation problem for the mixture parameters $\Theta = \{\pi_j, \theta_j\}_{j=1}^{K}$,

where the log-likelihood function is given by

$$L(Y|\Theta) = \sum_{i=1}^{N} \log\{\sum_{j=1}^{K} \pi_j p(y_i|\theta_j)\} \ . \tag{3}$$

The Expectation-Maximization (EM) algorithm [11] is an efficient framework for solving likelihood estimation problems for mixture models. It performs iteratively two steps: The E-step, where the current posterior probabilities of samples to belong to each cluster are calculated:

$$z_{ij}^{(t)} = P(j|y_i, \Theta^{(t)}) = \frac{\pi_j^{(t)} p(y_i|\theta_j^{(t)})}{f(y_i|\Theta^{(t)})} \ , \tag{4}$$

and the M-step, where the maximization of the expected value of the complete log-likelihood is performed. This leads to the following updated rules for the mixture parameters [4], [3]:

$$\pi_j^{(t+1)} = \frac{\sum_{i=1}^{N} z_{ij}^{(t)}}{N} \ , \tag{5}$$

$$\beta_j^{(t+1)} = \left[\sum_{i=1}^{N} z_{ij}^{(t)} X^T {\Sigma_j^{-1}}^{(t)} X\right]^{-1} X^T {\Sigma_j^{-1}}^{(t)} \sum_{i=1}^{N} z_{ij}^{(t)} y_i \ , \tag{6}$$

$${\sigma_{jl}^2}^{(t+1)} = \frac{\sum_{i=1}^{N} z_{ij}^{(t)} (y_{il} - [X\beta_j^{(t+1)}]_l)^2}{\sum_{i=1}^{N} z_{ij}^{(t)}} \ , \tag{7}$$

where $[.]_l$ indicates the l-th component of the T-dimensional vector that corresponds to location t_l. After convergence of the EM, the association of the N observable curves y_i with the K clusters is based on the maximum value of the posterior probabilities. The generative polynomial regression function is also obtained per each cluster, as expressed by the $(p+1)$-dimensional vectors of the regression coefficients β_j.

2.2. *The spatially variant regression mixture with smoothness prior*

In order to enforce spatial smoothness to the basic scheme of the regression mixture model we use a spatially varying approach, see for example [9]. This model, unlike the classical, assumes that the probabilities of the data labels π_{ij} are random variables, where i defines the spatial location and j the class. To handle this information we use a Markov random field (MRF) prior [13],[14], [15] that provides a convenient way of modeling the constraint in many computer vision and image processing problems,

i.e. that the probability of a node in the image field depends only on its neighboring nodes.

In particular, we assume that the mixture density function is given by the following equation

$$f(y_i|\Theta) = \sum_{j=1}^{K} \pi_{ij} p(y_i|\theta_j) \ , \tag{8}$$

where the mixture parameters are $\Theta = \{\{\pi_{ij}\}_{i=1}^{N}, \theta_j\}_{j=1}^{K}$. The probabilities of the pixel labels $\pi = \{\pi_{ij}\}$ satisfy the constraints: $\pi_{ij} \geq 0$, $\sum_j \pi_{ij} = 1$ and they follow the Gibbs distribution with a density function given by

$$p(\pi) = \frac{1}{Z} \exp(-\frac{\sum_{i=1}^{N} V_{\mathcal{N}_i}(\pi)}{\xi}) \ , \tag{9}$$

where Z is a normalizing constant, while ξ is the regularization parameter. The function $V_{\mathcal{N}_i}(\pi)$ denotes the clique potential function of the pixel label vectors $\{\pi_m\}$ within the neighborhood $\mathcal{N}_i$ of the ith-pixel and can be computed as follows

$$V_{\mathcal{N}_i}(\pi) = \sum_{m \in \mathcal{N}_i} \sum_{j=1}^{K} (\pi_{ij} - \pi_{mj})^2 \ . \tag{10}$$

The neighborhood $\mathcal{N}_i$ is the set of adjacent pixels m around pixel i ($|\mathcal{N}_i| = 8$ in the general case).

As explained in [10], it is advantageous in the above formulation to use a Gaussian-MRF with a different variance ξ_j at each cluster. Then, this prior is given by

$$p(\pi) \propto \prod_{j=1}^{K} \xi_j^{-N} \exp(-\frac{\sum_{i=1}^{N} \sum_{m \in \mathcal{N}_i} (\pi_{ij} - \pi_{mj})^2}{2\xi_j^2}) \ . \tag{11}$$

The advantages from this type of prior are twofold. First, the parameter ξ_j that capture spatial attributes enforce smoothness of different degree at each cluster and better adapt to the data. Also, as will be showed later, this prior allows the estimation of the values of the parameters ξ_j directly from data.

Using this prior the log likelihood of the MAP function is

$$L_{MAP}(\Theta|Y) = \log p(Y|\Theta) + \log p(\Theta) = \sum_{i=1}^{N} \log\{\sum_{j=1}^{K} \pi_{ij} p(y_i|\theta_j)\} + \log p(\pi) \ . \tag{12}$$

