Digital Signal Processing 59 (2016) 106–114

Contents lists available at ScienceDirect

Digital Signal Processing

www.elsevier.com/locate/dsp

Bayesian classification of hyperspectral images using spatially-varying Gaussian mixture model ☆

Koray Kayabol [a,*], Sezer Kutluk [b]

[a] Electronics Engineering Department, Gebze Technical University, Kocaeli, Turkey
[b] Electrical-Electronics Engineering Department, Istanbul University, Istanbul, Turkey

ARTICLE INFO

Article history:
Available online 26 August 2016

Keywords:
Gaussian mixture model
Spatially-varying
Spectral-spatial
Small sample size
Hyperspectral

ABSTRACT

We propose a spatially-varying Gaussian mixture model for joint spectral and spatial classification of hyperspectral images. The model provides a robust estimation framework for small sample size training sets. Defining prior distributions for the mean vector and the covariance matrix enables us to regularize the parameter estimation problem. More specifically, we can obtain invertible positive definite covariance matrices by the help of this regularization. Moreover, the proposed model also takes into account the spatial alignments of the pixels by using spatially-varying mixture proportions. The spatially-varying mixture model is based on spatial multinomial logistic regression. The classification results obtained on Indian Pines, Pavia Centre, Pavia University, and Salinas data sets show that the proposed methods perform better especially for small-sized training sets compared to the state-of-the-art classifiers.

1. Introduction

As one of the main topics of remote sensing research, hyperspectral images (HSIs) are used in many real-life applications including forest vegetation mapping and classification, urban and land usage, determination of the water resources, and classification of the crop species. The aim of this paper is supervised joint spectral and spatial classification of hyperspectral images. We propose a generative probabilistic model that is able to regularize the under-determined covariance matrix estimation problem for small sample size training data and provides a joint spectral and spatial classification framework as well.

Statistical mixture models are frequently used in classification problems [1]. Although GMM is a widely-used model, it is not much preferred in HSI classification since, in general, the length of the feature vectors, or spectral bands, is high, and the number of training samples is small. For an L-dimensional data vector, there are $L+(L^2+L)/2$ number of unknowns to be estimated for a single mixture component (L unknowns for mean vector and $(L^2+L)/2$ unknowns for covariance matrix). Since there is a smaller number of training samples compared to unknowns, the estimation problem becomes under-determined. To overcome the difficulty of under-determined estimation problem, Tadjudin and Landgrebe [2] propose a covariance matrix estimator called bLOOC that consists of a combination of three covariance estimators. In [3], the covariance matrices are constrained to be in a particular structure to reduce the number of parameters to be estimated. In [4], regularized linear discriminant analysis (LDA) is used. A dimension reduction before GMM classification is proposed in [5]. A Bayesian mixture model is first introduced in our previous work [6]. This study is an improved and extended version of [6] with different priors for the means and covariance matrices. The Bayesian framework enables us to regularize an under-determined estimation problem by defining appropriate prior distributions for the parameters. For example, using a few number of samples to estimate the covariance matrix of a Gaussian may cause a non-invertible covariance matrix. Since normal-inverse-Wishart prior is a conjugate prior for the unknown mean and covariance matrix of a Gaussian random vector [7], we use normal-inverse-Wishart prior to regularize the covariance matrix estimation problem. Without using a prior, direct maximum likelihood estimation of covariance matrix from hyperspectral data causes a non-invertible covariance matrix. By defining the prior distribution, we are able to obtain an invertible covariance matrix.

In this study, we focus on joint spectral and spatial classification of HSIs rather than pixel-wise spectral classification. Spectral-spatial classification is another challenge in HSI classification [8]. A general and intuitive approach for spectral-spatial classification of hyperspectral images consists of classification and segmentation as two successive, independent processes. K-means, k-nearest neighbors (KNN), support vector machine (SVM), and linear

☆ This work is supported by Scientific and Technological Research Council of Turkey (TUBITAK) under Project No. 114E535.

