Curve clustering with spatial constraints for analysis of spatiotemporal data

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

N. V. Tsekos
School of Medicine, Whashington University in St. Louis
E-mail: tsekosn@mir.wustl.edu

Abstract

In this paper we present a new approach for curve clustering designed for analysis of spatiotemporal data. Such kind of data contains both spatial and temporal patterns that we desire to capture. The proposed methodology is based on regression and Gaussian mixture modeling and the novelty of the herein work is the incorporation of spatial smoothness constraints in the form of a prior for the data labels. This enables the proposed model to take into account the underlying property of spatiotemporal data that spatially adjacent data points most likely should belong to the same cluster. A maximum a posteriori Expectation Maximization (MAP-EM) algorithm is used for learning this model. We present numerical experiments with simulated data where the ground truth is known in order to assess the value of the introduced smoothness constraint, and also with real cardiac perfusion MRI data. The results are very promising and demonstrate the value of the proposed constraint for analysis of such data .

1 Introduction

Clustering, apart from being on its own a challenging field of research, is useful to a wide spectrum of application areas, such as pattern recognition, machine learning, computer vision, bioinformatics, web mining, etc. Gaussian mixture modeling (GMM) 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 [3], [13], and [7]. 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 very long and so 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 by a parametric model and then to cluster the predictions by the model. Different types of parametric models have been used to capture the trends in such data. Polynomial and spline regression models are among the most commonly used ones [10] that have been successfully used for curve clustering in a number of diverse applications, ranging from gene clustering in bioinformatics to clustering of cyclone trajectories, see for example [8] [6] and [9].

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 [5], [1], and the references within. In such studies it is important to measure both the temporal characteristics of the grouped pixels and simultaneously to accurately classify the pixels in groups of similar temporal behavior.

Thus, for this type of data in determining class membership, apart from the distance between the coefficients of the model, it is also beneficial to use spatial constraints. Such constraints capture the prior knowledge that adjacent pixels most likely belong to the same class. The idea of combining GMM with spatial smoothness prior has been used previously with success for segmentation of natural images [4] and [14].

In this paper we extend this idea to the problem of time-sequence analysis via regression based curve clustering. In other words, we model the curves with a regression model and use a spatially variant Gaussian mixture model

with a Gaussian Markov random field (MRF) smoothness prior which is imposed on the labels of the data. Then, a maximum a posteriori expectation maximization algorithm (MAP-EM) [7], [13] is used to learn this model and cluster the data.

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-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. $e_i \sim \mathcal{N}(0, \Sigma)$, where $\Sigma = diag(\sigma_1^2, \ldots, \sigma_T^2)$. 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. to divide the set of curves y_i with $i = 1, \ldots N$ into K clusters where each cluster will contain curves obtained from the same 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 [7] is an efficient framework for solving likelihood estimation problems for GMMs. 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 [8, 10]:

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

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

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

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 [4]. 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 [11, 2, 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 N_i is the set of adjacent pixels m around pixel i ($|N_i| = 8$ in the general case).

As explained in [14], 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

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

Direct maximization of this function is difficult thus we resort to the EM methodology which maximizes the expected value of the MAP log likelihood of the complete data

$$\begin{aligned} Q(\Theta|\Theta^{(t)}) = \quad & \sum_{i=1}^{N} \sum_{j=1}^{K} z_{ij}^{(t)} \{\log \pi_{ij} + \log p(y_i|\theta_j)\} - \\ & - \log \xi_j - \frac{\sum_{m \in \mathcal{N}_i} (\pi_{ij} - \pi_{mj})^2}{2\xi_j^2} \,. \end{aligned} \tag{13}$$

Based on the EM methodology the E-step (Eq. 4) and the update rules during the M-step 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 (13) with respect to parameters π_{ij} equal to zero we take the following quadratic equation:

$$\pi_{ij}^2 - \widetilde{\pi}_{ij} \pi_{ij} - \frac{\xi_j^2}{|\mathcal{N}_i|} z_{ij}^{(t)} = 0 \,, \tag{14}$$

where $\widetilde{\pi}_{ij} = \frac{1}{|\mathcal{N}_i|} \sum_{m \in \mathcal{N}_i} \pi_{mj}$ 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 $\pi_{ij} \geq 0$:

$$\pi_{ij}^{(t+1)} = \frac{\widetilde{\pi}_{ij} + \sqrt{\widetilde{\pi}_{ij}^2 + 4\frac{\xi_j^2}{|\mathcal{N}_i|} z_{ij}^{(t)}}}{2} \,. \tag{15}$$

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. 15 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. 15) we project them on the constraint convex hull. For this projection an efficient convex quadratic programming algorithm presented in [4] is used.