Direct maximization of this function is difficult thus we resort to the EM methodology. In particular, during the E-step the posterior probabilities values are calculated

$$z_{ij}^{(t)} = P(j|y_i, \Theta^{(t)}) = \frac{\pi_{ij}^{(t)} p(y_i|\theta_j^{(t)})}{\sum_{j'=1}^{K} \pi_{ij'} p(y_i|\theta_j)} \ , \tag{13}$$

while the expected value of the MAP log likelihood of the complete data

$$Q(\Theta|\Theta^{(t)}) = \sum_{i=1}^{N}\sum_{j=1}^{K} z_{ij}^{(t)}\{\log \pi_{ij} + \log p(y_i|\theta_j)\} - \log \xi_j - \frac{\sum\limits_{m\in\mathcal{N}_i}(\pi_{ij}-\pi_{mj})^2}{2\xi_j^2} \ . \quad (14)$$

is maximized next during the M-step. The update rules for the regression parameters β_j and Σ_j (Eqs. 6 and 7, respectively) are exactly the same as in the case of ML configuration. However, inference of the probabilities of the pixel labels π_{ij} is not as straightforward. Setting the derivative of (14) with respect to parameters π_{ij} equal to zero we take the following quadratic equation:

$$\pi_{ij}^2 - \widetilde{\pi}_{ij}\pi_{ij} - \frac{\xi_j^2}{|\mathcal{N}_i|} z_{ij}^{(t)} = 0 \ , \quad (15)$$

where $\widetilde{\pi}_{ij} = \frac{1}{|\mathcal{N}_i|}\sum\limits_{m\in\mathcal{N}_i}\pi_{mj}$ is the mean value of the j-th cluster's probability of the spatial neighbors of the i-th pixel. The above quadratic expression has two roots, where we select only the root with the positive sign since it yields the constraint $\pi_{ij} \geq 0$:

$$\pi_{ij}^{(t+1)} = \frac{\widetilde{\pi}_{ij} + \sqrt{\widetilde{\pi}_{ij}^2 + 4\frac{\xi_j^2}{|\mathcal{N}_i|} z_{ij}^{(t)}}}{2} \ . \quad (16)$$

It must be noted that in the above equation the neighborhood $\mathcal{N}_i$ may include pixels with probabilities either updated ($\pi_{mj}^{(t+1)}$) or not ($\pi_{mj}^{(t)}$). However, these values of π_{ij} as computed by Eq. 16 are not the final solution since they do not satisfy the constraints $0 \leq \pi_{ij} \leq 1$ and $\sum_{j=1}^{K}\pi_{ij} = 1$. These constraint equations define a convex hull. Thus, after calculation of $\pi_{ij}^{(t+1)}$ (Eq. 16) we project them on the constraint convex hull. For this projection an efficient convex quadratic programming algorithm presented in [9] is used.

Finally, by taking the derivative of Q-function (Eq. 14) with respect to the smoothness parameters ξ_j we obtain the following update rule:

$$\xi_j^{2(t+1)} = \frac{1}{N}\sum_{i=1}^{N}\sum_{m\in\mathcal{N}_i}(\pi_{ij}^{(t+1)} - \pi_{mj}^{(t+1)})^2 \ . \quad (17)$$

As discussed earlier, the fact that MRF variances ξ_j^2 can be found in closed form is another advantage of the proposed prior inference framework.

2.3. *Model Selection*

The problem of selecting a statistical model with the correct complexity is fundamental in statistical modeling. In such cases one has to tradeoff between fitting accurately the data and the ability of the model to generalize. For the herein used mixture model the number of components K determines the complexity. In order

to address this problem models of varying order along with the Bayesian information criterion (BIC), has been used successfully in a number of applications; see for example [16], [12] and [17]. The BIC is given by

$$BIC = -2L(\hat{\Theta}_k) + D \times log(N) \ , \tag{18}$$

where $\hat{\Theta}_k$ is the $k-$order estimated model. The quantity D is the total number of model parameters. In our case, $D = N(K-1)+K(p+1)+KT = K(N+p+T+1)-N$, since there are three kinds of parameters, π_{ij}, β_j, σ_{jl}, and the constraint $\sum_j \pi_{ij} = 1$, $\forall i$. In other words BIC includes a penalty term along with the negative log-likelihood that depends on the number of model parameters. This term penalizes complex models with many parameters and thus counterbalances the negative log-likelihood term that increases monotonically with the number of parameters.

3. Experimental results

The performance of the proposed regression mixture model with smoothness constraints, referred to as *spatial RM*, is evaluated using a number of numerical experiments. We have considered both simulated spatiotemporal data with known ground truth, as well as real cardiac perfusion MRI sequences along with visual inspection by an expert. The herein *spatial RM* is compared with the simple regression mixture model without spatial smootheness, referred to as *simple RM*.

Both models were initialized using the same strategy. In particular, K curves from the data set were randomly selected and a simple least-square fit was used to obtain the $K \times (p+1)$ regression coefficients. The mixing parameters were selected as $\frac{1}{K}$ and the covariances were assumed spherical and equal to $\frac{1}{K}$ times the total variance of the data set. Then, one step of the EM was used to further refine these parameters. Based on these parameters the log-likelihood was evaluated. The parameters used for initialization were selected as the ones that gave the best value of the log-likelihood criterion after 200 random trials. Finally, it must be noted that during all experiments we have used polynomials of order $p = 8$.

3.1. *Experiments with simulated spatiotemporal data*

At first the performance of our method was evaluated with simulated spatiotemporal data with known ground truth. In particular, we used two piece-wise constant images in Fig. 1 that contain three ($K = 3$) and five ($K = 5$) distinct regions, respectively [15]. Simulated image sequences with spatial resolution 128×128 pixels resulting in $N = 16,384$ and $T = 50$ time frames were generated by varying the intensity of each region according to a predefined time pattern. Four different sets of such patterns were used in our study, namely $C1$, $C2$, $C3$ and $C4$, shown in Fig. 2. The first set $C1$ consists of a Gaussian, a Rayleigh and a logarithmic curve, the second set $C2$ contains more complex curves obtained by polynomials of degree $p = 6$, the third set $C3$ consists of a sinusoid and two 2-component Gaussian mixtures, one with a discontinuity adopted from [18], and finally the fourth set $C4$ contains the

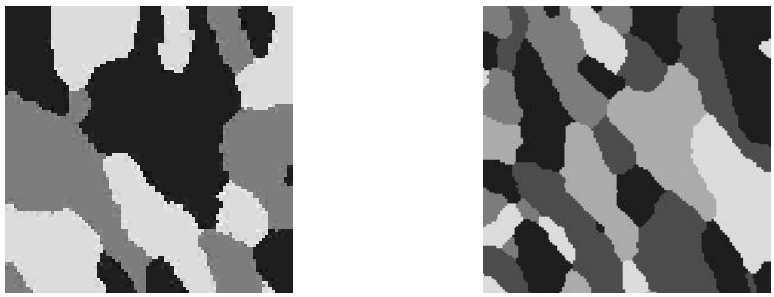


Fig. 1. The two test images (with $K = 3$ and $K = 5$ classes, respectively) used in the experiments.

