Enhancement of satellite imagery
based on Markov Random Fields

Yoji MORISAKI* and Ryuei NISHII**

*Graduate School of Engineering,
Hiroshima University, Kagamiyama 1–7–1, Higashi-Hiroshima 739–8521
Email: morisaki@mis.hiroshima-u.ac.jp

**Faculty of Integrated Arts and Sciences,
Hiroshima University, Kagamiyama 1–7–1, Higashi-Hiroshima 739–8521
Email: nishii@mis.hiroshima-u.ac.jp

Abstract — Consider contextual data fusion of multispectral data with different spatial resolutions. We introduce a new method for data fusion based on Markov random fields so as to take contextual information into account. The data of low resolution are corrected by the predictor merged with the contextual clustering result of high-resolution images using Gaussian Markov random fields.

1. Introduction

Data fusion with multispectral images with multiresolutions is an important and basic technology in satellite image analysis. The technology is widely required because many satellites are equipped with sensors of different resolutions. We consider a problem to correct ℓ-dimensional low-resolution images by h-dimensional high-resolution images. Note that both the images may be multivariate.

As an example of such situations, we suppose that multispectral low-resolution data and panchromatic high-resolution data are given, i.e., $\ell \geq 3$ and $h = 1$. For instance, the spatial resolution of visible lights of the satellite IKONOS is of 4m, whereas that of the panchromatic sensor is of 1m. In this case, $\ell = 3$ and $h = 1$. Also, the resolutions of SPOT are respectively 20m and 10m ($\ell = 3$, $h = 1$), and those of LANDSAT 7 are 30m and 15m ($\ell = 6$, $h = 1$).

The commonly-used method for the enhancement of colored images is the Hue-Saturation-Value (HSV) transform [1]. The spectral data corresponding to Red, Green and Blue (RGB) at the lower spatial resolution are converted to HSV, after which the values are replaced by the panchromatic values of the higher resolution. Then, the result is transformed back to RGB, and an enhanced colored image is obtained. However, the HSV method reduces spectral information, and is valid only when the number of images is three. Linear methods based on wavelet transform are also proposed, see e.g., [2]. Many references can be found therein. These methods may keep local information and preserve the spectral information better than the HSV method. Another method for image correction based on non-contextual clustering is proposed by [3]. This method may be powerful, but needs much computation and overfits to the data. A Bayes method is found in a review paper [4]. A fully-statistical approach is also proposed by multivariate normal distributions [5]. They derived a predictor based on the conditional expectation given the high-resolution data. However, this method also overfits to the data. A common drawback of these methods in the above is that spatial information, which is important for remote sensing data, is ignored. To take spatial information of images into account, another statistical approach so-called cokriging is proposed by [6].

In this paper, we propose a new method taking contextual information into account. Markov random fields (MRF) provide a theoretically robust and mathematically tractable way of characterizing contextual information, and are widely used in image segmentation and restoration. See, e.g., Chapter 13 of [7] and references therein for MRF. First, we classify all pixels of high-resolution images into some clusters by Gaussian MRF (GMRF). Then, the value/vector of low resolution is corrected by a predictor based on the information of clusters. Here, we derive predictors based on the conditional expectation given the high-resolution data in the similar way to [5].

In Section 2, we introduce notations and distributional assumptions on spectrum data. Section 3 derives predictors for low resolution based on the conditional distribution. A MRF for contextual clustering is introduced in Section 4. Contextual prediction procedure based on the ICM (iterated conditional modes) algorithm due to [8] is presented in

	2			6	2	5
3	0	1		3	0	1
	4			7	4	8

1	2	3	4
5	6	7	8
9	10	11	12
13	14	15	16

Fig. 1. Four-adjacent-pixels window (W4) and all-square-pixels window (W8) of the center pixel 0.

Fig. 2. All-square-pixels window (W(16)).

Section 5. The proposed procedure is examined through the actual LANDSAT data, and compared with the standard methods in Section 6. Conclusions are given in Section 7.

2. Distributional assumptions on spectrum data

Fig. 1 illustrates two neighborhood of the center pixel 0 consisting of four or eight pixels. The four-adjacent-pixels window and the all-square-pixels window will be referred as W4 and W8 respectively.

We suppose that observations of spectrums with high resolution are available at all pixels, whereas observations of low-resolution spectrums are not available at all pixels. They may be only observed at each window. Our aim is to correct such low-resolution data by the high-resolution data observed at the neighborhood W4 or W8.