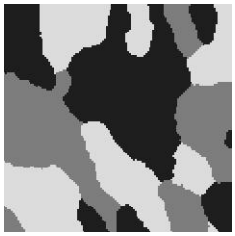
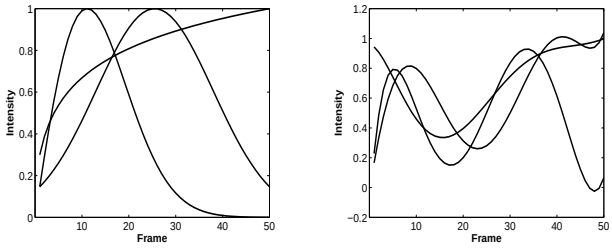


Figure 1. The 3-class test image used in the experiments.



Set A *Set B*

Figure 2. Two sets of the three signal patterns used for generating the image sequences of the piece-wise constant image in Fig. 1

Finally, by taking the derivative of Q-function (Eq. 13) with respect to the smoothness parameters ξ_j we obtain the following update rule:

$$\xi_j^{2(t+1)} = \frac{1}{N} \sum_{i=1}^{N} \sum_{m \in \mathcal{N}_i} (\pi_{ij}^{(t+1)} - \pi_{mj}^{(t+1)})^2 \,. \qquad (16)$$

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.

3 Experimental results

The performance of the proposed regression mixture model with smoothness constraints, referred to as *spatial RM* in the experiment section of this paper, 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 where visual inspection by an expert confirmed the validity of the segmentation results. Comparison has made with the simple regression mixture, referred to this section as *simple RM*. Both methods were initialized using the same strategy. In particular, for each curve y_i we estimate first the $p+1$ regression parameters as given by the simple least-square solution, and then we applied the K-means algorithm to this parameter space (we perform 10 different runs of the K-means and keep the best solution). Initial values for the regression parameters β_j are found directly from the optimum centers of the K-means solution, while the diagonal covariance matrices Σ_j are initialized indirectly from the samples that belong to each discovered cluster.

3.1 Simulated spatiotemporal data

At first, we test the performance of our method with simulated spatiotemporal data with known ground truth. In particular, we have used the piece-wise constant image in Fig. 1 that contains three distinct regions ($K = 3$) [15]. Simulated image sequences with $T = 50$ time frames were generated by varying the intensity of each region of the image according to a predefined time pattern. Two sets of such patterns were used in our study, namely *Set A* and *Set B*, shown in Fig. 2: The first one consists of a Gaussian, a Rayleigh and a logarithmic curve, while the second set contains more complex curves obtained by polynomials of degree $p = 6$. The generated sequences of frames were also corrupted by additive white Gaussian noise with properly defined variance in order to obtain signal to noise ratios (SNR) varying between $16dB$ and $-18dB$. Some characteristic examples of the obtained frames are shown in Fig.3.

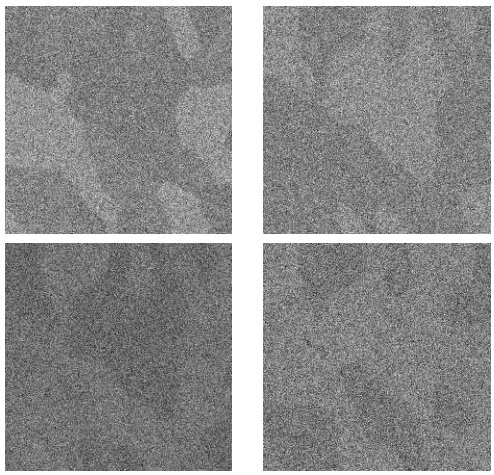


Figure 3. Sample frames from the simulated sequence *Set B* with $SNR = 0dB$.

