CONTEXTUAL CLUSTERING AND UNMIXING OF GEOSPATIAL DATA BASED ON GAUSSIAN MIXTURE MODELS AND MARKOV RANDOM FIELDS

Sawamura, Yoko
Graduate School of Mathematics, Kyushu University

Nishii, Ryuei
Faculty of Mathematics, Kyushu University

Nakamoto, Akiko
Graduate School of Mathematics, Kyushu University

Kawaguchi, Shuji
Bioinformatics and Systems Engineering Division, RIKEN

他

https://doi.org/10.5109/21045

出版情報：Bulletin of informatics and cybernetics. 41, pp.39-49, 2009-12. 統計科学研究会

CONTEXTUAL CLUSTERING AND UNMIXING OF GEOSPATIAL DATA BASED ON GAUSSIAN MIXTURE MODELS AND MARKOV RANDOM FIELDS

by

Yoko Sawamura,[*] Ryuei Nishii,[†] Akiko Nakamoto,[‡]
Shuji Kawaguchi,[§] and Tomohiko Ozaki

Reprinted from the Bulletin of Informatics and Cybernetics
Research Association of Statistical Sciences,Vol.41

FUKUOKA, JAPAN
2009

[*] Graduate School of Mathematics, Kyushu University, Motooka, Fukuoka, Japan
[†] Faculty of Mathematics, Kyushu University, Motooka, Fukuoka, Japan
[‡] Graduate School of Mathematics, Kyushu University, Motooka, Fukuoka, Japan
[§] Bioinformatics and Systems Engineering Division, RIKEN, Suehiro, Tsurumi, Yokohama, Japan

Bulletin of Informatics and Cybernetics, Vol. 41, 2009

CONTEXTUAL CLUSTERING AND UNMIXING OF GEOSPATIAL DATA BASED ON GAUSSIAN MIXTURE MODELS AND MARKOV RANDOM FIELDS

By

Yoko Sawamura[¶], **Ryuei Nishii**[‖], **Akiko Nakamoto**[**], **Shuji Kawaguchi**[††] **and Tomohiko Ozaki** [‡‡]

Abstract

In supervised and unsupervised image classification, it is known that contextual classification methods based on Markov random fields (MRFs) improve the performance of non-contextual classifiers. In this paper, we consider the unsupervised unmixing problem with the introduction of a new MRF. First, spectral vectors observed at mixels are assumed to follow Gaussian mixtures. Second, vectors representing fractions of categories are supposed to follow an MRF over the observed area. Then, we derive an unsupervised unmixing method, which is also useful for unsupervised classification. When evaluated using a synthetic data set and a benchmark data set for classification, the proposed method performed well.

Key Words and Phrases: contextual clustering, Gaussian mixture, MRF, unmixing

1. Introduction

Image classification is an important issue, especially, in the remote sensing community. For the case where training data are available, many classification methods have been studied (see Lu and Weng (2007) for a comprehensive review referring to many relevant papers). Furthermore, a machine learning approach has also been discussed, e.g., Nishii and Eguchi (2005); Kawaguchi and Nishii (2007). It is now known that contextual classification methods based on Markov random fields (MRFs) improve the performance of non-contextual classifiers.

When training data are unavailable, the application of clustering methods to feature vectors (e.g., K-means methods) is used to detect homogeneous regions in an image. In this paper, we consider a soft-classification problem without training data. Suppose that we are required to estimate fractions of categories covering each pixel for a multispectral image without training data. This issue, called **unmixing**, is usually solved by a linear equation derived by assuming that the observed feature vector is composed of a convex combination of category reflectance signatures, see Nielsen (2001). Other approaches such as independent component analysis are also promising. Contextual unmixing methods for hyperspectral data have also been proposed, e.g., Nasciment and Dias (2003), Jia and Qian (2007), and Moussaoui et al. (2007).

The purpose of this paper is to refine previous results by Nishii et al. (2008), which are based on Gaussian mixture distributions and an MRF for fraction vectors. The remaining part

[¶] Graduate School of Mathematics, Kyushu University, Motooka, Fukuoka, Japan
[‖] Faculty of Mathematics, Kyushu University, Motooka, Fukuoka, Japan
[**] Graduate School of Mathematics, Kyushu University, Motooka, Fukuoka, Japan
[††] Bioinformatics and Systems Engineering Division, RIKEN, Suehiro, Tsurumi, Yokohama, Japan
[‡‡] Graduate School of Mathematics, Kyushu University, Motooka, Fukuoka, Japan

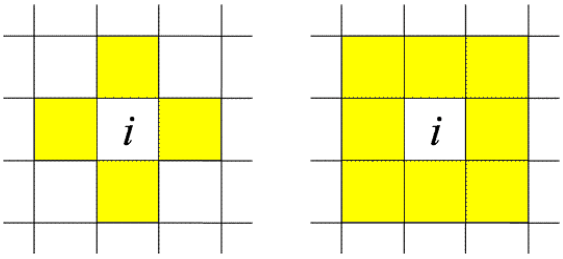


Figure 1: Pixel i and its four neighbors ($d(i, j) = 1$), and eight neighbors ($d(i, j) \leq \sqrt{2}$)

of the paper is organized as follows. Section 2 states our assumptions regarding feature vectors as well as fraction vectors. Section 3 gives the way how to maximize conditional densities. A method of estimating each fraction vector is proposed in Section 4. Section 5 looks at the possibility of a generalization of the distributional assumptions. The proposed unmixing method is numerically examined using two data sets in Section 6. Section 7 concludes the paper.