Let

$$\boldsymbol{Z}_i = \begin{pmatrix} \boldsymbol{H}_i \\ \boldsymbol{L}_i \end{pmatrix} \; : \; (h + \ell) \times 1, \quad i = 0, 1, \ldots, 8, \tag{1}$$

be a random vector corresponding to spectrum data at the ith pixel, where $\boldsymbol{H}_i$ represents an $h \times 1$ random vector of high resolution, and $\boldsymbol{L}_i$ represents an $\ell \times 1$ random vector of low resolution. Suppose that observations on $\boldsymbol{H}_i$ of high resolution, say $\boldsymbol{h}_j$, are available for $i = 0, 1, \ldots, 8$, whereas each $\boldsymbol{L}_i$ of low resolution is not observed. We regard the average vector $\sum_{j=0}^{4} \boldsymbol{L}_j / 5 \equiv \overline{\boldsymbol{L}}_5$ or $\sum_{j=0}^{8} \boldsymbol{L}_j / 9 \equiv \overline{\boldsymbol{L}}_9$ is only observed at the window W4 or W8. We also assume that each pixel in the window W4 or W8 belongs to one of g possible clusters denoted by $C_1, \ldots, C_g$. A label of the cluster at the pixel i is denoted by y_i. Each y_i is supposed to be a realization of a discrete random variable Y_i. The distribution of the cluster label Y_i is discussed in section 4. In sections 2 and 3, we assume that y_i are known.

Suppose that the conditional distribution of $(\boldsymbol{H}_i', \boldsymbol{L}_i')'$ given $Y_i = y_i$ is an $(h + \ell)$-variate normal distribution with mean vector $(\boldsymbol{\mu}(y_i)', \boldsymbol{\nu}(y_i)')'$ and common variance-covariance matrix $\begin{pmatrix} \Sigma_{hh} & \Sigma_{h\ell} \\ \Sigma_{\ell h} & \Sigma_{\ell\ell} \end{pmatrix}$. This distribution is denoted by

$$\begin{pmatrix} \boldsymbol{H}_i \\ \boldsymbol{L}_i \end{pmatrix} \Big| \, y_i \sim N_{h+\ell} \left(\begin{pmatrix} \boldsymbol{\mu}(y_i) \\ \boldsymbol{\nu}(y_i) \end{pmatrix}, \begin{pmatrix} \Sigma_{hh} & \Sigma_{h\ell} \\ \Sigma_{\ell h} & \Sigma_{\ell\ell} \end{pmatrix} \right) \; \text{for } i = 0, 1 \ldots, 8. \tag{2}$$

Further, we assume that they are independent.

3. Prediction for low resolution based on the conditional expectation

A linear predictor for low resolution is proposed by employing multivariate normal distributions for spectrum data [5]. They derived a predictor based on the conditional expectation given the high-resolution data. By the similar way to their method, we shall derive predictors for low resolution in the windows W4 and W8. We note that our predictors are different from that of [5] in that mean vectors on respective clusters are incorporated.

A. Prediction in four-adjacent-pixels window

Consider the window W4 of Fig. 1 consisting of four pixels. Under the conditional normal assumption (2) and the conditional independence of $(\boldsymbol{H}_i', \boldsymbol{L}_i')'|y_i$ for $i = 0, 1 \ldots, 4$, the joint distribution of the vector $\boldsymbol{Z} \equiv$

$(\boldsymbol{H}_0', \boldsymbol{L}_0', \ldots, \boldsymbol{H}_4', \boldsymbol{L}_4')' : (5h + 5\ell) \times 1$ given $\boldsymbol{y}_5$ is the multivariate normal distribution with mean vector $\boldsymbol{\mu}$ and covariance matrix Σ, where

$$
\boldsymbol{y}_5 = \begin{pmatrix} y_0 \\ y_1 \\ \vdots \\ y_4 \end{pmatrix}, \quad \boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}(y_0) \\ \boldsymbol{\nu}(y_0) \\ \vdots \\ \boldsymbol{\mu}(y_4) \\ \boldsymbol{\nu}(y_4) \end{pmatrix} \quad \text{and} \quad \Sigma = \begin{pmatrix} \Sigma_{hh} & \Sigma_{h\ell} & \cdots & O & O \\ \Sigma_{\ell h} & \Sigma_{\ell\ell} & \cdots & O & O \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ O & O & \cdots & \Sigma_{hh} & \Sigma_{h\ell} \\ O & O & \cdots & \Sigma_{\ell h} & \Sigma_{\ell\ell} \end{pmatrix}. \tag{3}
$$

See, e.g., [9] for the definition and properties of multivariate normal distributions.

We shall derive the joint distribution of $\boldsymbol{L}_0$ and all random vectors whose observations are available. The random vector $(\boldsymbol{H}_0', \boldsymbol{H}_1', \ldots, \boldsymbol{H}_4', 5\overline{\boldsymbol{L}}_5', \boldsymbol{L}_0')' : (5h + 2\ell) \times 1$ is expressed by a linear mapping $A\boldsymbol{Z}$ with a matrix A defined by

$$
A = \begin{pmatrix} I_h & O & O & O & \cdots & O & O \\ O & O & I_h & O & \cdots & O & O \\ \vdots & \vdots & \vdots & \vdots & & \vdots & \vdots \\ O & O & O & O & \cdots & I_h & O \\ O & I_\ell & O & I_\ell & \cdots & O & I_\ell \\ O & I_\ell & O & O & \cdots & O & O \end{pmatrix} : (5h + 2\ell) \times (5h + 5\ell). \tag{4}
$$

