A modular eigen subspace scheme for high-dimensional data classification

Yang-Lang Chang\textsuperscript{a,*}, Chin-Chuan Han\textsuperscript{b}, Fan-Di Jou\textsuperscript{c}, Kuo-Chin Fan\textsuperscript{c}, K.S. Chen\textsuperscript{d}, Jeng-Horng Chang\textsuperscript{e}

\textsuperscript{a} Department of Computer Science and Information Engineering, St. John’s and St. Mary’s Institute of Technology, Taipei, Taiwan, ROC
\textsuperscript{b} Department of Computer Science and Information Engineering, Chung Hua University, Hsinchu, Taiwan, ROC
\textsuperscript{c} Institute of Computer Science and Information Engineering, National Central University, Chung-Li, Taiwan, ROC
\textsuperscript{d} Center for Space and Remote Sensing Research, National Central University, Chung-Li, Taiwan, ROC
\textsuperscript{e} Department of Information Management, Tatung College, Taipei, Taiwan, ROC

Available online 6 February 2004

Abstract

In this paper, a novel filter-based greedy modular subspace (GMS) technique is proposed to improve the accuracy of high-dimensional data classification. The proposed approach initially divides the whole set of high-dimensional features into several arbitrary number of highly correlated subgroups by performing a greedy correlation matrix reordering transformation for each class. These GMS can be treated as not only a preprocess of GMS filter-based classifiers but also a unique feature extractor to generate a particular feature subspaces for each different class presented in high-dimensional data. The similarity measures are next calculated by projecting the samples into different modular feature subspaces. Finally, a GMS filter-based architecture based on the mean absolute errors criterion is adopted to build a non-linear multi-class classifier. The proposed GMS filter-based classification scheme is developed to find non-linear boundaries of different classes for high-dimensional data. It not only significantly improves the classification accuracy but also dramatically reduces the computational complexity of feature extraction compared with the conventional principal components analysis. Experimental results demonstrate that the proposed GMS feature extraction method suits the GMS filter-based classifier best as a classification preprocess. It significantly improves the precision of high-dimensional data classification.

Keywords: Greedy modular subspaces; High-dimensional data classification; Mean absolute errors; Principal components analysis

1. Introduction

With the evolution of remote sensing technology, an increasing number of spectral bands becomes available. Data can be collected in a few multispectral bands to as many as several hundreds hyperspectral bands, even thousands of ultraspectral bands. A lot of attention has been focused on the developing of high-dimensional classification devoted to earth remote sensing. The increment of such high-dimensional data greatly enhances the information content, but provides a challenge to the current techniques for analyzing such data [1].

The most common problem of high-dimensional classification is how to achieve the best class separability without the restriction of limited number of training samples [2,3]. There are numerous techniques developed for feature extraction to reduce the
dimensionality without losing class separability. Most of them focus on the estimation of statistics at full dimensionality to extract the classification features. For example, the conventional principal components analysis (PCA) assumes the covariances of different classes are equal and the potential differences between class covariances are not utilized. In contrast, we propose a modular feature subspace method based on a dimension reduction preprocessing scheme, a greedy modular subspace (GMS) method, to overcome the dependency on the global statistics as much as possible and take the advantage to preserve the potential separability of different classes. The GMS provides a very good discrimination among the classes of interest using the idea of modular eigenspaces description which had been proved to have very high recognition rates in the face recognition community [4]. The approach uses a greedy condensed correlation coefficient matrix reordering transformation to find a set of high correlated GMS. Reordering the bands regardless of the original order of wavelengths and spectrally adjacent bands in high-dimensional data set is an important characteristic of GMS. The proposed GMS approach divides the whole set of high-dimensional features into several arbitrary number of highly correlated subgroups. It makes good use of these highly corrected band subgroups to speed up the computational time of PCA.

After finding a set of the most correlated modular subgroups \( \Phi_k \), an eigenspace projection similarity measure is next executed. Consequently, a GMS filter-based classifier can be constructed by these distinguishable GMS sets. Finally, a classification map is obtained by calculating the distances between test samples and tanning samples using the GMS filter-based classifier. Fig. 1 illustrates the flow diagram of this scheme. There are some high-dimensional data feature extraction and reduction techniques presented recently. Jimenez and Landgrebe proposed a projection pursuit method to utilize a projection index function to select potential interesting projections by the local optimization-over-projection directions of the index of performance [2]. A best-bases feature extraction algorithm highlighted by Kumar et al. used an extended local discriminant bases method to reduce the feature spaces [3]. Both of them used the pairwise classifier to divide multi-class problem into a set of two-class ones. In this paper, we introduce a new non-linear filter-based multi-class classification scheme recently proposed in our previous works [5–7] to enhance the precision of image classification significantly by exploring the capabilities of multi-class classifier rather than pairwise one. The main characteristic of GMS filter-based classifier is its exhaustive discrete and non-linear properties. Compared with conventional feature extraction methods, the features of GMS are very suitable for filter-based classifier. These facts will be best demonstrated in Section 3.