2. Assumptions for unsupervised contextual unmixing

Let $\mathcal{D}$ be a set of pixels in a multispectral image. The pixels are numbered from 1 to n, and $\mathcal{D}$ is denoted by $\{1, \ldots, n\}$. Note that the pixels are small unit areas on the surface of the earth. Assume that d-dimensional feature vector $\boldsymbol{x}_i$ is available at pixel i in $\mathcal{D}$. Now, each pixel is supposed to be a mixel of G land-cover categories $C_1, \ldots, C_G$. Our interest here is to estimate vector $\boldsymbol{f}_i = (f_{i1}, ..., f_{iG})^T$ denoting **fractions of the categories** C_g covering pixel i. Obviously, the fraction vector $\boldsymbol{f}_i$ should meet the following conditions:

$$f_{ig} \geq 0 \;\text{ and }\; \sum_{g=1}^{G} f_{ig} = 1. \tag{1}$$

Note again that we are dealing with the case **without training data** (contextual unmixing or contextual end-member detection). We will derive a contextual unmixing method under the following three assumptions regarding features and fractions.

Assumption 1: (Gaussian mixtures for feature vectors)
Let $\boldsymbol{f} = (f_1, ..., f_G)^T$ denote a vector of fractions of the categories. Then, it is assumed that the conditional distribution of feature vector $\boldsymbol{X} = (X_1, ..., X_d)^T$ given fraction vector $\boldsymbol{f}$ is expressed by a Gaussian mixture:

$$p(\boldsymbol{x} \mid \boldsymbol{f}) = \sum_{g=1}^{G} f_g \phi(\boldsymbol{x}; \boldsymbol{\mu}_g, \Sigma_g) \tag{2}$$

where $p(\boldsymbol{x}|\boldsymbol{f})$ denotes a conditional probability density function, and $\phi(\boldsymbol{x}; \boldsymbol{\mu}, \Sigma)$ is a probability density function of Gaussian distribution $N_d(\boldsymbol{\mu}, \Sigma)$ defined by $\phi(\boldsymbol{x}; \boldsymbol{\mu}, \Sigma) = (2\pi)^{-d/2}|\Sigma|^{-1/2}$ $\times \exp\left\{-(\boldsymbol{x} - \boldsymbol{\mu})^T\, \Sigma^{-1}(\boldsymbol{x} - \boldsymbol{\mu})/2\right\}$.

Assumption 2: (MRF for category-fraction vectors)
Let $\{\boldsymbol{F}_1, ..., \boldsymbol{F}_n\}$ be a set of random vectors denoting category fractions. Assume that the conditional distribution of $\boldsymbol{F}_i$ given all other fraction vectors $\boldsymbol{f}_l; l \neq i$ depends only on all of the

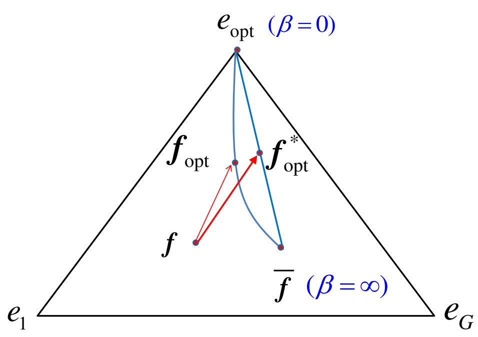


Figure 2: Optimal and quasi-optimal points

fraction vectors of its neighbors $\{\boldsymbol{f}_j \mid j \sim i\}$, where "$j \sim i$" means pixel j satisfies $0 < d(i, j) \le r$. (Figure 1 gives two neighborhoods with $r = 1, \sqrt{2}$.) Here "$d(i, j) \le r$" denotes that the distance between i and j does not exceed a pre-assigned radius r. More exactly, the joint distribution of $\boldsymbol{F}_i$ is expressed by

$$p(\boldsymbol{f}_1, ..., \boldsymbol{f}_n) = \frac{1}{Z} \exp\left\{-\beta \sum_{i \sim j}\sum \|\boldsymbol{f}_i - \boldsymbol{f}_j\|^2\right\} \tag{3}$$

where Z is a normalizing factor and $\beta \ge 0$ denotes **a granularity** of MRF. If $\beta = 0$, the fraction vectors $\boldsymbol{F}_1, ..., \boldsymbol{F}_n$ are spatially independent. If β is large, the spatial dependency becomes strong.

Assumption 3: (Conditional independence)
Let $\{\boldsymbol{X}_1, ..., \boldsymbol{X}_n\}$ be a set of random feature vectors. We assume that the conditional distribution of $\{\boldsymbol{X}_1, ..., \boldsymbol{X}_n\}$ given feature vectors $\{\boldsymbol{F}_1, ..., \boldsymbol{F}_n\} = \{\boldsymbol{f}_1, ..., \boldsymbol{f}_n\}$ is simply decomposed as

$$p(\boldsymbol{x}_1, ..., \boldsymbol{x}_n \mid \boldsymbol{f}_1, ..., \boldsymbol{f}_n) = \prod_{i=1}^{n} p(\boldsymbol{x}_i \mid \boldsymbol{f}_i). \tag{4}$$

3. Maximization of conditional densities

To estimate all fraction vectors, it is desirable if we can find the maximizer of the joint posterior density $p(\boldsymbol{f}_1, ..., \boldsymbol{f}_n \mid \boldsymbol{x}_1, ..., \boldsymbol{x}_n)$ w.r.t. $\boldsymbol{f}_1, ..., \boldsymbol{f}_n$. However, this task needs the exact expression of the normalizing factor Z in the formula (3). Therefore, the local posterior density is maximized instead of the joint posterior.