The formulas (3) and (4) show that the random vector $A\boldsymbol{Z}$ given $\boldsymbol{y}_5$ follows the multivariate normal distribution with mean vector $A\boldsymbol{\mu}$ and covariance matrix $A\Sigma A'$, i.e.,

$$
\begin{pmatrix} \boldsymbol{H}_0 \\ \vdots \\ \boldsymbol{H}_4 \\ 5\overline{\boldsymbol{L}}_5 \\ \boldsymbol{L}_0 \end{pmatrix} \Big| \boldsymbol{y}_5 \sim N_{5h+2\ell} \left(\begin{pmatrix} \boldsymbol{\mu}(y_0) \\ \vdots \\ \boldsymbol{\mu}(y_4) \\ 5\overline{\boldsymbol{\nu}}_5 \\ \boldsymbol{\nu}(y_0) \end{pmatrix}, \begin{pmatrix} \Sigma_{hh} & \cdots & O & \Sigma_{h\ell} & \Sigma_{h\ell} \\ \vdots & \ddots & \vdots & \vdots & \vdots \\ O & \cdots & \Sigma_{hh} & \Sigma_{h\ell} & O \\ \Sigma_{\ell h} & \cdots & \Sigma_{\ell h} & 5\Sigma_{\ell\ell} & \Sigma_{\ell\ell} \\ \Sigma_{\ell h} & \cdots & O & \Sigma_{\ell\ell} & \Sigma_{\ell\ell} \end{pmatrix} \right), \tag{5}
$$

where $\overline{\boldsymbol{\nu}}_5 = \sum_{j=0}^{4} \boldsymbol{\nu}(y_j)/5$. Using the joint distribution (5), we can derive the conditional distribution of $\boldsymbol{L}_0$ as follows. The covariance matrix $A\Sigma A'$ appearing in (5) is partitioned as $A\Sigma A' = \begin{pmatrix} \Xi_{11} & \Xi_{12} \\ \Xi_{21} & \Sigma_{\ell\ell} \end{pmatrix}$ of size $\Xi_{11} : (5h + \ell) \times (5h + \ell)$ and $\Xi_{12} : (5h + \ell) \times \ell$. After some algebra we have

$$
\Xi_{21}\Xi_{11}^{-1} = \frac{1}{5}(4\Sigma_{\ell h}\Sigma_{hh}^{-1}, -\Sigma_{\ell h}\Sigma_{hh}^{-1}, \ldots, -\Sigma_{\ell h}\Sigma_{hh}^{-1}, I_\ell), \tag{6}
$$

$$
\Sigma_{\ell\ell} - \Xi_{21}\Xi_{11}^{-1}\Xi_{12} = 4\Sigma_{\ell\ell\cdot h}/5, \tag{7}
$$

where $\Sigma_{\ell\ell\cdot h}$ denotes a conditional variance defined by $\Sigma_{\ell\ell\cdot h} = \Sigma_{\ell\ell} - \Sigma_{\ell h}\Sigma_{hh}^{-1}\Sigma_{h\ell}$. From the joint distribution (5) and the relations (6) and (7), the conditional distribution of $\boldsymbol{L}_0$ is derived as

$$
\begin{aligned}
\boldsymbol{L}_0 &\big| \{\boldsymbol{h}_0, \boldsymbol{h}_1, \ldots, \boldsymbol{h}_4, \overline{\boldsymbol{\ell}}_5, \boldsymbol{y}_5\} \\
&\sim N_\ell \left(\overline{\boldsymbol{\ell}}_5 + \Sigma_{\ell h}\Sigma_{hh}^{-1}(\boldsymbol{h}_0 - \overline{\boldsymbol{h}}_5) + \{\boldsymbol{\nu}(y_0) - \Sigma_{\ell h}\Sigma_{hh}^{-1}\boldsymbol{\mu}(y_0)\} - (\overline{\boldsymbol{\nu}}_5 - \Sigma_{\ell h}\Sigma_{hh}^{-1}\overline{\boldsymbol{\mu}}_5), \ 4\Sigma_{\ell\ell\cdot h}/5 \right),
\end{aligned} \tag{8}
$$

where $\overline{\boldsymbol{h}}_5 = \sum_{j=0}^{4} \boldsymbol{h}_j/5$, $\overline{\boldsymbol{\mu}}_5 = \sum_{j=0}^{4} \boldsymbol{\mu}(y_j)/5$.