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

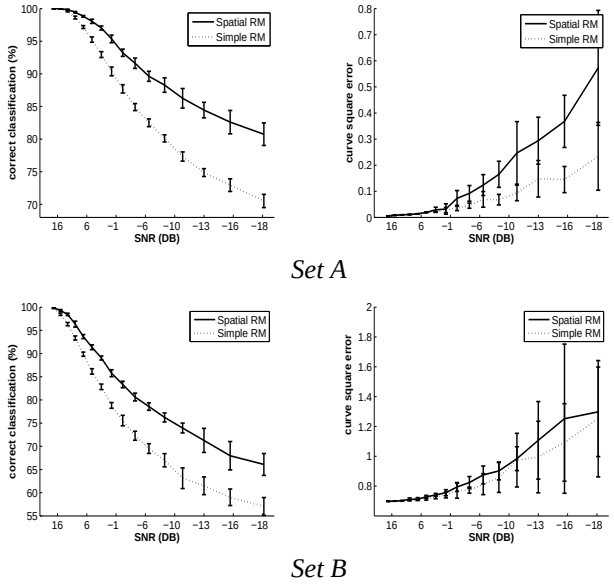


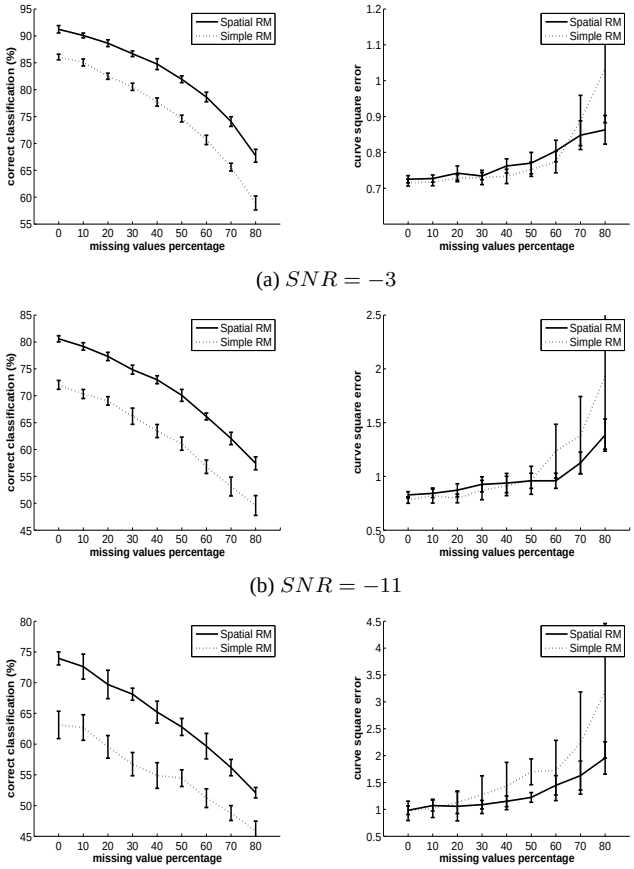
Figure 4. Percentage of correct classification and curve square error as a function of the SNR for the two sets of signal patterns.

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.

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

This quantifies the performance of the proposed methodology to capture the temporal patterns of the data. In order to obtain a meaningful comparison, for each problem (SNR value) we performed 20 different executions of both methods with different seeds and kept records of the mean value and the standard deviation of the two evaluation criteria.

Figure 4 presents the evolution of these quantities as a function of the SNR for the two sets of sequences based on the temporal patterns (*Set A* and *Set B*). As can be observed, the proposed methodology improves significantly classification accuracy as compared to simple RM for the entire range of examined SNR, especially in lower values. On the other hand, this study on both sets showed that the introduction of the spatial information to the regression model does not seem to improve the capability of estimating the temporal patterns, since the calculated curve square error was always larger with the constraints. However, this difference was insignificant.



(a) $SNR = -3$

(b) $SNR = -11$

(c) $SNR = -16$

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

3.2 Results with missing data

This methodology is also beneficial in the missing data case where, for certain time instances, there may be missing measurements at certain pixel locations. An example of such a situation is in functional MRI where isolated spiking on the radio frequency coils can cause variable temporal and spatial missing data. This is an interesting application of regression mixtures, since any other conventional clustering methodology cannot be used in a straightforward manner. More specifically, this scenario refers to the case when each temporal pattern is of variable length, i.e. $y_i = \{y_{il}\}_{l=1}^{T_i}$. In other words, there are values for some time instances that are missing. Then, the polynomial regression model of y_i is given by

$$y_i = X_i\beta + e_i \ , \qquad (17)$$

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

$$\beta_j^{(t+1)} = \left[\sum_{i=1}^{N} z_{ij}^{(t)} X_i^T \Sigma_j^{-1(t)} X_i \right]^{-1} \sum_{i=1}^{N} z_{ij}^{(t)} X_i^T \Sigma_j^{-1(t)} y_i \ . \tag{18}$$

Additional experiments have been conducted in an attempt to study the behavior of our methodology in situations with missing data. In particular, we have introduced a parameter that provides the percentage of the time samples that will be removed from the observations. Since these removed samples are chosen at randomly, no spatial location is privileged with respect to any other. Therefore, the percentage of the missing data expresses the probability that a time sample may not be present at a spatial location. Experiments were conducted on the signal pattern *Set B* 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 is superior in comparison with the simple RM model for all levels of missing data. However, the study showed a significant improvement of the curve square error (CSE) criterion since the spatial RM method provides lower CSE values, especially in noisy environments with a lot of missing data. This observation suggests that the proposed prior offers a more robust regression approach in the absence of data.

