# SEMI-SUPERVISED CONTEXTUAL CLASSIFICATION AND UNMIXING OF HYPERSPECTRAL DATA BASED ON MIXTURE DISTRIBUTIONS

*R. Nishii* [1*], *T. Ozaki* [2], *and Y. Sawamura* [3]

[1]Faculty of Mathematics, Kyushu University, Fukuoka, Japan, nishii@math.kyushu-u.ac.jp
[2]Graduate School of Mathematics, Kyushu University, Fukuoka, Japan, ma208035@math.kyushu-u.ac.jp
[3]Graduate School of Mathematics, Kyushu University, Fukuoka, Japan, ma207038@math.kyushu-u.ac.jp

## ABSTRACT

This paper considers image unmixing of hyperspectral data with a small training data set. We propose a semi-supervised contextual unmixing method for hyperspectral data. Gaussian mixture models and a novel MRF (Markov random field) are assumed for distributions of feature vectors and category fraction vectors, respectively. Then, we derive a semi-supervised unmixing method applicable for hyperspectral data through EM algorithm and ICM method. The proposed method is examined through artificial and real data sets, and shows a good performance.

***Index Terms*—** contextual unmixing, Gaussian mixture, MRF, semi-supervised classification

## *1. Introduction*

Contextual image classification is a classification method based on adjacency relationship of the pixels as well as category-conditional spectral distributions. It is known that contextual classifiers derived by Markov random fields (MRF) improve non-contextual classifiers successfully. When training data are unavailable, clustering methods to feature vectors like $K$-means methods are used for detection of homogeneous regions in the image. Furthermore, contextual clustering methods were also proposed, see [1].

Estimation problem of the category fractions covering each mixel is called **unmixing**. It is usually solved by the linear equation derived by the assumption that the observed feature vector is composed by a convex combination of the category reflectance signatures. Other approach, e.g., Independent Component Analysis may be a promising method. Contextual unmixing methods and methods for hyperspectral data have been also proposed, see [2] and [3].

Now, consider the situation such that a low-spatial resolution multispectral image is given, and we are required to estimate fractions of categories in each mixel. Nishii et al. (2008) proposed unsupervised contextual unmixing methods based on Gaussian mixtures for the feature vectors and MRF for the fraction vectors. We will extend the unsupervised contextual unmixing method into the case with a small training data set. This situation arises frequently because collection of training data is expensive and time-consuming. But, the training data are very helpful to obtain an appropriate classification even if they are small. See [4] for a review paper on semi-supervised classification.

## *2. Assumptions for contextual unmixing*

Let $\mathcal{D}$ be a set of pixels in a multispectral image. The pixels are numbered from 1 to $n$, and $\mathcal{D} = \{1, \ldots, n\}$. Note that the pixels are small unit areas in 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 $\boldsymbol{f}_j = (f_{j1}, ..., f_{jG})^t$ denoting **fractions of the categories** $C_g$ covering unlabelled pixel $j$. A set of the fraction vectors is defined as

$$\Delta = \left\{ \boldsymbol{f} = (f_1, ..., f_G)^T \in \mathbb{R}^G \;\middle|\; f_g \geq 0,\ \textstyle\sum_{g=1}^G f_g = 1 \right\}.$$

Note again that we are dealing with the case such that **a small training data set** is given (contextual unmixing or contextual endmember detection). So, suppose that the set of pixels $\mathcal{D}$ is partitioned into two sets: a labelled data set $\mathcal{L} = \left\{(\boldsymbol{x}_i, \boldsymbol{f}_i) \in \mathbb{R}^d \times \Delta \mid i = 1, ..., l\right\}$ and an unlabelled data set $\mathcal{U} = \left\{(\boldsymbol{x}_j, \cdot\,) \in \mathbb{R}^d \times \Delta \mid j = l+1, ..., n\right\}$ where $n = l + u$. Implicitly, we assume that sizes of two sets have the relation $l << u$. Our primary interest is to estimate the un-observed labels $\boldsymbol{f}_j,\ j = l+1, ..., n$.

To estimate the unknown fraction vectors, the following distributional assumptions are made.

**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 distribu-

*This work is supported by Grants-in-Aid for Scientific Research (B) 19300096 from the Ministry of Education, Culture, Sports, Science and Technology, Japan.

tion of feature vector $\boldsymbol{X} = (X_1, ..., X_d)^t$ given fraction vector $\boldsymbol{f}$ is expressed by a Gaussian mixture:

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

where $\phi(\boldsymbol{x}; \boldsymbol{\mu}, \Sigma)$ is a Gaussian probability density function of $N_d(\boldsymbol{\mu}, \Sigma)$.