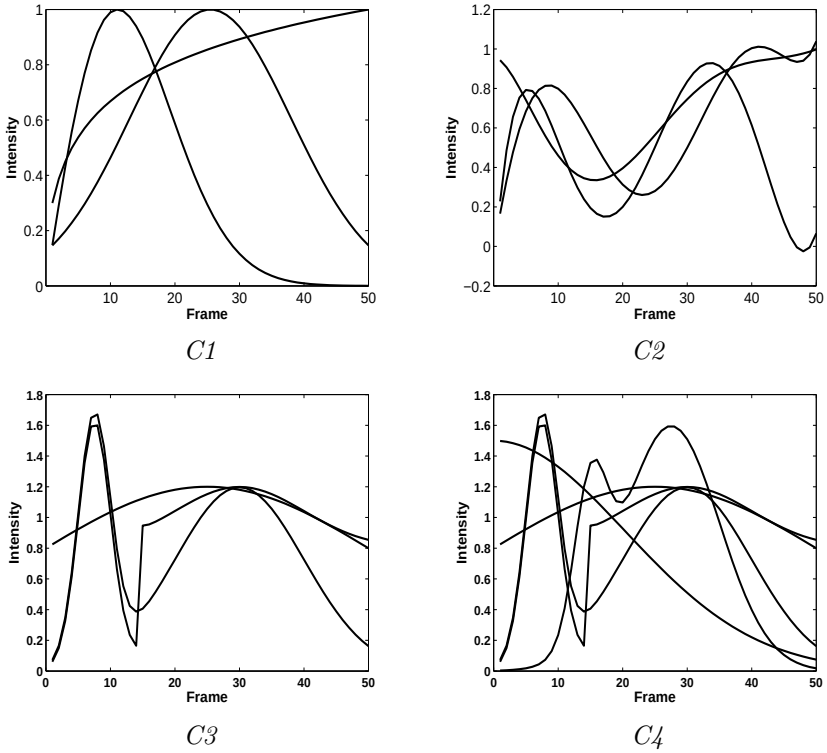


Fig. 2. Four different sets of the signal patterns used for generating the image sequences of the piece-wise constant images in Fig. 1. The configurations *C1, C2, C3* correspond to the case with $K = 3$ segments, while the pattern $C4$ has used for the case with $K = 5$ clusters.

three curves of the set $C3$ along with a Gaussian and an exponential function. The generated sequences of frames were also corrupted by additive white Gaussian noise

with variance selected to yield signal-to-noise-ratio (SNR) between $8dB$ and $-8dB$. Some characteristic examples of the obtained frames are shown in Fig.3.

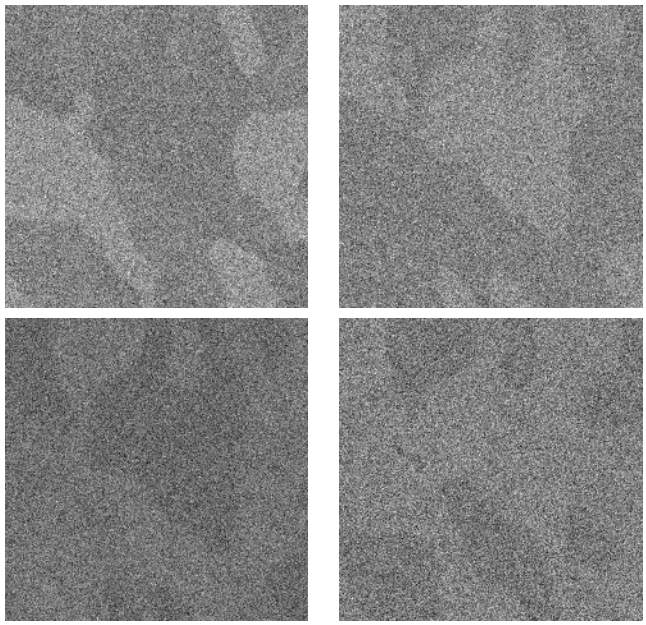


Fig. 3. Sample frames from the simulated sequence $C2$ with $SNR = 0dB$.

To quantify the performance of the proposed method two evaluation criteria were used: a) The percentage of correctly classified pixels that quantifies the ability of our methodology to capture the spatial patterns of the data. b) The curve square error CSE, that is the sum of squared errors between real curves ($\{r_{jl}\}$) and the estimated curves ($\{[X\hat{\beta}_j]_l\}$), i.e.

$$CSE = \sum_{j=1}^{K} CSE_j \quad \text{, where } CSE_j = \sum_{l=1}^{T} (r_{jl} - [X\hat{\beta}_j]_l)^2,$$

was used to quantify the ability of the proposed methodology to capture the temporal patterns of the data. For each problem we performed 50 runs of both methods with different noise realizations and evaluated the mean value and the standard deviation of the above two performance criteria.