The rest of this paper is organized as follows. The proposed supervised GMS filter-based classification scheme is presented in Section 2. In Section 3, experimental results are demonstrated to verify the feasibility and validity of the proposed approach. Finally, concluding remarks are given in Section 4.

2. Methodology

2.1. PCA and high-dimensional data analysis

PCA has become the mainstream algorithm to advance the progress of high-dimensional data analysis in many fields such as noise-whitened, feature extraction, dimensionality reduction, data compression and target discrimination [8,9]. The goal of PCA is...
to find the principal components with their directions along the maximum variances of a data matrix. Chang et al. demonstrated a band selection method that used a loading factors matrix eigen-analysis-based prioritization and a divergence-based band decorrelation [8]. It fully exploited the usefulness of eigen-analysis for feature extraction in high-dimensional data classification.

The conventional low-dimensional classification methods that used single-pixel-based similarity measures such like Gaussian maximum likelihood and minimum Euclidean distance do not work well in high-dimensional classification [10]. Instead, the spectral correlation of pixel vectors for high-dimensional data sets can provide richer features to improve the classification accuracy. In addition, Chang and Chiang [11] showed that the use of second-order statistics, such as PCA that utilizes covariance matrix, can improve hyperspectral classification performance more efficiently. In this paper, we also apply the second-order statistics to the proposed GMS method to improve the classification accuracy.

2.2. Correlation matrix pseudo-color map

A visual scheme to display the magnitude of correlation matrix for emphasizing the second-order statistics of high-dimensional data was proposed by Lee and Landgrebe [12]. Fig. 2 shows a correlation matrix pseudo-color map (CMPM) in which the gray scale is used to represent the magnitude of its corresponding correlation matrix. It is also equal to a modular subspace set \( \Phi^k \). Different material classes have different value sequences of correlation matrices. It can be treated as the special sequence codes for feature extractions.

We define a correlation submatrix \( c_{\Phi_k}(m_l \times m_l) \) which belongs to the \( l \)th modular subspace \( \Phi^k_l \) of a complete set \( \Phi^k = (\Phi^k_1, \ldots, \Phi^k_l, \ldots, \Phi^k_{n_k}) \) for a class \( \omega_k \), where \( m_l \) and \( n_k \) represent, respectively, the number of bands (feature spaces) in modular subspace \( \Phi^k_l \) and the total number of modular subspaces for a complete set \( \Phi^k \), i.e. \( l \in \{1, \ldots, n_k\} \) as shown in Fig. 2. The original correlation matrix \( c_X(m_t \times m_t) \) is decomposed into \( c_{\Phi_1}(m_1 \times m_1), \ldots, c_{\Phi_l}(m_l \times m_l), \ldots, c_{\Phi_{n_k}}(m_{n_k} \times m_{n_k}) \) for the class \( \omega_k \). There are \( m_t! \) (the factorial of \( m_t \)) possible combinations to construct one complete set exactly. The \( m_l \) represents the total number of original bands:

\[
m_t = \sum_{l=1}^{n_k} m_l, \tag{1}
\]

where \( m_l \) represents the \( l \)th feature subspace of a complete set for a class \( \omega_k \).

![Correlation matrix pseudo-color map (CMPM)](image)

![Correlation matrix](image)

Fig. 2. An example illustrates a CMPM with different gray levels and its corresponding correlation matrix with different correlation coefficients in percentage (white = 100; black = 0) for the class \( \omega_k \). Note that four squares with fine black borders represent the highly correlated modular subspaces which have higher correlation coefficient compared with their neighborhood.
2.3. Greedy modular subspaces

We define a complete modular subspace (CMS) set which is composed of all possible combinations of the CMPM. The original correlation matrix \( c_{\chi k}(m_t \times m_t) \) is decomposed into \( n_k \) correlation submatrices \( c_{\Phi k l}(m_1 \times m_1), \ldots, c_{\Phi k l}(m_{l_1} \times m_{l_1}), \ldots, c_{\Phi k l}(m_{l_nk} \times m_{l_nk}) \) to build a CMPM for the class \( \omega_k \). There are \( m_t! \) different kinds of CMPMs in a CMS set as shown in Fig. 3. Each different CMPM is associated with a unique sequence of band order. It needs \( m_t! \) swapping operations by band order to find a complete and exhaustive CMS set. In Fig. 4, a visual interpretation is introduced to highlight the relations between swapping and rotating operations in terms of band order. The swapping operations which exchange the horizontal and vertical correlation coefficient lists row-by-row and column-by-column simultaneously in the correlation matrices \( c_{\chi k l} \) is equivalent to the behavior of rotating.