We propose a predictor for $\boldsymbol{L}_0$ of low resolution based on the conditional distribution (8) by

$$
\widehat{\boldsymbol{L}}_A = \overline{\boldsymbol{\ell}}_5 + \Sigma_{\ell h}\Sigma_{hh}^{-1}(\boldsymbol{h}_0 - \overline{\boldsymbol{h}}_5) + \{\boldsymbol{\nu}(y_0) - \Sigma_{\ell h}\Sigma_{hh}^{-1}\boldsymbol{\mu}(y_0)\} - (\overline{\boldsymbol{\nu}}_5 - \Sigma_{\ell h}\Sigma_{hh}^{-1}\overline{\boldsymbol{\mu}}_5). \tag{9}
$$

The term $\{\boldsymbol{\nu}(y_0) - \Sigma_{\ell h}\Sigma_{hh}^{-1}\boldsymbol{\mu}(y_0)\} - (\overline{\boldsymbol{\nu}}_5 - \Sigma_{\ell h}\Sigma_{hh}^{-1}\overline{\boldsymbol{\mu}}_5)$ denotes a difference between two conditional expectations. The predictor omitting the term is similar to that of [5]. We note that there is no concept of the cluster in their method.

B. Prediction in all-square-pixels window I

Next, we consider spectral data observed in the window W8, see the right hand side of Fig. 1. By the similar discussion in the preceding subsection A, the predictor at the center pixel is given by

$$
\widehat{\boldsymbol{L}}_B = \overline{\boldsymbol{\ell}}_9 + \Sigma_{\ell h}\Sigma_{hh}^{-1}(\boldsymbol{h}_0 - \overline{\boldsymbol{h}}_9) + \{\boldsymbol{\nu}(y_0) - \Sigma_{\ell h}\Sigma_{hh}^{-1}\boldsymbol{\mu}(y_0)\} - (\overline{\boldsymbol{\nu}}_9 - \Sigma_{\ell h}\Sigma_{hh}^{-1}\overline{\boldsymbol{\mu}}_9), \tag{10}
$$

where $\overline{\boldsymbol{h}}_9 = \sum_{j=0}^{8} \boldsymbol{h}_j / 9$, $\overline{\boldsymbol{\mu}}_9 = \sum_{j=0}^{8} \boldsymbol{\mu}(y_j)/9$, $\overline{\boldsymbol{\nu}}_9 = \sum_{j=0}^{8} \boldsymbol{\nu}(y_j)/9$.

C. Prediction in all-square-pixels window II

In this subsection, we reconsider the local window used in [5], which consists of m pixels. Each pixel is numbered from $i = 1$ to m. In the sequel, the local window will be referred as W(m). Fig. 2 shows the local window in the case of $m = 16$, i.e., W(16).

It is assumed that each $\boldsymbol{H}_i$ of high resolution is observed for $i = 1, 2, \ldots, m$. Concerning $\boldsymbol{L}_i$ of low resolution, however, only averaged values $\sum_{j=1}^{m} \boldsymbol{L}_j / m \equiv \overline{\boldsymbol{L}}_m$ are observed in the window W(m). Observations of $\boldsymbol{H}_i$ and $\overline{\boldsymbol{L}}_m$ are respectively denoted by $\boldsymbol{h}_i$ and $\overline{\boldsymbol{\ell}}_m$. Under these assumptions, we derive the predictor for $\boldsymbol{L}_i$ of low resolution by incorporating mean vectors on respective clusters as follows:

$$\widehat{\boldsymbol{L}}_C = \overline{\boldsymbol{\ell}}_m + \Sigma_{\backslash h}\Sigma_{hh}^{-1}(\boldsymbol{h}_i - \overline{\boldsymbol{h}}_m) + \{\boldsymbol{\nu}(y_i) - \Sigma_{\backslash h}\Sigma_{hh}^{-1}\boldsymbol{\mu}(y_i)\} - (\overline{\boldsymbol{\nu}}_m - \Sigma_{\backslash h}\Sigma_{hh}^{-1}\overline{\boldsymbol{\mu}}_m), \tag{11}$$

where $\overline{\boldsymbol{h}}_m = \sum_{j=1}^{m} \boldsymbol{h}_j / m$, $\overline{\boldsymbol{\mu}}_m = \sum_{j=1}^{m} \boldsymbol{\mu}(y_j)/m$, $\overline{\boldsymbol{\nu}}_m = \sum_{j=1}^{m} \boldsymbol{\nu}(y_j)/m$.

4. Spatial configuration due to Markov random field

In this section, we consider a model due to MRF for contextual clustering. We suppose that there is a two-dimensional scene S consisting of n pixels. All pixels in S are numbered from $i = 1$ to n. A feature vector of high resolution at pixel i and its observation are respectively denoted by $\boldsymbol{H}_i : h \times 1$ and $\boldsymbol{h}_i : h \times 1$. It is assumed that each pixel belongs to one of g possible clusters denoted by $C_1, \ldots, C_g$. A label of the cluster covering the pixel i is denoted by Y_i. This implies that Y_i is a discrete random variable taking cluster labels from 1 to g. Put

$$\boldsymbol{H} \equiv \begin{pmatrix} \boldsymbol{H}_1 \\ \vdots \\ \boldsymbol{H}_n \end{pmatrix} : hn \times 1, \quad \boldsymbol{Y} \equiv \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix} : n \times 1 \ \text{and} \ \boldsymbol{Y}_{(i)} \equiv \boldsymbol{Y} \ \text{with deleted} \ Y_i : (n-1) \times 1. \tag{12}$$