Figure 4 presents the evolution of these quantities as a function of the SNR for the four sets of sequences based on the temporal patterns of Fig. 2. As can be observed, the proposed methodology improves significantly classification accuracy as compared to the *simple RM* for the entire range of examined SNR, especially for the lower values. For this emperiments in all four data sets the introduction of spatial smoothness does not improve its capability to estimate the temporal patterns. The

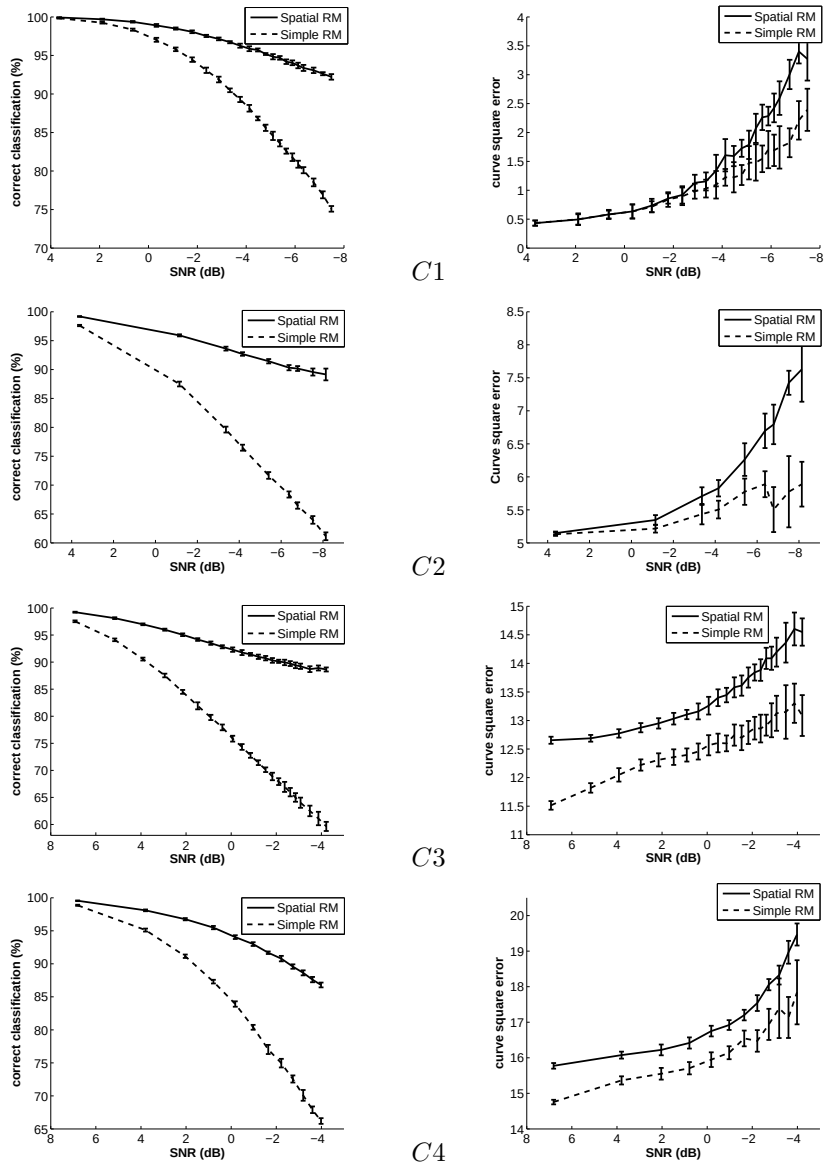


Fig. 4. Percentage of correct classification and curve square error as a function of the SNR for the four sets of signal patterns. We have $C1, C2$ and $C3$ with $K = 3$ and $C4$ with $K = 5$.

calculated CSE was always slightly worse for the model with the spatial constraints.

This was expected, since it is well known that in such regression problems spatial smoothness although reduces the variance at the expense of introducing bias to the estimates of the parameters [19]. Both the variance and the bias contribute to mean-square-error like metrics as the herein used CSE. The smoothness constraint is very

beneficial in cases where the problem is ill-posed, i.e. when the variance of the parameters estimated based only on the data is large. On the opposite side, in cases where the number of data points is quite large as during the above experiments ($N = 16,384$), the variance of the estimates of the regression parameters becomes smaller. Thus, the smoothness constraint, although it improves spatial segmentation, does not help reduce CSE [19]. In what follows we show experiments with missing data where the value of the smoothness constraint becomes apparent for the estimation of the temporal patterns also.

3.1.1. *Experiments with missing data*

An important application of this methodology is in situations with missing data. In such cases for certain time instances there are missing spatial measurements. Such a situations can be found in many applications. For example, in tracking from video sequences, occlusions that block the view of the objects in certain image frames result in missing data. Furthermore, isolated spiking on the radio frequency coils can cause variable temporal and spatial missing data in functional MRI. In such cases use of regression mixtures is very beneficial since other conventional clustering methodologies cannot be used in a straightforward manner. For this case each temporal pattern is of variable length, i.e. $y_i = \{y_{il}\}_{l=1}^{T_i}$. Then, the polynomial regression model of y_i is given by