Consider the conditional density of $\boldsymbol{f}_i$ given feature vector $\boldsymbol{x}_i$ and fraction vectors $\{\boldsymbol{f}_j \mid j \sim i\}$. From Assumptions 1-3, it holds that

$$\begin{aligned} p\left(\boldsymbol{f}_i \mid \boldsymbol{x}_i, \boldsymbol{f}_j, j \sim i\right) &\propto \sum_{g=1}^{G} f_{ig}\phi(\boldsymbol{x}_i; \boldsymbol{\mu}_g, \Sigma_g) \exp\{-\beta\|\boldsymbol{f}_i - \overline{\boldsymbol{f}}_i\|^2\} \\ &= \boldsymbol{f}_i^T \boldsymbol{\phi}_i \exp\{-\beta\|\boldsymbol{f}_i - \overline{\boldsymbol{f}}_i\|^2\} \end{aligned} \tag{5}$$

where $\boldsymbol{\phi}_i = (\phi(\boldsymbol{x}_i; \boldsymbol{\mu}_1, \Sigma_1), ..., \phi(\boldsymbol{x}_i; \boldsymbol{\mu}_G, \Sigma_G))^T$, $\overline{\boldsymbol{f}}_i \equiv \sum_{j:j\sim i} \boldsymbol{f}_j / N_i$, and N_i is the number of neighbors of pixel i. Note that $\overline{\boldsymbol{f}}_i$ meets the conditions (1) since it is the averaged vector of neighboring fractions. Now, our aim is to maximize the posterior (5) w.r.t. $\boldsymbol{f}_i$ subject to the conditions (1).

If $\beta = 0$, the conditional density (5) is maximized at one of edge points $\boldsymbol{e}_1 = (1, 0, ..., 0)^T, ..., \boldsymbol{e}_G = (0, ..., 0, 1)^T$. Let $\boldsymbol{e}_{\text{opt}} = (0, ..., 1, ..., 0)^T$ be the optimal point, where mass 1 is assigned to the category maximizing $\phi(\boldsymbol{x}_i; \boldsymbol{\mu}_g, \Sigma_g)$. Furthermore, it is obvious that the optimal fraction vector converges to $\overline{\boldsymbol{f}}_i$ as β tends to infinity. This implies that the optimal point at positive β may be found in the segment connecting two points $\boldsymbol{e}_{\text{opt}}$ and $\overline{\boldsymbol{f}}_i$. The optimal point found in the segment would not be optimal (quasi-optimal), but the point is expected to be close to the optimal point.

The optimal point $\boldsymbol{f}_{\text{opt}}$ is numerically obtained from Matlab function `fmincon`. However, the calculation time using `fmincon` is one hundred times of that using the quasi-optimum method.

Figure 2 illustrates the optimal points $\boldsymbol{e}_{\text{opt}}$ at $\beta = 0$ and $\overline{\boldsymbol{f}}_i$ as $\beta \to \infty$. The quasi-optimal point $\boldsymbol{f}^*_{\text{opt}}$ is found in the segment connecting two points, and the current fraction vector $\boldsymbol{f}_i$ is updated to $\boldsymbol{f}^*_{\text{opt}}$.

An actual procedure to derive the quasi-optimal point $\boldsymbol{f}^*_{\text{opt}}$ is given as follows. Put $\boldsymbol{f}_i = w\boldsymbol{e}_{\text{opt}} + (1 - w)\overline{\boldsymbol{f}}_i$ in the formula (5) for $0 \le w \le 1$. Omitting suffix i, we define the following target function of w:

$$g(w) = \left\{ w\boldsymbol{e}_{\text{opt}}^T \boldsymbol{\phi} + (1 - w)\overline{\boldsymbol{f}}^T \boldsymbol{\phi} \right\} \exp\{-\beta w^2 \|\boldsymbol{e}_{\text{opt}} - \overline{\boldsymbol{f}}\|^2\}$$

where $\boldsymbol{\phi} = (\phi(\boldsymbol{x}; \boldsymbol{\mu}_1, \Sigma_1), ..., \phi(\boldsymbol{x}; \boldsymbol{\mu}_G, \Sigma_G))^T$. If $\beta\|\boldsymbol{e}_{\text{opt}} - \overline{\boldsymbol{f}}\| = 0$, the optimal value maximizing $g(w)$ is given by $w = 1$. If $\beta\|\boldsymbol{e}_{\text{opt}} - \overline{\boldsymbol{f}}\| > 0$, the optimal value of w can be derived by solving the differential equation $dg(w)/dw = 0$, equivalently,

$$\boldsymbol{e}_{\text{opt}}^T \boldsymbol{\phi} - \overline{\boldsymbol{f}}^T \boldsymbol{\phi} - 2\beta \|\boldsymbol{e}_{\text{opt}} - \overline{\boldsymbol{f}}\|^2 \cdot w \left\{ w\boldsymbol{e}_{\text{opt}}^T \boldsymbol{\phi} + (1 - w)\overline{\boldsymbol{f}}^T \boldsymbol{\phi} \right\} = 0.$$

The solution w is easily found through a root of a quadratic equation for $0 < w < 1$.