**Assumption 2:** (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 $\{\boldsymbol{f}_1, ..., \boldsymbol{f}_n\}$ is simply decomposed as

$$p(\boldsymbol{x}_1, ..., \boldsymbol{x}_n \mid \boldsymbol{f}_1, ..., \boldsymbol{f}_n) = \prod_{i=1}^{n} p(\boldsymbol{x}_i \mid \boldsymbol{f}_i). \tag{2}$$

**Assumption 3:** (MRF for category-fraction vectors)

Let $\{\boldsymbol{F}_1, ..., \boldsymbol{F}_n\}$ be a set of random vectors denoting category fractions. We assume that set $\{\boldsymbol{F}_1, ..., \boldsymbol{F}_n\}$ follows MRF with the joint density:

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

where $Z$ is a normalizing factor, and "$i \sim j$" means that a pair of pixels has distance less than $r$. The parameter $\beta \geq 0$ called **a granularity** gives the spatial dependency of MRF. If $\beta = 0$, the fraction vectors $\boldsymbol{F}_1, ..., \boldsymbol{F}_n$ are independent. If $\beta$ is large, the spatial dependency becomes strong.

### 3. *Parameter estimation and unmixing*

Let $L_\lambda = L_\lambda(\boldsymbol{\mu}_1, ..., \boldsymbol{\mu}_G, \Sigma)$ be a weighted sum of two log likelihoods given by

$$L_\lambda(\boldsymbol{\mu}_1, ..., \boldsymbol{\mu}_G, \Sigma) = \sum_{i=1}^{l} \log\left\{\sum_{g=1}^{G} f_{ig} \phi(\boldsymbol{x}_i; \boldsymbol{\mu}_g, \Sigma)\right\} + \lambda \sum_{j=l+1}^{n} \log\left\{\sum_{g=1}^{G} \widehat{f}_{jg} \phi(\boldsymbol{x}_j; \boldsymbol{\mu}_g, \Sigma)\right\}$$

where $0 < \lambda \leq 1$ is a tuning parameter and $n = l + u$.

Small $\lambda$ puts less contribution of the unlabelled data for parameter estimation. In the case with $l << u$, a possible choice is to take $\lambda \leqq l/u < 1$.

The target function $L_\lambda$ to be maximized is used for estimation of Gaussian parameters. Using the estimates, the fraction vectors are estimated by maximizing the posterior density $p(\boldsymbol{f}_{l+1}, ..., \boldsymbol{f}_n | \boldsymbol{x}_{l+1}, ..., \boldsymbol{x}_n)$. This is done by the iterative conditionam modes (ICM) algorithm, see [1]. We employ the method in the semi-supervised case. ICM method and maximization of $L_\lambda$ are repeated until the convergence of the estimated fractions $\widehat{f}_{jg}$.

**Table 1.** Error rates with tradeoff parameter $\lambda$ and training size $l$. Unlabelled data size is $u = 8281$

| $\lambda \cdot l/u$ | $10^{-6}$ | 0.2 | 0.5 | 0.75 | 1 | 1.333 | 2 | 5 | $10^6$ |
|---|---|---|---|---|---|---|---|---|---|
| $l = 60$ | .6803 | .6714 | .6601 | .6582 | .6428 | .6437 | .6369 | .5580 | .5380 |
| $l = 120$ | .5026 | .5009 | .5046 | .5039 | .5026 | .5043 | .5115 | .5217 | .5813 |
| $l = 180$ | .4971 | .4967 | .4984 | .4974 | .4949 | .4967 | .5138 | .5319 | .5773 |
| $l = 240$ | .4714 | .4740 | .4808 | .4925 | .4897 | .4995 | .5087 | .5333 | .5446 |
| $l = 300$ | .4626 | .4628 | .4650 | .4676 | .4689 | .4697 | .4820 | .5060 | .5277 |

### 4. *Numerical experiments*

The proposed contextual unmixing method is applied to the artificial data set in Nishii et al. (2008) with $G = 3$, $d = 4$ and $u = 8281$. Accuracy of unmixing is evaluated by the error rate defined by the averaged value of absolute differences of the true and estimated fractions.

Table 1 examines the effect of the trade-off parameter $\lambda$ and the training data size $l$ to classification accuracy. This table implies the followings: When the training size is medium, an appropriate selection of $\lambda$ improve the classification. When the training size is large, the parameter estimation based on the training data only shows the best accuracy. Furthermore, it is shown that contextual unmixing is superior to non-contextual unmixing.

Similar results are obtained for the IEEE benchmark data set grss_dfc_0006 provided for supervised classification.

### 5. *Discussion*

We have proposed the semi-supervised contextual unmixing method based on Gaussian mixture and MRF. The fraction vectors are estimated by ICM method, and the Gaussian parameters are estimated by the weighted sum of the log-likelihoods. The method is applied to two data sets, and it is shown that contextual information improves the unmixing accuracy significantly.

The most important future task is to derive the optimal determination of the tradeoff parameter $\lambda$ as well as granularity $\beta$. At this stage, it is recommended to take $\lambda = l/u$.

## References


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

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

[3] S. Moussaoui 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.

[4] X. Zhu (2008). "Semi-supervised learning literature survey," *Computer Sciences TR* 1530, University of Wisconsin - Madison.