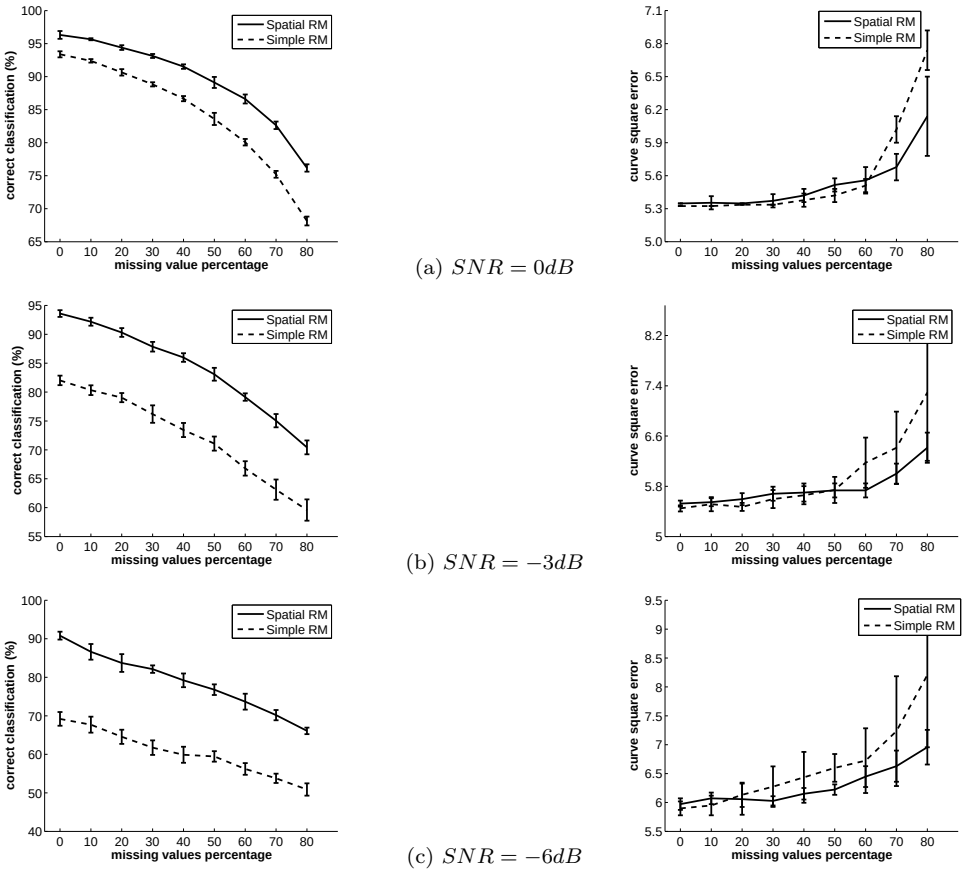
$$y_i = X_i \beta + e_i \ , \tag{19}$$

where the time regressor X_i is a Vandermonde matrix with a variable number rows which is equal to existing time measurements at each spatial location. The use of the EM algorithm for estimating the regression parameters differs only in the M-step in Eq. (6) and is given by:

$$\beta_j^{(t+1)} = \left[\sum_{i=1}^{N} z_{ij}^{(t)} X_i^T {\Sigma_j^{-1}}^{(t)} X_i \right]^{-1} \sum_{i=1}^{N} z_{ij}^{(t)} X_i^T {\Sigma_j^{-1}}^{(t)} y_i \ . \tag{20}$$

Additional experiments have been conducted in an attempt to study the behavior of our methodology in this case. Since the removed samples are chosen at randomly, no spatial location is privileged with respect to any other and the percentage of the missing data expresses the probability that a time sample may not be present at any spatial location.

Experiments were conducted on the second signal pattern $C2$ by varying the percentage of missing data between 0% and 80% and (again) for various SNR levels. Figure 5 illustrates the mean values and standard deviations of the two evaluation criteria for three such SNR values. As expected, the classification accuracy of the *spatial RM* model again is superior in comparison with the *simple RM* model for all levels of missing data. However, in this study a significant improvement of the CSE criterion was also observed in noisy environments especially when the percentage missing data was large. This is in agreement with our previous explanation suggested



(a) $SNR = 0dB$

(b) $SNR = -3dB$

(c) $SNR = -6dB$

Fig. 5. Comparative results in terms of two evaluation criteria in various levels of missing data.

that the benefits of the smoothness prior are more apparent in cases where the available data alone cannot estimate accurately the model parameters.

3.1.2. *Experiments with BIC model selection*

In order to test the value of the BIC-based model selection approach, we performed additional experiments. In particular, for every set of signal patterns we created 50 different datasets by using a different noise realization. Then we measured the number of components K found based on the BIC model selection criterion. Figure 6 presents the relative frequencies plots of the values of K found for the datasets $C3$ and $C4$ for two different noise levels. It is clear that the number of clusters K in high noise environments can be accurately deduced most of the times. However, as the value of SNR decreases this becomes harder as observed by the widening of the relative frequency plots.