4. Proposal of contextual unmixing methods and evaluation

Based on the discussion in Section 2, we propose the following EM algorithm for estimating the fraction vectors.

Contextual unmixing procedure

S1. Fix $\beta > 0$, and a natural number G.

S2. Find the initial estimates, $\boldsymbol{\mu}_g^{(0)}$, $\Sigma_g^{(0)}$ and $\boldsymbol{f}_i^{(0)}$, using normal mixture models under the spatially-independent assumption on the feature vectors $\{\boldsymbol{x}_1, ..., \boldsymbol{x}_n\}$.

S3. Repeat the following steps from $t = 0$ until the convergence of the estimated parameters.

(a) M-step: (Repeat for $i = 1, ..., n$)
Through the method developed in Section 3, find the quasi-optimal fraction $\boldsymbol{f}_i = \boldsymbol{f}_i^{(t+1)}$ which maximizes the following conditional density:

$$p\left(\boldsymbol{f}_i \middle| \boldsymbol{x}_i, \boldsymbol{f}_j^{(t)}, j \sim i\right) \propto \sum_{g=1}^{G} f_{ig} \phi(\boldsymbol{x}_i; \boldsymbol{\mu}_g^{(t)}, \Sigma_g^{(t)}) \exp\left(-\beta \|\boldsymbol{f}_i - \overline{\boldsymbol{f}}_i^{(t)}\|^2\right)$$

where $\overline{\boldsymbol{f}}_i^{(t)} \equiv \sum_{j:j\sim i} \boldsymbol{f}_j^{(t)}/N_i$.

(b) E-step: Update parameters $\boldsymbol{\mu}_g^{(t+1)}$ and $\Sigma_g^{(t+1)}$ by re-estimated fractions $\boldsymbol{f}_i^{(t+1)}$ as

$$\boldsymbol{\mu}_g^{(t+1)} = \frac{\sum_{i=1}^n r_{ig}\boldsymbol{x}_i}{\sum_{i=1}^n r_{ig}}, \quad \Sigma_g^{(t+1)} = \frac{\sum_{i=1}^n r_{ig}(\boldsymbol{x}_i - \boldsymbol{\mu}_g^{(t)})(\boldsymbol{x}_i - \boldsymbol{\mu}_g^{(t)})^T}{\sum_{i=1}^n r_{ig}}$$

where $r_{ig} \equiv f_{ig}^{(t+1)}\phi(\boldsymbol{x}_i \mid \boldsymbol{\mu}_g^{(t)}, \Sigma_g^{(t)})$.

The proposed contextual method is applicable for unmixing as well as clustering. Our method for clustering is evaluated as follows.

1. Apply the unmixing procedure to a data set.
2. Allocate each pixel to the category with the maximum fraction.
3. Calculate the error rate defined by

$$\text{clustering error rate} = \frac{\text{the number of pixels categorized false}}{\text{total number of pixels}}.$$

Furthermore, unmixing is evaluated as follows.

1. Apply the unmixing procedure to a data set.
2. Calculate the error rate defined by

$$\text{unmixing error rate} = \frac{1}{2n}\sum_{i=1}^{n}\sum_{g=1}^{G}|f_{ig} - \hat{f}_{ig}|.$$

It is easily seen that the unmixing error rate coincides with the clustering error rate when fractions $\boldsymbol{f}_i$ and $\hat{\boldsymbol{f}}_i$ are given by one of the edge points $\boldsymbol{e}_1 = (1, 0, ..., 0)^T, \ldots, \boldsymbol{e}_G = (0, ..., 0, 1)^T$.

5. Generalization of distributional assumptions

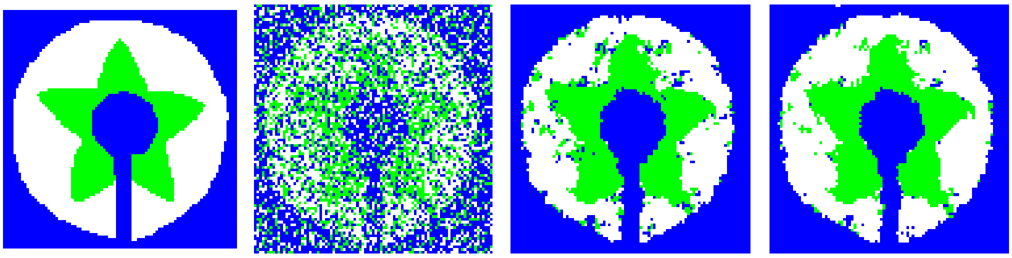
The conditional distribution of feature vector $\boldsymbol{X}$ given fraction vector $\boldsymbol{f}$ is easily extended in a general mixture distribution defined by

$$p(\boldsymbol{x} \mid \boldsymbol{f}) = \sum_{g=1}^{G} f_g h_g(\boldsymbol{x}; \boldsymbol{\theta}_g) \tag{6}$$