We assume that the conditional distribution of $\boldsymbol{H}_i$ given $Y_i = y_i$ is an h-variate normal distribution with mean vector $\boldsymbol{\mu}(y_i)$ and common variance-covariance matrix Σ_{hh}, written as $\boldsymbol{H}_i|y_i \sim N_h(\boldsymbol{\mu}(y_i), \Sigma_{hh})$. We also assume that $\boldsymbol{H}_i|y_i$ for $i = 1, \ldots, n$ are independent. Hence, the conditional density of $\boldsymbol{H}|\boldsymbol{y}$ with $\boldsymbol{y} \equiv (y_1, \ldots, y_n)'$ is given by

$$f(\boldsymbol{h}|\boldsymbol{y}) = \prod_{i=1}^{n} f(\boldsymbol{h}_i|y_i) \equiv \prod_{i=1}^{n} |2\pi\Sigma_{hh}|^{-1/2} \exp\Big[- \{\boldsymbol{h}_i - \boldsymbol{\mu}(y_i)\}' \Sigma_{hh}^{-1} \{\boldsymbol{h}_i - \boldsymbol{\mu}(y_i)\}/2 \Big]. \tag{13}$$

Further, we assume that the cluster-indicator vector $\boldsymbol{Y}$ follows a locally-dependent MRF, i.e., the conditional probability of Y_i given all labels except y_i is given by

$$\Pr\Big\{Y_i = k \mid \boldsymbol{y}_{(i)}\Big\} = \frac{\exp\left\{-\beta \sum_{j \in U_r(i)} \Delta(k, y_j)\right\}}{\sum_{k^*=1}^{g} \exp\left\{-\beta \sum_{j \in U_r(i)} \Delta(k^*, y_j)\right\}} \quad \text{for } \beta, \ r \geq 0, \tag{14}$$

where $\Delta(k, k^*) = \{\boldsymbol{\mu}(k) - \boldsymbol{\mu}(k^*)\}' \Sigma_{hh}^{-1} \{\boldsymbol{\mu}(k) - \boldsymbol{\mu}(k^*)\}$ for $k, k^* = 1, \ldots, g$ denote the squared Mahalanobis distance, $U_r(i)$ indicates a neighborhood of the pixel i forming a disc centered at i with radius r, and $\boldsymbol{y}_{(i)} : (n-1) \times 1$ denotes a realization of $\boldsymbol{Y}_{(i)}$, see (12). Fig. 3 illustrates the neighborhood $U_1(i)$ with radius $r = 1$. The parameter β is a measure of spatial dependency of the MRF.

5. Contextual prediction based on the ICM algorithm

In the framework of the previous section, the problem is to estimate the cluster-indicator vector $\boldsymbol{Y}$ of the scene S by using the observed feature data $\boldsymbol{h}$ of high resolution. The estimation of $\boldsymbol{Y}$ can be performed by taking $\widehat{\boldsymbol{Y}}$ which maximizes the posterior probability

$$\Pr\{\boldsymbol{Y} = \boldsymbol{y} \mid \boldsymbol{h}\}. \tag{15}$$

That is, $\widehat{\boldsymbol{Y}}$ is taken to be the mode of the posterior distribution of $\boldsymbol{Y}$. It is therefore referred to as the maximum a posteriori (MAP) estimate. The maximization of (15) with respect to $\boldsymbol{y}$ is clearly formidable because there are g^n possible values of $\boldsymbol{y}$. Simulated annealing due to [10] can be used to approximate the MAP estimate. However, this algorithm needs much computation. In this paper, we use the ICM algorithm by [8] instead.

The ICM algorithm is based on the maximization of $\Pr\{Y_i = y_i \mid \boldsymbol{h}, \ \boldsymbol{y}_{(i)}\}$ with respect to y_i. If we remove all terms that are independent of y_i, it follows from Bayes' theorem that

$$\Pr\Big\{Y_i = y_i \mid \boldsymbol{h}, \ \boldsymbol{y}_{(i)}\Big\} \propto f(\boldsymbol{h}_i | y_i) \Pr\Big\{Y_i = y_i \mid \boldsymbol{y}_{(i)}\Big\}. \tag{16}$$

Under the assumptions (13) and (14), maximizing (16) is equivalent to minimizing

$$\big\{\boldsymbol{h}_i - \boldsymbol{\mu}(y_i)\big\}' \Sigma_{hh}^{-1}\big\{\boldsymbol{h}_i - \boldsymbol{\mu}(y_i)\big\} + 2\beta \sum_{j \in U_r(i)} \Delta(y_i, y_j). \tag{17}$$

An estimate $\widehat{y}_i$ is obtained based on the target function (17) iteratively. First, an initial estimate $\boldsymbol{y}^{(0)}$ of $\boldsymbol{y}$ is obtained, for example, by using a non-contextual rule. In our case, the K-means clustering algorithm is used for an initial segmentation of the given scene S. See, e.g., [11] for K-means method. Let $\boldsymbol{y}^{(t)}$ be the estimate of $\boldsymbol{y}$ after completion of the t-th cycle of the ICM algorithm. The $(t+1)$-st cycle of the ICM algorithm for the iterative computation of the estimate of $\boldsymbol{y}$ and the prediction for low resolution is performed in the following steps.