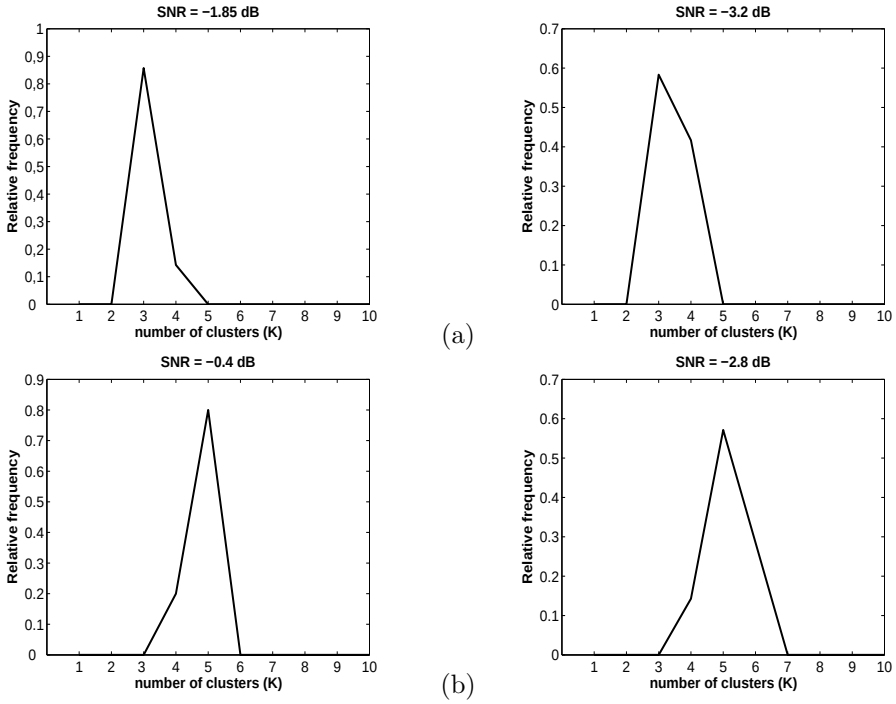


Fig. 6. Relative frequencies of the number of clusters found by applying the BIC model selection method to the regression mixture model in two noise values for the simulated $C3$ (a) and $C4$ (b) datasets of Fig. 2.

3.2. *Cardiac perfusion MRI sequence segmentation*

We also tested our method on two *real sets* of spatiotemporal data from in vivo dynamic cardiac magnetic resonance imaging (MRI) studies. Our goal was to segment the anatomies of the heart based on their hemodynamic coherence, as well as to measure the time behavior of each segment. The MRI data was obtained from an instrumented pig with an intracoronary catheter inserted into the left main coronary artery. Perfusion images were acquired with a heavily T1-weighted "contrast enhanced with saturations and multiple inversions" (CESMIR) prepared fast gradient recalled echo sequence [20], with TR/TE/a = 2.2/1.2/20; FOV = 200 X 200 mm2; slice = 5 mm; matrix = 96×96; and central phase-encoding order. Dynamic imaging was performed by acquiring a series of 100 to 200 2D MR images aligned along an oblique long axis views to image the left ventricle (LV) from the base to the apex of the heart [21]. After 15-20 precontrast images, 2 mL of 0.125 mM Gd-DTPA (Omniscan; Amersham Health, Princeton, NJ, USA) were administered at 2 mL/second with a power injector followed by a normal saline flush.

Figure 7 illustrates representative MR images from this sequence, illustrating the differential enhancement of the left circumflex (LCx) and left anterior descending

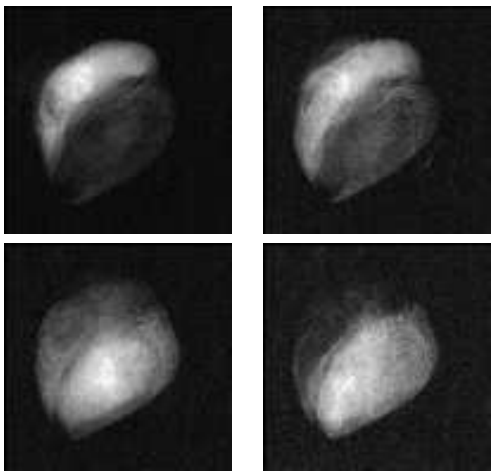
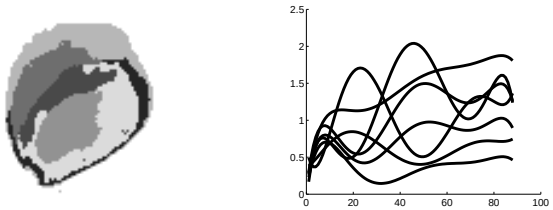
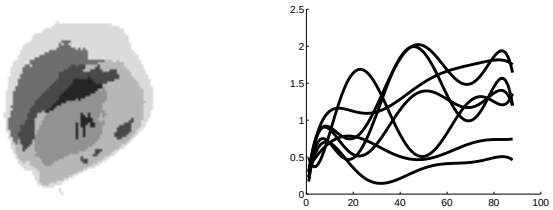


Fig. 7. Sample frames from the cardiac perfusion MRI sequence used in the experiments.



(a) Spatial RM



(b) Simple RM

Fig. 8. Segmentation results of the cardiac perfusion MRI sequence together with the estimated curves using (a) the proposed spatial RM and b) the simple RM.

(LAD) coronary arteries (that are directly supplied contrast enhanced blood from the left main), the perfused myocardium and finally the right and left ventricles. Figure 8 shows the results of segmentation with the two methods performed with a predefined number of clusters K = 7. An expert in cardiac MRI inspected and evaluated the two methods using the original MR images as reference. Comparison of the two methods demonstrated that (a) both provide efficient segmentations of the heart ventricles, myocardium and arteries, and (b) the *spatial RM* results offer a far better spatial coherence of the cardiac segments (e.g., note in Fig. 7a the

artifactual segments in the antero-basal area of the myocardium that are absent in Fig. 8 b). The segmentation efficiency of the *spatial RM* method offers excellent capabilities for segmenting out tissue based on its spatiotemporal features. This has many potential applications especially in the emerging field of interventional and functional MRI, for optimizing the assessment and quantification of myocardial perfusion, including the generation of perfusion maps, and the generation of masks for 3D reconstruction of multislice perfusion or vascular MRI.

The proposed methodology was used with a second set of cardiac perfusion MRI images. Figure 9 shows representative frames of two slices (out of seven) from this study. Slice 1 corresponds to an apical and slice 2 corresponds to a midsection (apical-to-basal) oblique short axis view of the heart. With the CESMIR sequence, the signal intensity (SI) in the pre-contrast images (frames 1) is substantially reduced for improved dynamic range of the contrast enhancement. After infusion of the contrast agent (frames 2 to 4), the segment of the myocardium that is directly perfused by the left main coronary artery uptakes contrast agent and demonstrate increased signal intensity (SI). In this particular animal, we observe enhancement of the anterior, antero-septal and antero-lateral walls.