where $h_g(\boldsymbol{x}; \boldsymbol{\theta}_g)$ denotes a class-conditional probability density function specified by unknown parameter vector $\boldsymbol{\theta}_g$ for $g = 1, ..., G$. Furthermore, the joint distribution of the fraction vectors is extended as

$$p(\boldsymbol{f}_1, ..., \boldsymbol{f}_n) = \frac{1}{Z}\left\{\frac{\Gamma(\alpha_1 + \cdots + \alpha_G)}{\Gamma(\alpha_1)\cdots\Gamma(\alpha_G)}\right\}^n \prod_{i=1}^{n}\prod_{g=1}^{G} f_{ig}^{\alpha_g - 1} \exp\left\{-\beta \sum_{i}\sum_{\sim j} \|\boldsymbol{f}_i - \boldsymbol{f}_j\|^\gamma\right\} \tag{7}$$

where γ is a positive parameter and $\alpha_g \geqq 1$ denotes a parameter of the Dirichlet distribution. Note that the marginal distribution of the fraction follows the Dirichlet distribution.



(a) True labels (b) $(\beta, e) = (0, 0.419)$ (c) $(r, \beta, e) = (1, 3, 0.106)$ (d) $(r, \beta, e) = (1, 3, \mathbf{0.075})$

Figure 3: Clustering of the synthetic data with three categories. True labels (a), non-contextual clustering (b), and contextual clustering (c). The "r" and "e" respectively denote the radius of neighborhoods and the clustering error rate.

6. Numerical examples

In this section, the method is examined through a synthetic data set and an IEEE benchmark data set provided for supervised classification. For the initial parameter estimation in S2 of Section 4, `EM_GM` of the Matlab Toolbox[*] is used.

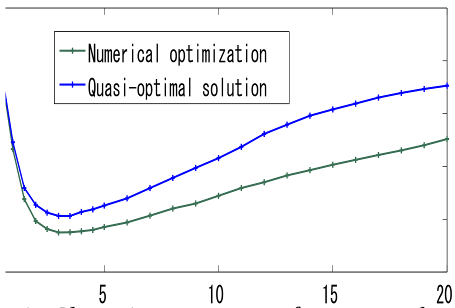


Figure 4: Clustering error rates for two methods against β for the synthetic data (radius $r = 1$)

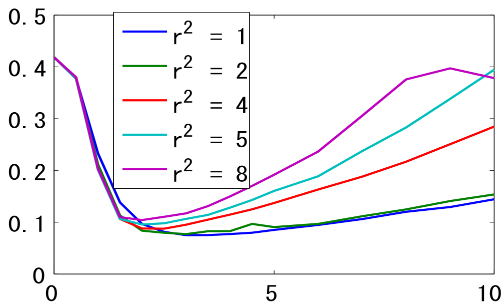


Figure 5: Clustering error rates against β with radiuses r for the synthetic data

6.1. Application to a synthetic data set

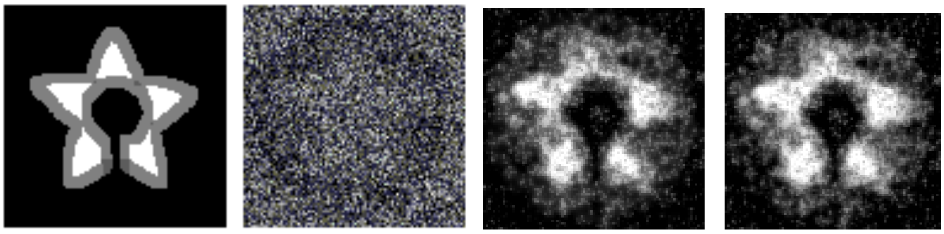
Clustering

The proposed method is applied to four-dimensional data ($d = 4$) generated over the image shown in Figure 3 (a) of size 91×91. There are three categories ($G = 3$) with $n = 8281$ samples following $N_4(\boldsymbol{\mu}_g, I)$ with $\boldsymbol{\mu}_1 = (0\ 0\ 0\ 0)^T$, $\boldsymbol{\mu}_2 = (1\ 1\ 0\ 0)^T / \sqrt{2}$, and $\boldsymbol{\mu}_3 = (1.0498\ -0.6379\ 0\ 0)^T$. (See Nishii and Eguchi (2005) for a detailed explanation of the data.)

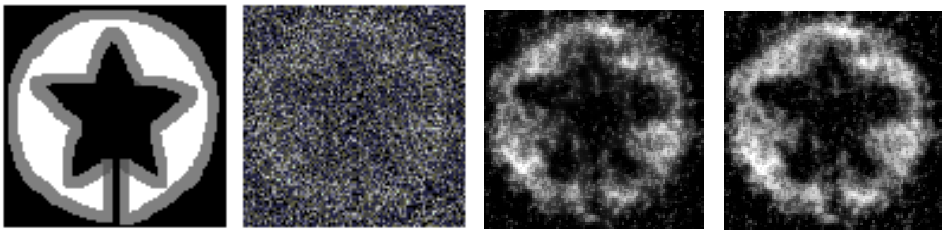
Figure 3 compares non-contextual clustering and contextual clustering. The clustering image (b) obtained by the non-contextual method is very poor, whereas the proposed methods enable very clear images (c) and (d), which were respectively derived by the quasi-optimization and the numerical optimization.