1. Expectation-step: The mean vectors $\boldsymbol{\mu}^{(t)}(k)$ and $\boldsymbol{\nu}^{(t)}(k)$ for $k = 1, \ldots, g$ and the common covariance matrix $\begin{pmatrix} \Sigma_{hh}^{(t)} & \Sigma_{h\ell}^{(t)} \\ \Sigma_{\ell h}^{(t)} & \Sigma_{\ell\ell}^{(t)} \end{pmatrix}$ are estimated by using the current estimate $\boldsymbol{y}^{(t)}$, the high-resolution data $\boldsymbol{h}$ and the low-resolution data $\boldsymbol{\ell}$.

2. Maximization-step: At pixel i, $\widehat{y}_i^{(t+1)}$ is chosen to minimize (17) with $\boldsymbol{\mu}^{(t)}(y_i)$ and $\Sigma_{hh}^{(t)}$ for given β and r. After application of the process to each pixel in turn, the current estimate $\boldsymbol{y}^{(t)}$ of $\boldsymbol{y}$ is updated to $\boldsymbol{y}^{(t+1)}$.

3. Prediction-step: The low-resolution values $\boldsymbol{\ell}_i$ at each pixel are corrected consecutively by the predictor (9) or (10) or (11) in the moving window W4 or W8 or W(m).

The algorithm is repeated for a prespecified number of cycles, or until convergence of the corrected images.

We now consider the selection of optimal parameters β and radius r in the locally dependent MRF. In our spatially-dependent case, we calculate the posterior probability at pixel i defined by

$$p_i(k; \beta, \ r) \equiv \Pr\Big\{Y_i = k \mid \boldsymbol{h}, \ \boldsymbol{y}_{(i)}\Big\} = \frac{f(\boldsymbol{h}_i | k) \Pr\Big\{Y_i = k \mid \boldsymbol{y}_{(i)}\Big\}}{\displaystyle\sum_{k^* = 1}^{g} f(\boldsymbol{h}_i | k^*) \Pr\Big\{Y_i = k^* \mid \boldsymbol{y}_{(i)}\Big\}} \quad \text{for} \quad k = 1, \ldots, g. \tag{18}$$

The parameters β and r are optimized by using the posterior probability (18) in the following steps.

1. Take various values for $\beta > 0$ and radius $r > 0$.

   - We estimate the cluster-indicator vector which maximizes the joint density as
   $$\widehat{\boldsymbol{y}}(\beta, \ r) \equiv \ \arg\max_{\boldsymbol{y}} \ \Big[\ f(\boldsymbol{h}|\boldsymbol{y}) \Pr\{\boldsymbol{Y} = \boldsymbol{y}\} \ \Big] \tag{19}$$

   for given β and r. The probability $\Pr\{\boldsymbol{Y} = \boldsymbol{y}\}$ depends on β and radius r. The ICM algorithm is used for this problem as stated above.

   - Calculate posterior probabilities $p_i(\beta, \ r) \equiv p_i(\widehat{y}_i(\beta, \ r); \beta, \ r)$ for $i = 1, \ldots, n$ defined in (18). Then, we put the averaged value by

   $$\overline{p}(\beta, \ r) = \left\{ \prod_{i=1}^{n} p_i(\beta, \ r) \right\}^{\frac{1}{n}}. \tag{20}$$

2. Find the parameters that maximize the average of posterior probabilities by
$$(\widehat{\beta}, \ \widehat{r}) \equiv \ \arg\max_{\beta, \ r} \ \overline{p}(\beta, \ r). \tag{21}$$

Finally, we obtain the best clustering result by $\widehat{\boldsymbol{y}}(\widehat{\beta}, \ \widehat{r})$.
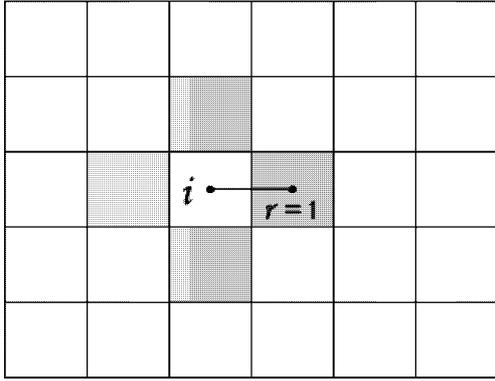
Fig. 3. Neighborhood $U_1(i)$ of a pixel i with radius $r = 1$.
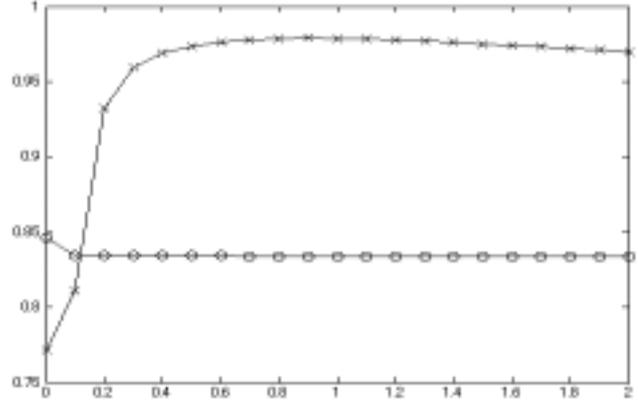
Fig. 4. RSSE (–o–) and $\overline{p}(\beta, \sqrt{8})$ (–×–) versus parameter β.