Figure 10 shows results of the segmentation of those two slices with the proposed method using a $K = 5$. In both slices the five clusters reflect anatomical areas with differential SI responses, which are consistent with the aforediscussed differential uptake of contrast agent, as well as differences in the pre-contrast SI. The important finding is that the method effectively segments the perfused territories (cluster 4 in slice 1 and cluster 2 in slice 2). This observation can also be appreciated from the SI time curves, in Fig. 10 C) and F), which show that those two clusters correspond to high signal enhancement. Notably, the efficiency of the algorithm is manifested by the clear delineation of smaller anatomical structures such as the papillary muscle (PM) in slice 2. In contrast, the non-perfused portion of the myocardium (i.e. posterior and postero-septal segments) is separately clustered and demonstrates no signal enhancement.

The presented in vivo application of the proposed approach demonstrated high efficiency by accurately segmenting the perfused areas of the myocardium, while it categorized other anatomical structures in clusters primarily based on their pre-contrast SI. The achievement of such good fidelity segmentation is the most important step toward generating three dimensional (3D) perfusion maps for use in both diagnosis, as well as for planning and assessing the performance of therapeutic interventions. In view of such a potential clinical use of this method, we also selected this particular animal model: with intracoronary (localized) infusion of contrast agent only the portion of the tissue that is perfused by the left main coronary artery is enhanced. As a result, a well defined anatomical area is artificially defined that is appropriate for segmentation by clustering. An additional benefit of this model is the use of highly diluted contrast agent, so that during first pass and re-circulation, the right and left ventricles do not show any significant enhancement [21], [22], thereby reducing the complexity of tissue spatio-temporal properties. This protocol also

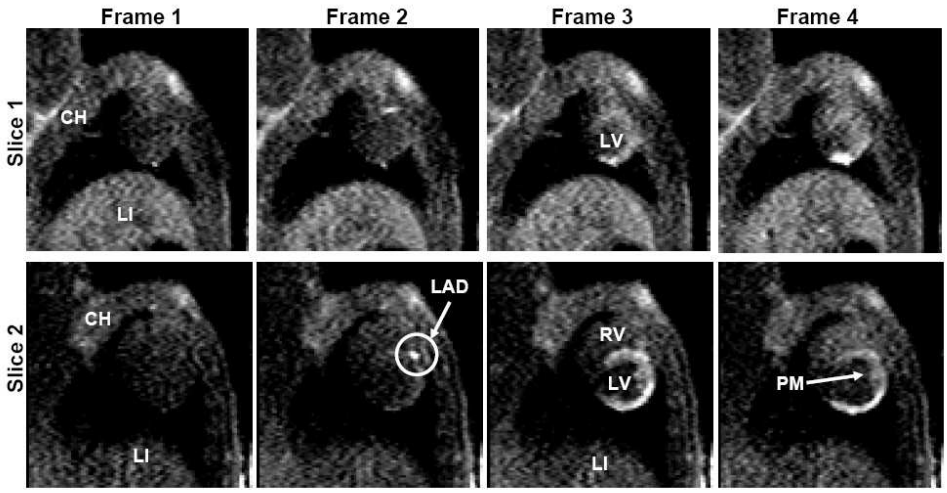


Fig. 9. Representative frames from a myocardial perfusion MRI study depicting four frames of two slices (out of seven). Frames 1 were collected before the infusion of contrast agent (pre-contrast). Frames 2, 3 and 4 were collected to different instances after intracoronary infusion of Gd-DTPA into the left main artery. Note the high signal intensity of the portion of the left main during the passage of the agent before the agent distributes into the myocardium. (CH = chest wall, LI = liver, LV = Left Ventricle, RV = Right Ventricle, LAD = Left Anterior Descending coronary artery, PM = Papillary Muscle)

simulates the appearance of the heart with infarction, secondary to coronary artery stenosis.

4. Conclusions and Future Work

In this paper we presented a methodology based on regression mixture modeling for analysis of spatiotemporal data. The main feature of this approach is the incorporation of a spatial smoothness prior for capturing spatial information. This approach was tested using numerical experiments with both simulated and real cardiac perfusion MRI data spatiotemporal data. These experiments demonstrated that the proposed spatial smoothness constraint always improves the ability of the regression mixture model to discover spatial patterns in the data. As far as the temporal patterns are concerned, we demonstrated that spatial smoothness constraints are gainful only in cases where the available data are not sufficient to accurately estimate the model parameters. We also demonstrated that the proposed methodology can be very beneficial for cardiac perfusion MRI applications since it has the ability to capture the anatomical structures of the perfused areas of the myocardium.

In the future we plan to extend this stochastic model to automatically detect the order of the regressor that is necessary to model the underlying temporal pattern and also to study other functional models for regression such as splines. Moreover,