* Corresponding author.
E-mail addresses: koray.kayabol@gtu.edu.tr (K. Kayabol), sezer.kutluk@gmail.com (S. Kutluk).

http://dx.doi.org/10.1016/j.dsp.2016.08.010
1051-2004/

and quadratic discriminant analysis can be given as examples of mostly-used classification methods. In this kind of approaches, an image smoothing algorithm after the classification is needed in order to have a smooth classification map. A widely used smoothing model for image segmentation is a discrete-valued Markov random field called Potts model [9]. Potts model is already used in hyperspectral image segmentation in [10,11]. An SVM classification followed by an MRF segmentation is proposed by Tarabalka et al. [12]. Since the output of an SVM classifier is not a probability measure, in [12] the output of SVM is converted to probability measures as proposed in [13]. These probabilities are then used as likelihood in MRF segmentation. In generative probabilistic models, the likelihood term represents the spectral information, while the prior term models the spatial interaction of the pixels. Gamba and Trianni [14] and Plaza et al. [15] propose Markov random field (MRF) prior model for segmentation of HSI. They use a kind of Gaussian mixture models for likelihood. In this study, we propose a generative probabilistic model and supervised learning algorithm that performs contextual classification. For contextual classification of hyperspectral images rather than using MRF, we propose a spatially-varying Gaussian mixture model. By spatially-varying, we mean that the mixture proportions of the Gaussian distributions are changed according to pixel locations.

Spatially-varying mixture models have already been used in optical, medical and synthetic aperture radar images for segmentation and classification. For example, in [16–19] an MRF prior is used for mixture proportions to achieve a spatially-varying mixture model. In [20] and [21], a latent Gaussian random field is proposed such that the mixture proportions are connected to the class labels by a Multinomial Logistic (MNL) function. In [22] and [23] a spatial auto-logistic regression model is defined on class labels for synthetic aperture radar image classification. To the best of our knowledge, spatially-varying mixture models are first used for HSI classification problem in our previous short paper [6].

The paper has two main contributions listed as follows: 1) to overcome the small sample size problem, a Bayesian Gaussian mixture model with normal-inverse-Wishart prior is used for HSI classification, and 2) spatially-varying mixture model allows joint spectral and spatial classification of HSIs. Organization of the paper is as follows. Section 2 and 3 respectively present the proposed Bayesian models and the related classification algorithm. The experimental results are reported in Section 4. Section 5 summarizes the conclusion and future work.

2. Spatially-varying Gaussian mixture model

For an HSI with N pixels and L spectral bands, we denote each spectral vector by $\mathbf{s}_n \in \mathbf{R}^L$ where $n = 1, \ldots, N$ is the lexicographically ordered pixel indices.

Assuming that there are K number of land cover classes in the HSI, we define a K-dimensional label vector $\mathbf{z}_n \in \{0, 1\}^K$ for each pixel. The binary label vector $\mathbf{z}_n$ has the property that $\sum_{k=1}^{K} z_{n,k} = 1$ which means it indicates only one of the K classes by assigning its related element to 1. We can write $\mathbf{z}_n \in \mathcal{Z} = \{[1, 0, \ldots, 0], [0, 1, \ldots, 0], \ldots, [0, 0, \ldots, 1]\}$.

For all pixels, the joint conditional density of $\mathbf{s}_n$'s and $\mathbf{z}_n$'s is written as

$$p(\mathbf{s}_{1:N}, \mathbf{z}_{1:N} | \theta_{1:K}, \beta) = p(\mathbf{s}_{1:N} | \mathbf{z}_{1:N}, \theta_{1:K}) p(\mathbf{z}_{1:N} | \beta) \tag{1}$$

where $\theta_{1:K}$ and β are the parameters of the densities. By assuming that $\mathbf{s}_n$ vectors are conditionally independent given the labels, $\mathbf{z}_n$ vectors, the joint density in (1) can be simplified as

$$p(\mathbf{s}_{1:N}, \mathbf{z}_{1:N} | \theta_{1:K}, \beta) = \left[\prod_{n=1}^{N} \prod_{k=1}^{K} p(\mathbf{s}_n | \theta_k)^{z_{n,k}} \right] p(\mathbf{z}_{1:N} | \beta) \tag{2}$$

where $p(\mathbf{z}_{1:N} | \beta)$ is the prior density of the class labels. We assume that $\mathbf{s}_n$ vectors are independent but the hidden labels, $\mathbf{z}_n$ vectors, are spatially dependent. To introduce spatial information into (2), we need a dependent probabilistic model for $\mathbf{z}_{1:N}$ rather than independent and identically distributed multinomial model.