Figure 4 compares two procedures based on the quasi-optimal method and the numerical optimization discussed in Section 3. Clustering error rates are plotted against granularity β with

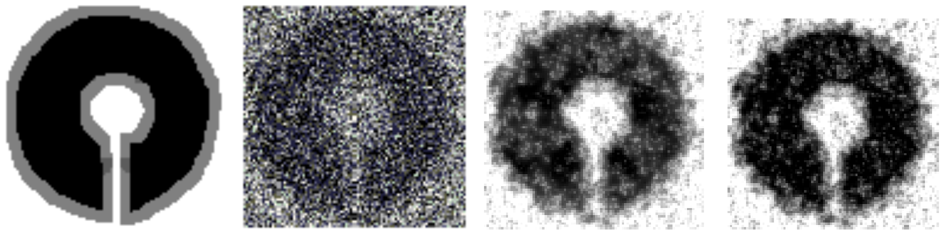
[*] http://www.mathworks.com/matlabcentral/fileexchange/loadCategory.do



(a) True fraction of C_1 (d) $(\beta, e) = (0, 0.464)$ (g) $(\beta, e) = (3, 0.271)$ (j) $(\beta, e) = (3, \mathbf{0.233})$



(b) True fraction of C_2 (e) $(\beta, e) = (0, 0.464)$ (h) $(\beta, e) = (3, 0.271)$ (k) $(\beta, e) = (3, \mathbf{0.233})$



(c) True fraction of C_3 (f) $(\beta, e) = (0, 0.464)$ (i) $(\beta, e) = (3, 0.271)$ (l) $(\beta, e) = (3, \mathbf{0.233})$

Figure 6: Unmixing of synthetic data. Images based on the true fractions f_1, f_2, f_3 (a)-(c), non-contextual unmixing (d)-(f), contextual unmixing (g) - (i) and (j) - (l). The "e" denotes the unmixing error rate.

radius $r = 1$. The non-contextual clustering ($\beta = 0$) with a clustering error rate of 0.210 is greatly improved. The minimum value is attained at $\beta = 3$, giving a clustering error rate of 0.106 by the quasi-optimal method. The numerical optimization is time-consuming, but improves the error rate to 0.075 . This method is superior to the quasi-optimization, as expected, but, the difference is not so large.

Figure 5 depicts clustering error rates against β based on numerical optimization with radiuses $r = 1,\ \sqrt{2},\ 2,\ \sqrt{5},\ \sqrt{8}$. As can be seen, the method with $r = 1$ or $\sqrt{2}$ performed well.

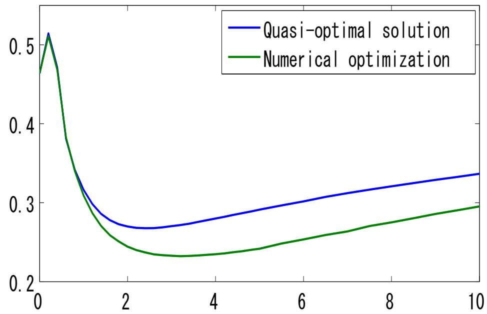


Figure 7: Unmixing error rate against β with radius $r = 1$ for the synthetic data

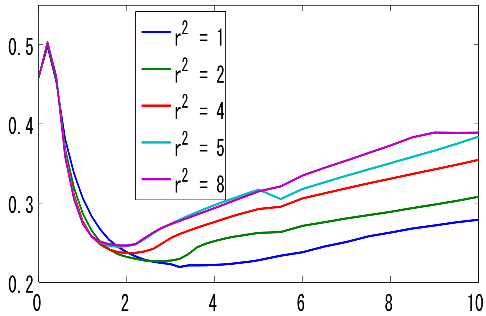


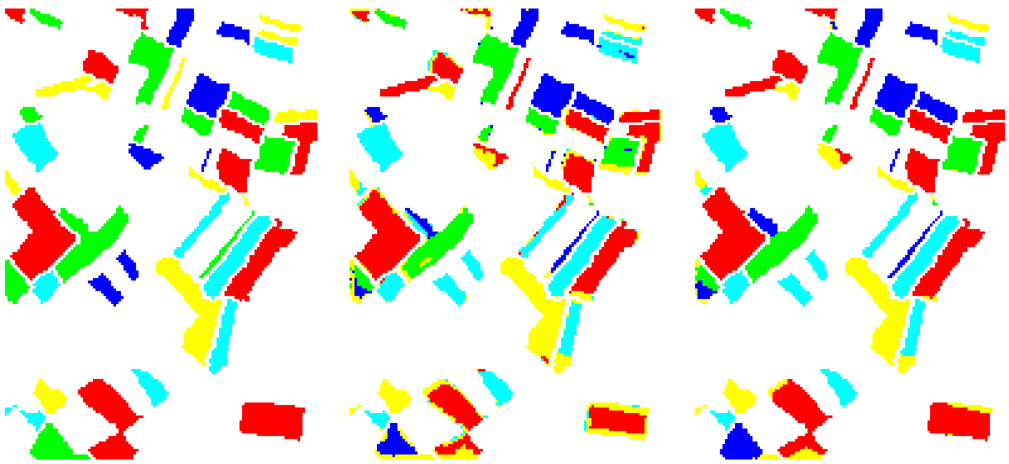
Figure 8: Unmixing error rates against β with radiuses r for the synthetic data

Unmixing

The proposed method can also be applied to four-dimensional data generated by the mixture of the Gaussian populations. Figures 6 (a) - (c) give the true fraction rates of the three categories.