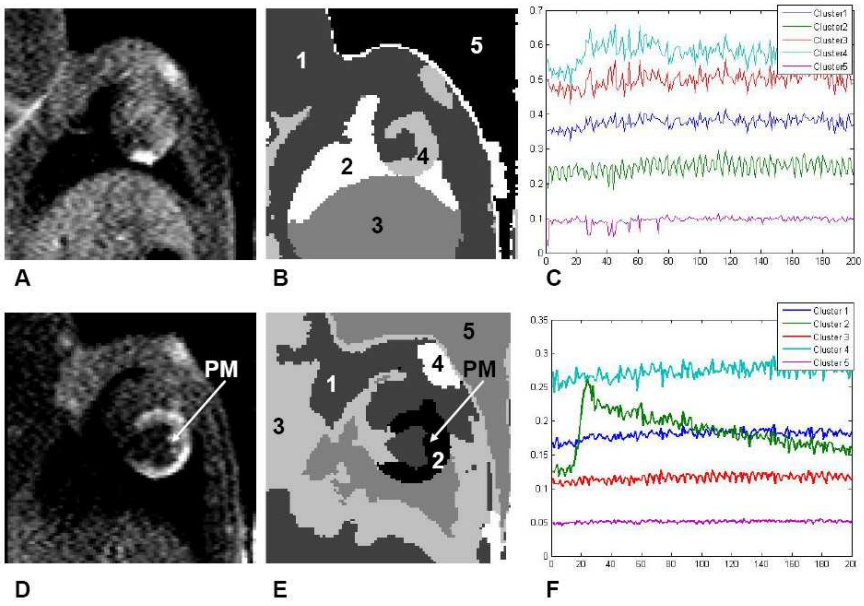


Fig. 10. Segmentation results with the proposed approach for the two slices shown in Figure 6. (A, D) original MRI frames. (B, E) Segmentation masks with K = 5. (C, F) Signal Intensity time curves for each one of the segments.

other interesting medical image applications for this methodology include the extraction of brain activation functional maps from functional MRI and monitoring the physiologic motion of tissue for guiding image based interventions and surgeries. Finally, we also plan to test the value of our approach in surveillance and tracking applications from video sequences [8].

Acknowledgments

This work was supported in part by the National Institutes of Health grant RO1HL067924 (NVT).

References

1. C.M. Bishop. *Pattern Recognition and Machine Learning*. Springer, 2006.
2. G. M. McLachlan and D. Peel. *Finite Mixture Models*. New York: John Wiley & Sons, Inc., 2001.
3. S. J. Gaffney and P. Smyth. Curve clustering with random effects regression mixtures. In C. M. Bishop and B. J. Frey, editors, *Proc. of the Ninth Intern. Workshop on Artificial Intelligence and Statistics*, 2003.
4. W. S. DeSarbo and W. L. Cron. A maximum likelihood methodology for clusterwise linear regression. *Journal of Classification*, 5(1):249–282, 1988.
5. D. Chudova, S. Gaffney, E. Mjolsness, and P. Smyth. Mixture models for translation-invariant clustering of sets of multi-dimensional curves. In *Proc. of the Ninth ACM*

SIGKDD Intern. Conf. on Knowledge Discovery and Data Mining, pages 79–88, Washington, DC, 2003.
6. S. J. Gaffney. *Probabilistic curve-aligned clustering and prediction with regression mixture models*. PhD thesis, Department of Computer Science, University of California, Irvine, 2004.
7. J. Brankov, N. Galatsanos, Y. Yang, and M. Wernick. Segmentation of dynamic PET or fMRI images based on a similarity metric. *IEEE Trans. on Nuclear Science*, 50(5):1410–1414, 2003.
8. G. Antonini and J. Thiran. Counting pedestrians in video sequences using trajectory clustering. *IEEE Trans. on Circuits and Systems for Video Technology*, 16(8):1008–1020, 2006.
9. K. Blekas, A. Likas, N. Galatsanos, and I. E. Lagaris. A spatially-constrained mixture model for image segmentation. *IEEE Trans. on Neural Networks*, 16(2):494–498, 2005.
10. C. Nikou, N. Galatsanos, and A. Likas. A class-adaptive spatially variant finite mixture model for image segmentation . *IEEE Trans. on Image Processing*, 14(4):1121–1130, 2007.
11. A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum Likelihood from incomplete data via the EM algorithm. *J. Roy. Statist. Soc. B*, 39:1–38, 1977.
12. C. Fraley and A. E. Raftery. Bayesian regularization for normal mixture estimation and model-based clustering. *The Computer Journal*, 41:578–588, 1998.
13. P. J. Green. Bayesian Reconstructions from Emission Tomography Data Using a Modified EM Algorithm. *IEEE Trans. on Medical Imaging*, 9(1):84–93, 1990.
14. J. Besag, J. York, and A. Mollie. Bayesian image restoration with two applications in spatial statistics (with discussion). *Annals of the Institute of Statistical Mathematics*, 43:1–59, 1991.
15. Y. Zhang, M. Brady, and S. Smith. Segmentation of Brain MR Images Through a Hidden Markov Random Field Model and the Expectation-Maximization Algorithm. *IEEE Trans. on Medical Imaging*, 20(1):45–57, 2001.
16. G. Schwarz. Estimating the dimensions of a model. *Annals of Statistics*, 6:461–464, 1978.
17. L. Wasserman. Bayesian model selection and model avaraging. *Journal of Mathematical Psychology*, 44(1):92–107, 2000.
18. S.A. Wood, W. Jiang, and M. Tanner. Bayesian mixture of splines for spatially adaptive nonparametric regression. *Biometrika*, 89(3):513–528, 2002.
19. N. P. Galatsanos and A. K. Katsaggelos. Methods for choosing the regularization parameter and estimating the noise variance in image restoration and their relation. *IEEE Trans. on Image Processing*, 1(3):322–336, 1992.
20. D. Gui and N. V. Tsekos. Fast magnetization-driven preparation for imaging of contrast-enhanced coronary arteries during intra-arterial injection of contrast agent. *J. Magn. Reson. Imaging*, 24:1151–1158, 2006.
21. N. V. Tsekos, P. K. Woodard, G. J. Foster, P. Moustakidis, T. L. Sharp, P. Herrero, and R. J. Gropler. Dynamic coronary MR angiography and first-pass perfusion with intracoronary administration of contrast agent. *J Magn. Reson.*, 16:311–319, 2002.
22. N. V. Tsekos, E. Atalar, D. Li, R. A. Omary, J. M. Serfaty, and P. K. Woodard. Magnetic resonance imaging-guided coronary interventions. *J Magn. Reson.*, 19:734–749, 2004.