By using the estimate $\widehat{\boldsymbol{y}}$ of $\boldsymbol{y}$, the mean vectors $\boldsymbol{\mu}(k)$ and $\boldsymbol{\nu}(k)$ for $k = 1, \ldots, g$ and the regression vector $\Sigma_{\cdot h}\Sigma_{hh}^{-1}$ of our predictors are estimated. And then, the low-resolution values at each pixel are corrected by the predictor (9) or (10) or (11) incorporating the mean vectors on respective clusters. After applying the process iteratively, enhanced multispectral images which preserve the spectral information can be obtained.

6. Enhancement of images based on a panchromatic image

Many satellites are equipped with panchromatic sensors whose spatial resolution is finer than other sensors, see (a) and (c) of Fig. 6 for example. The panchromatic images aim to supply supplementary information for the low-resolution images.

We consider the improvement of spatial resolution of low-resolution multispectral images by using a high-resolution panchromatic image. For example, the spatial resolution of visible lights of LANDSAT 7 ETM+ sensor is 30m, and that of the panchromatic sensor is 15m, as noted. Our predictors given by the formulas (9), (10) and (11) are employed for enhancement of the colored image, and compared with the HSV method, the conditional expectation [5] and the cokriging method [6] through LANDSAT 7 images of Hasselt, Belgium taken at Oct. 18, 1999.

In order to compare the results numerically, the original panchromatic image with resolution 15m and the visible bands 1 to 3 with resolution 30m are degraded to 30m and 60m resolution respectively. Using the high-resolution panchromatic image, the low-resolution multispectral images of size 150×150 are corrected. And then, the results are compared based on the ratio of sums of squared errors (RSSE):

$$\text{RSSE} = \frac{\displaystyle\sum_{b=1}^{3}\sum_{i,\,j=1}^{150}\left(\widehat{L}_{ij}^{b} - L_{ij}^{b}\right)^2}{\displaystyle\sum_{b=1}^{3}\sum_{i,\,j=1}^{150}\left(\overline{L}_{ij}^{b} - L_{ij}^{b}\right)^2}, \tag{22}$$

where L_{ij}^{b}, $\overline{L}_{ij}^{b}$ and $\widehat{L}_{ij}^{b}$ for each band $b = 1, 2, 3$ respectively denote original values to be predicted, averaged values of the original values by 2×2 pixels, and the predicted values at pixels of the image under consideration.

Table I tabulates optimal parameters β and maximum averaged posterior probabilities $\overline{p}(\beta, r)$ defined at (20) under the given radiuses r. RSSEs between different methods are also shown in Table I. We note that the number of clusters selected in this application is five, i.e., $g = 5$. This table shows that our method in the window W(4) with $r = \sqrt{8}$ and $\beta = 0.8$ is best. We also see that the best result gives the maximum averaged posterior probability in the third column of the table. Fig. 4 shows RSSEs and averaged posterior probabilities $\overline{p}(\beta, \sqrt{8})$ for varying $\beta = 0.0 : 0.1 : 2.0$. We see from this figure that averaged posterior probability increases as RSSE decreases. These results imply that our strategy for choosing the optimal parameter succeeded in finding the best result.

Fig. 6 shows a portion of the town of Hasselt, which includes urban area, arable land and river. Each figure is enlarged to observe in detail. The figure (c) of Fig. 6 gives a 2×2 averaged image of the original image (b). Using the panchromatic image (a) and the clustering result of Fig. 5, the low-resolution colored image (c) is corrected by the three methods: Our method in the window W(4), conditional expectation due to [5], and HSV show the figures (d), (e) and (f), respectively. We see from the figure (f) that the color of the image is quite different from that of the

TABLE I
The ratios of sums of squared errors due to correction methods
through the images of Hasselt of size 150×150

r	$\widehat{\beta}(r)$	$\overline{p}(\widehat{\beta}(r),\ r)$	RSSE (%)		
			W4	W8	W(4)
1	1.0	50.42	87.83	95.87	84.32
$\sqrt{2}$	0.4	89.49	87.17	95.09	83.51
2	0.6	95.88	87.18	95.12	83.80
$\sqrt{5}$	0.9	97.38	87.01*	94.96**	83.48*
$\sqrt{8}$	0.8	97.87**	87.01	95.00*	83.39**
3	0.6	97.73*	87.00**	95.00*	83.65
$\sqrt{10}$	0.2	95.71	87.04	95.09	83.67
$\sqrt{13}$	1.0	97.51	87.11	95.11	83.70
Conditional expectation			85.73		
Cokriging			88.86		
HSV			261.84		