Figures 6 (d) - (f) show poor non-contextual unmixing images with an error rate of 0.464. Figures 6 (g) - (i) show that the contextual unmixing method with quasi-optimization greatly improves image quality. Figures 6 (j) - (l) show that the method based on the numerical optimization gives the best result with an error rate of 0.233.

Figure 7 compares the numerical optimization and the quasi-optimization, revealing that the former is superior. Furthermore, Figure 8 depicts clustering error rates against β based on numerical optimization with five radiuses. The method with $r = 1$ gives the best result.



(a) True labels of five categories (b) $(\beta, e) = (0, 0.267)$ (c) $(r, \beta, e) = (\sqrt{8}, 20, \mathbf{0.187})$

Figure 9: Clustering of the benchmark data. True labels (a), non-contextual clustering (b), and contextual clustering (c). The "r" and "e" respectively denote the radius of neighborhoods and the clustering error rate.

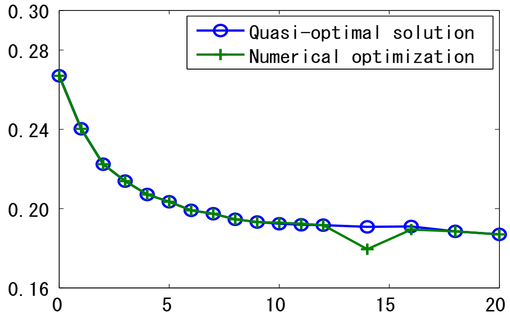


Figure 10: Clustering error rate against β with radius $r = \sqrt{8}$ for the benchmark data

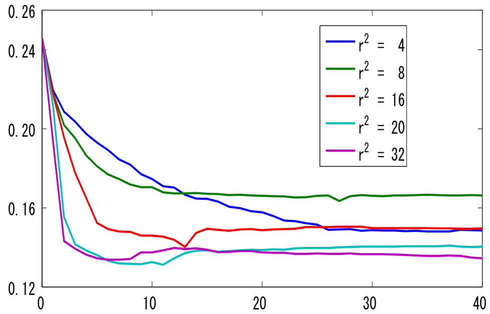
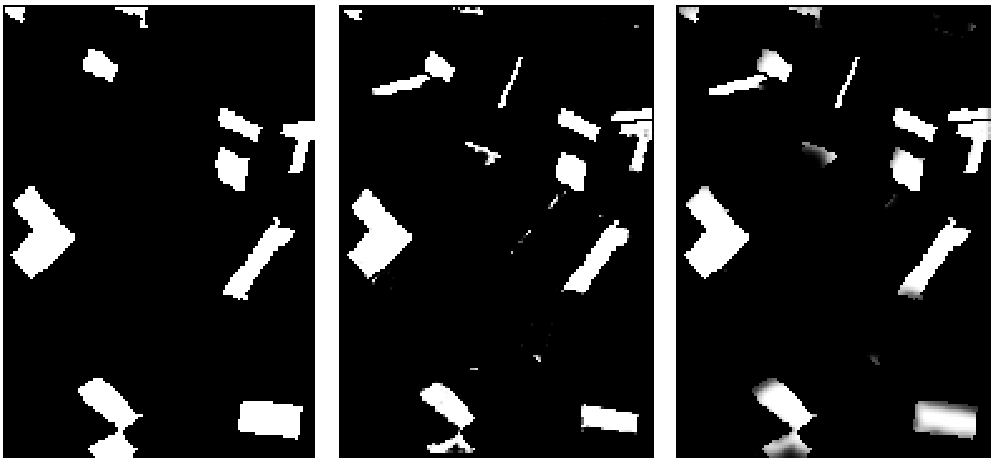


Figure 11: Clustering error rates against β with radiuses r for the benchmark data



(a) True labels of C_1 (b) $(\beta, e) = (0, 0.269)$ (c) $(r, \beta, e) = (\sqrt{8}, 9, \mathbf{0.210})$

Figure 12: Unmixing of the first category "beet". True fraction (a), non-contextual unmixing (b), and contextual unmixing (c). The "e" denotes the unmixing error rate.

6.2. Application to benchmark data set grss_dfc_0006

The data set grss_dfc_0006[†] consists of samples acquired by ATM and SAR ($d = 15$) with five agricultural categories (beet, potato etc, $G = 5$) in Feltwell, UK. The data with $n = 5760$ samples are observed in a rectangular region of size 350×250.

Clustering

Figure 9 gives the true labels with five categories (a), non-contextual clustering (b), and contextual clustering (c). The error rate of non-contextual clustering (b) is improved by the proposed contextual method from 0.267 to 0.187. Figure 10 compares the two optimization procedures. In this case, the quasi-optimal method performs similarly to the numerical optimization, except for the case of $\beta = 14$. This may come from a feature of the data set such that each patch is

[†] The data set grss_dfc_0006 is provided by the IEEE GRS-S Data Fusion Committee.

separated and covered by one of the five categories. Figure 11 depicts clustering error rates against β based on numerical optimization with radiuses $r = 2$, $\sqrt{8}$, 4, $\sqrt{20}$, $\sqrt{32}$. It is seen that the method with $r = \sqrt{20}$ or $\sqrt{32}$ works well. The main reason for a fairly large radius giving small error rates is that all patches with the same category are separated.