There is one optimal CMPM in a CMS set as shown in Fig. 3. This optimal CMPM is defined as a specific CMPM which is composed of a set of modular subspaces \( \Phi^k_l, l \in \{1, \ldots, l_1, \ldots, n_k\} \in \Phi^k \). It has the highest correlated relations inside each individual modular subspace \( \Phi^k_l \). It tends to reach the condition that the high correlated blocks (with high correlation coefficient values) are put together adjacently, as near as possible, to construct the optimal modular subspace set \( \Phi^k = (\Phi^k_1, \ldots, \Phi^k_{nk}) \) in the diagonal of CMPM. It is too expensive to make an exhaustive computation for a large amount of \( m_t \) to find an optimal CMPM in a CMS set. In order to overcome this drawback, we develop a fast searching algorithm, called greedy modular subspaces transformation (GMST), based on the fact that highly correlated bands often appear adjacently for remote sensing high-dimensional data sets [13] to construct an alternative greedy CMPM (modular subspace) instead of the optimal CMPM. This new greedy CMPM is defined as a GMS feature module set \( \Phi^k = (\Phi^k_1, \ldots, \Phi^k_{i}, \ldots, \Phi^k_{nk}) \). It cannot only reduce the redundant operations in finding the greedy CMPM but also provides an efficient method to construct GMS sets which have very suitable properties for feature extractions of high-dimensional data sets.

![Fig. 3. (a) The initial CMPM for four original bands (A, B, C and D), \( m_t = 4 \), is applied to the exhaustive swapping operations by band order. (b) An example shows a CMS set which is composed of 24 (\( m_t! = 4! \)) possible CMPMs. The CMPM with a dotted-line square is the optimal CMPM in a CMS set.](image-url)
Fig. 4. The rotation operation between the Block \( K \) and Block 2 is performed by swapping their horizontal and vertical correlation coefficient lists row-by-row and column-by-column. Note that a pair of blocks (squares) switched by this swapping operation in terms of band order should have the same size of any length along the diagonal of correlation matrices. The Fixed Blk 1 and Fixed Blk 2 will rotate 90° at the same locations.

In this algorithm, every positive correlation coefficient \( c_{i,j} \), which is an absolute value, is compared with a correlation coefficient threshold value \( 0 \leq t_c \leq 1 \) by band order adjacently. A greedy searching iteration is initially carried out at \( c_{0,0} \in c_{\Phi_k}^{(m \times m)} \) as a new seed to build a GMS set. All attributes of \( c_{\Phi_k}^{(m \times m)} \) are first set as available except \( c_{0,0} \). The proposed GMST is stated as follows.

- **Step 1.** GMST moves the “current” \( c_{i,j} \) to the next adjacent column \( c_{i,j+1} \) which will act as the “current” \( c_{i,j} \); i.e., \( c_{i,j} \rightarrow c_{i,j+1} \). A new greedy modular feature subspace \( \Phi_k^0 \) is initialized by a seed correlation coefficient \( c_{d,d} \) if the last column index is reached (i.e., \( j = m-1 \)), where \( d \) is the first “available” row or column subindex in the diagonal list of the correlation matrix \( [c_{1,1}, \ldots, c_{m-1,m-1}] \). Thus a modular feature subspace \( \Phi_k^0 \) is constructed with all “used” correlation coefficients and a new greedy modular feature subspace \( \Phi_k^{j+1} \) operation is next performed iteratively with a new seed correlation coefficient \( c_{i,j} = c_{d,d} \) which is the first “available” row or column subindex in the diagonal list of the correlation matrix \( [c_{1,1}, \ldots, c_{m-1,m-1}] \).

- **Step 2.** If \( c_{i,j} \) is larger than \( t_c \), then go to step 3. Otherwise, go to step 1.

- **Step 3.** Simultaneously, Swap \( c_{\ast,j} \) with \( c_{d,\ast} \) and \( c_{j,\ast} \) with \( c_{d,\ast} \), respectively, where \( d \) is the first “available” row or column subindex in the diagonal list of the correlation matrix \( [c_{1,1}, \ldots, c_{m-1,m-1}] \). The asterisk symbol “\( \ast \)” indicates any row or column subindex in the correlation matrix. Figs. 4 and 5 show this mechanism graphically. The attributes of \( c_{d,\ast} \) and \( c_{\ast,d} \) are then marked “used”. Then let \( c_{i\leftrightarrow d,\ast} \) and \( c_{\ast,d\leftrightarrow i} \in \Phi_k^j \), where \( i \leftrightarrow d \) means including all coefficients between subindex \( i \) and subindex \( d \). Go to step 1.