** and * denote the best and the second best values.
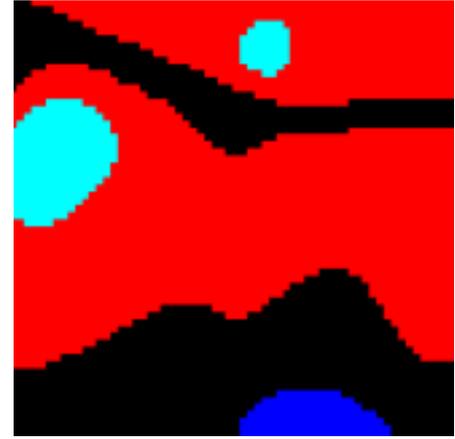
Fig. 5. Clustering result ($g = 5$, $r = \sqrt{8}$, $\beta = 0.8$).

(a) Panchromatic image

(b) Original color image

(c) 2×2 averaged image

(d) Our method W(4) ($g = 5$, $r = \sqrt{8}$, $\beta = 0.8$)

(e) Conditional expectation

(f) HSV

Fig. 6. Data fusion of the panchromatic image (a) and the low-resolution colored image (c) based on our method (d), the conditional expectation (e) and the HSV method (f). The figure (c) is derived by 2×2 averaging procedure of the true image (b). LANDSAT 7 images of size 60×60 courtesy ESA 1999 – distribution Eurimage, Hasselt, Belgium.

averaged image (c). This means that spectral degradation has been caused by the HSV method. This point can be also confirmed by Table I, because the assessed value due to HSV is poor. On the other hand, figures (d) and (e) give sharpened images preserving the spectral characteristics of the averaged image (c).

7. Concluding Remarks

In this paper, the use of GMRF for contextual data fusion of multispectral data with different spatial resolutions has been studied. We proposed a new method for data fusion based on GMRF. The features of our procedure are as follows.

- The predictors which incorporate mean vectors on respective clusters are obtained based on the conditional normal distribution. They are applicable for any dimensions of spectrum data.

- The contextual clustering based on the ICM algorithm is performed. Then, the low-resolution images are corrected by our predictor combined with the clustering result.

- The optimal parameters β and radius r in the locally dependent MRF are successfully selected by using the averaged posterior probability $\overline{p}(\beta, r)$ as a criterion of the selection.

Finally, the most important but difficult problem is to estimate the regression vector $\Sigma_{\cdot h} \Sigma_{hh}^{-1}$ of our predictors because the fluctuations in $\boldsymbol{\ell}_i$ are much smaller than in $\boldsymbol{L}_i$. The predictor based on the better estimate for the regression vector would improve our approach.

Acknowledgment

References

[1] A. R. Smith, "Color gamut transformation pairs," Computers Graphics, vol. 12, pp. 12–19, 1978.

[2] J. Núñez, X. Otazu, O. Fors, A. Prades, V. Palà, and R. Arbiol, "Multiresolution-based image fusion with additive wavelet decomposition," IEEE Trans. Geosci. Remote Sensing, vol. 37, no. 3, pp. 1204–1211, 1999.

[3] B. Zhukov, D. Oertel, and F. Lanzl, "A multiresolution multisensor technique for satellite remote sensing," Proc. 1995 Int. Geosci. Remote Sensing Symp., vol. I, pp. 51–53, 1995.

[4] B. Zhukov, D. Oertel, F. Lanzl, and G. Reinhäckel, "Unmixing-based multisensor multiresolution image fusion," IEEE Trans. Geosci. Remote Sensing, vol. 37, no. 3, pp. 1212–1226, 1999.

[5] R. Nishii, S. Kusanobu, and S. Tanaka, "Enhancement of low spatial resolution image based on high resolution bands," IEEE Trans. Geosci. Remote Sensing, vol. 34, no. 5, pp. 1151–1158, 1996.

[6] Y. Morisaki and R. Nishii, "Spatial enhancement of thermal infrared image based on cokriging," Proc. Symp. 22nd Remote Sensing for Environmental Sciences, pp. 23–28, 2000.

[7] G. J. McLachlan, Discriminant Analysis and Statistical Pattern Recognition. New York: Wiley, 1992.

[8] J. Besag, "On the statistical analysis of dirty pictures," J. R. Stat. Soc. B, vol. 48, pp. 259–302, 1986.

[9] C. R. Rao, Linear Statistical Inference and Its Applications, 2nd ed. New York: Wiley, 1976, ch. 8, Section A.

[10] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distribution, and Bayesian restoration of images," IEEE Trans. Pattern Anal. Machine Intell., vol. PAMI-6, pp. 721–741, 1984.

[11] A. K. Jain and R. C. Dubes, Algorithms for Clustering Data. Englewood Cliffs, NJ: Prentice-Hall, 1988.