Unmixing

Figure 12 gives fractions of the first category "beet". The error rate of non-contextual unmixing (b) is improved by the proposed contextual method from 0.269 to 0.210.

Figure 13 compares the two optimizing procedures with radius $r = \sqrt{8}$. The quasi-optimal solution performs similarly to the numerical optimization method. Note also that the optimal granularity β is around seven, and this is smaller than the optimal value for clustering. Figure 14 depicts unmixing error rates against β based on numerical optimization with five radiuses. Again, the method with $r = \sqrt{20}$ or $\sqrt{32}$ works well.

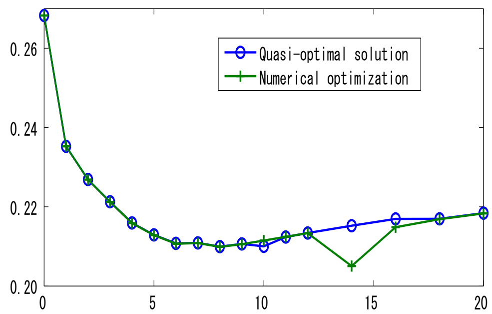


Figure 13: Unmixing error rate against β with radius $r = \sqrt{8}$ for the benchmark data

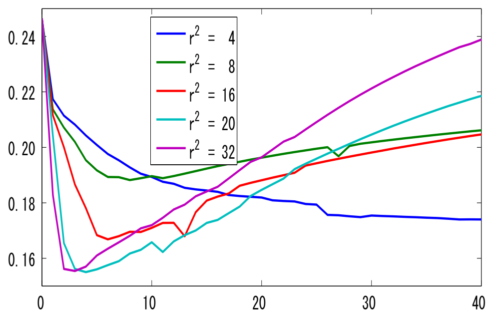


Figure 14: Unmixing error rates against β with radiuses r for the benchmark data

7. Discussion

We have proposed a contextual clustering/unmixing method based on Gaussian mixture and an MRF. Each fraction vector is estimated by maximizing the local conditional posterior density through an EM algorithm. We have also proposed a quasi-optimal procedure to maximize the posterior. It is possible to extend the Gaussian assumption into mixture of general densities. Also, the MRF is extendable into a more general one.

The method was applied to two data sets and evaluated as a contextual clustering/unmixing method. Two data sets discussed in Section 5 were already analyzed by supervised contextual classifiers by Nishii and Eguchi (2005). Naturally enough, our unsupervised method could not exceed the capabilities of their supervised classifiers, but it performs satisfactorily. We also found that contextual information significantly improves clustering/unmixing accuracy.

Our numerical experiments are carried out when the true number G of categories is known. If we choose a small number G, true categories close each other are unified and constitute a new category. This may **no**t lead to severe problems. However, if we choose a large number G, there are several true categories which are divided into two or more. Furthermore, this subdivision affects all the true categories.

Thus, an important future issue is an efficient determination of G as well as the granularity β and the radius r. Furthermore, comparison of the proposed method with existing unmixing methods is also an important issue.

Acknowledgement

The authors are indebted to the editor and reviewers for valuable comments after the careful reading. This research was supported in part by Grants-in-Aid for scientific research (B) 19300096 from the Ministry of Education, Culture, Sports, Science and Technology of Japan.

References

Jia, S. and Qian, Y. (2007). Spectral and spatial complexity-based hyperspectral unmixing, *IEEE Transactions on Geoscience and Remote Sensing*. **45**(12), 3867–3879.

Kawaguchi, S. and Nishii, R. (2007). Hyperspectral image classification by Bootstrap AdaBoost with random decision stumps, *IEEE Transactions on Geoscience and Remote Sensing*. **45**(11), 3845–3851.

Lu, D. and Weng, Q. (2007). A survey of image classification methods and techniques for improving classification performance, *International Journal of Remote Sensing*. **28**(5), 823–870.

Moussaoui, S. et al. (2008). On the decomposition of Mars hyperspectral data by ICA and Bayesian positive source separation, *Neurocomputing Electronic Edition*. **71**(10-12), 2194–2208.

Nascimento, J. M. P. and Dias, J. M. B. (2003). Does independent component analysis play a role in unmixing hyperspectral data? *Pattern Recognition and Image Analysis, ser. Lecture Notes in Computer Science*, F. j. Perales, A. Campilho, and N. P. B. A. Sanfeliu, Eds., **2652**, 616–625.

Nielsen, A. A. (2001). Spectral mixture analysis: linear and semi-parametric full and iterated partial unmixing in multi- and hyperspectral image data, *International Journal of Computer Vision*. **42**(1-2), 17 – 37.

Nishii, R. and Eguchi, S. (2005). Supervised image classification by contextual AdaBoost based on posteriors in neighborhoods, *IEEE Transactions on Geoscience and Remote Sensing*. **43**(11), 2547–2554.

Nishii, R., Kawaguchi, S., Nakamoto, A. and Sawamura, Y. (2008). Contextual unmixing of geospatial data based on Gaussian mixture models and Markov random fields, *Proc. of IEEE IGARSS 2008*, CD-ROM.

Received February 27, 2009
Revised September 18, 2009