3.3 Cardiac perfusion MRI sequence segmentation

We also applied our method to a real set of spatiotemporal data from an in vivo dynamic cardiac magnetic resonance imaging (MRI) study. 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 study was performed on an instrumented pig with an intracoronary catheter inserted into the left main coronary artery. Dynamic imaging was performed by acquiring a series of 100 2D MR images aligned along an oblique long axis view to image with a magnetization prepared pulse sequence (TR/TE/a = 2.2/1.2/20; FOV = 200 X 200 mm2; slice = 5 mm; matrix = 96x96) [12]. After 15 precontrast images, 4 mL of 0.125 mM Gd-DTPA (Omniscan; Amersham Health, Princeton, NJ, USA) were administered at 2 mL/second.

Figure 6 illustrates representative MR images from this sequence, illustrating the differential enhancement of the left circumflex (LCx) and left anterior descending (LAD)

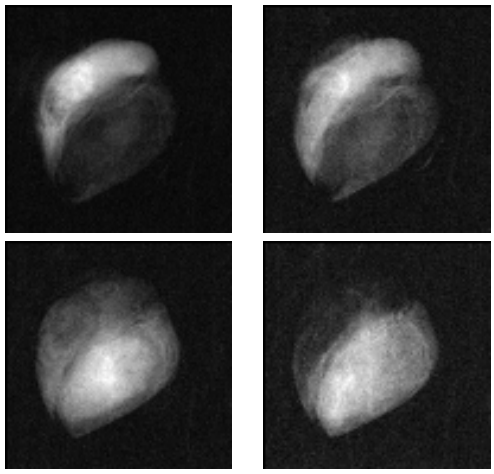
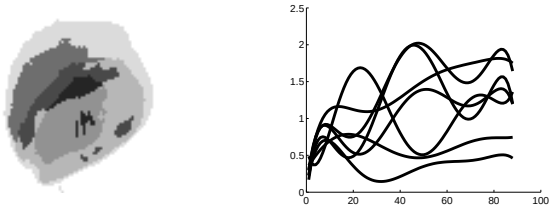
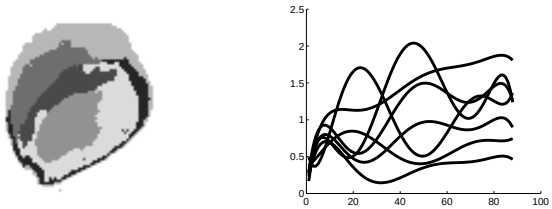


Figure 6. Sample frames from the cardiac perfusion MRI sequence used in the experiments.



(a) Simple RM



(b) Spatial RM

Figure 7. Segmentation result of the cardiac perfusion MRI sequence into $K = 7$ segments using simple RM (a) and the proposed spatial RM (b). The estimated $p = 4$-order curves are also shown.

coronary arteries (that are directly supplied contrast enhanced blood from the left main), the perfused myocardium and finally the right and left ventricles. Figure 7 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. 7 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.

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 demonstrated using numerical experiments to be effective in discovering both the spatial and the temporal patterns of spatiotemporal data. Furthermore, it was demonstrated to be robust to both noise and missing data.

In the future we plan to extend this stochastic model in two directions. First, to automatically detect the order of the regressor that is necessary to model the underlying temporal pattern. Second, to automatically find the number of components of the mixture model that is necessary to capture the spatial component of the data. Moreover, other interesting medical image applications for studying include the extraction of brain activation functional maps from functional MRI and monitoring the physiologic motion of tissue for monitoring and guiding image guided interventions and surgeries. Finally, we also plan to explore our approach in surveillance and tracking applications from video sequences [1].

Acknowledgments

This work was partially supported by Interreg IIIA (Greece-Italy) Grant (No I2101005)

References

[1] 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.

[2] 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.

[3] C. Bishop. *Pattern Recognition and Machine Learning*. Springer, 2006.

[4] 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.

[5] 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.

[6] 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.

[7] 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.

[8] W. S. DeSarbo and W. L. Cron. A maximum likelihood methodology for clusterwise linear regression. *Journal of Classification*, 5(1):249–282, 1988.

[9] S. J. Gaffney. *Probabilistic curve-aligned clustering and prediction with regression mixture models*. PhD thesis, Department of Computer Science, University of California, Irvine, 2004.

[10] 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.

[11] P. J. Green. Bayesian Reconstructions from Emission Tomography Data Using a Modified EM Algorithm. *IEEE Trans. on Medical Imaging*, 9(1):84–93, 1990.

[12] 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.

[13] G. M. McLachlan and D. Peel. *Finite Mixture Models*. New York: John Wiley & Sons, Inc., 2001.

[14] 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.

[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.