The density in (2) defines a mixture model. In order to show the mixture model apparently, we need to write the probability density of a single pixel and its label conditioned on the rest of the pixels in the image

$$p(\mathbf{s}_n, \mathbf{z}_n | \mathbf{s}_{\bar{n}}, \mathbf{z}_{\bar{n}}, \theta_{1:K}, \beta) = p(\mathbf{s}_n | \mathbf{z}_n, \theta_{1:K}) p(\mathbf{z}_n | \mathbf{z}_{\bar{n}}, \beta) \tag{3}$$

where $\bar{n}$ is the complement of n with respect to set $\{1, 2, \ldots, N\}$, i.e. $\bar{n} = \{1, 2, \ldots, N\} \backslash \{n\}$ and β is the smoothing parameter. Assuming a Markov property that only the neighbor pixels are dependent, (3) can be written as

$$p(\mathbf{s}_n, \mathbf{z}_n | \mathbf{s}_{\tilde{n}}, \mathbf{z}_{\tilde{n}}, \theta_{1:K}, \beta) = p(\mathbf{s}_n | \mathbf{z}_n, \theta_{1:K}) p(\mathbf{z}_n | \mathbf{z}_{\tilde{n}}, \beta) \tag{4}$$

where of $\tilde{n}$ is the set of pixels around the nth pixel. We may write the marginal density of $\mathbf{s}_n$ using the joint density in (4) as follows:

$$\begin{aligned} p(\mathbf{s}_n | \mathbf{z}_{\tilde{n}}, \theta_{1:K}, \beta) &= \sum_{\mathbf{z}_n} \prod_{k=1}^{K} [p(\mathbf{s}_n | \theta_k) \omega_{n,k}]^{z_{n,k}} \\ &= \sum_{k=1}^{K} p(\mathbf{s}_n | \theta_k) \omega_{n,k} \end{aligned} \tag{5}$$

where $\omega_{n,k}$ is the spatially-varying mixture proportions and is related to label prior as follows:

$$\omega_{n,k} = p(z_{n,k} = 1 | z_{\tilde{n},k}, \beta) \tag{6}$$

From (5), the spatially-varying mixture model can be seen apparently. In the following subsections, we give the details of the densities used in the proposed model.

2.1. Bayesian Gaussian mixture model

We assume that each class in the data is generated by a component of a mixture of multivariate Gaussian distributions. This assumption tells us that a feature vector related to a pixel in the image is a sample from one of the K multivariate Gaussian distributions. Therefore, the distribution of the kth class is a multivariate Gaussian given below

$$p(\mathbf{s}_n | \mathbf{m}_k, \Sigma_k) = \mathcal{N}(\mathbf{s}_n | \mathbf{m}_k, \Sigma_k) \tag{7}$$

where $\mathbf{m}_k$ and Σ_k are the mean vector and the covariance matrix of the kth class, respectively. For the kth class, the parameter set is defined to be $\theta_k = \{\mathbf{m}_k, \Sigma_k\}$. We define a normal-inverse-Wishart prior for $\mathbf{m}_k$ and Σ_k, i.e.

$$p(\mathbf{m}_k, \Sigma_k) = \mathcal{N}\left(\mathbf{m}_k \middle| \mathbf{m}_0, \frac{1}{\lambda} \Sigma_k\right) \mathcal{W}^{-1} (\Sigma_k | (\tau + L + 1)\Psi, \tau) \tag{8}$$

where $\mathbf{m}_0$, Ψ and τ are the parameters of the priors. The expressions of the probability density functions (pdfs) are given in Appendix A.

2.2. Spatially varying mixture proportions

Since we assume that the pixel labels are not independent, we should define the conditional density of a single pixel $z_{n,k}$ conditioned on its neighbor pixels $z_{\tilde{n},k}$. We use an auto-logistic regression model for spatially dependent class labels. According to auto-logistic regression, the conditional probability of a class label can be given as follows [24]:

$$p(z_{n,k} | z_{\tilde{n},k}, \beta) \propto e^{\beta \left(z_{n,k} + z_{n,k} \sum_{m \in \tilde{n}} z_{m,k}\right)} \tag{9}$$