Eventually, a GMS set \( \Phi_k = (\Phi_k^1, \ldots, \Phi_k^j, \ldots, \Phi_k^n) \) is composed. For convenience, we reorder every \( \Phi_k^j \) in \( \Phi_k \) according to their amount of feature bands in
order. Based on the properties of correlation matrix, the GMST attempts to reorder $c_{ik}(m_t \times m_t)$ by directly swapping the positions of $c_{ij}$, rows by rows and columns by columns symmetrically regardless of the original bands order in the CMPM.

In this visualization scheme, we can calculate the GMS efficiently and bypass the redundant procedures of reordering the original row data order before a new CMPM is composed. The computational complexity for finding the optimal CMPM by exhaustive swapping operation is on the order of $O(m_t!)$ and it is $O((m_t) \times (m_t - 1)/2)$ for constructing a GMS based on the GMST, where $m_t$ means the total number of the bands. For example, there are $m_t (=35)$ bands of test data sets in our experiments and the total number of the candidate CMPMs in a CMS set is $35! (=1.033 \times 10^{67})$. In the same condition, the total number of the candidate GMS is $(35 \times 34)/2 (=595)$. The candidate CMPMs in a CMS set is almost $1.736 \times 10^{37}$ times the computational complexity of
the GMS. Fig. 5 illustrates the original CMPM and the reordering CMPM after the GMST.

This GMST preserves the original information of a correlation matrix. The whole set of high-dimensional data is first divided into a highly correlated GMS set \( \Phi^i \) for class \( c_i \) regardless of the original band’s order. Every different class has its singular ordering sets of features for different classes to overcome this problem and improves the classification accuracy.

The proposed GMS method provides a unique feature for a distinguishable class in a specific type of high-dimensional sensor systems. Let \( m_i \) be the number of feature bands in \( \Phi^i \). Each \( \Phi^i \) square \( (m_i \times m_i) \) is filled with an average value of the correlation coefficients \( c_{ij} \) inside the \( \Phi^i \) as shown in Fig. 5(c). Furthermore, GMS reduces more PCA computations than the conventional PCA. The eigen-decomposition complexity for PCA is \( O(m \times m) \) and that for GMS is \( O(\sum_{i=1}^n m_i^2) \) [14]. Compared to the conventional PCA, it not only dramatically improves the computational complexity but also consequently increases the accuracy of high-dimensional data classification. Even the redundancy reduction property of PCA, it not only dramatically improves the computational complexity but also consequently increases the accuracy of high-dimensional data classification. Furthermore, GMS reduces more PCA computations than the conventional PCA. The eigen-decomposition complexity for PCA is \( O(m \times m) \) and that for GMS is \( O(\sum_{i=1}^n m_i^2) \) [14].

There are some merits of using this scheme. (1) It downsizes the number of bands in each GMS to speed up the computation compared to conventional PCA. (2) High correlated GMS makes PCA work more efficiently due to the redundancy reduction property of PCA. (3) The highly related modular features for the dedicated training class tends to have a regular similar variance rather than the whole set of bands if the scale of the data is not well calibrated. It takes the advantage of avoiding the bias problems occurred by conventional PCA [14]. (4) Most classifiers seek only one feature space: the principal subspace \( W \) feature spaces) \( \Phi^i = \{\Phi^j\}_{j=1}^W \) and orthogonal complement one \( \Phi^{i'} = \{\Phi^{j'}\}_{j=W+1}^W \), where \( \Phi^{j'} \) is the \( j \)-th eigenvector of \( \Phi^j \). A residual reconstruction error (RRE) is defined as:

\[
\epsilon_k(x) = \sum_{j=W+1}^W y_j^2 = ||x||^2 - \sum_{j=1}^W y_j^2,
\]

where \( \bar{x} = x - \bar{x} \) is the mean-normalized vector of sample \( x \) and \( y_j \) is the projected value of \( x \) by the eigenvector \( \Phi^{j'}_i \). Here, \( \epsilon_k(x) \) represents the distance between the query sample \( x \) and a GMS \( \Phi^i \) of the training samples. We multiply the eigen percentage \( w_i^j(x) \) by \( \epsilon_k(x) \) to preserve the weight of variance after KLT. A weighted RRE is calculated as follows:

\[
e_k(x) = \sum_{j=W+1}^W w_i^j(x)\epsilon_k(x),
\]

where \( w_i^j(x) \) is an eigen percentage of \( \Phi^j_i \). The weighted RRE vector \( e_k^i(x) = (e_k^1(x), \ldots, e_k^W(x), \ldots) \) is then generated. The drawing interpretation is depicted in Fig. 6.

### 2.4. Modular eigen subspace projection

The GMS can effectively utilize the Karhumen–Loève Transform (KLT) to extract the most valuable information from the highly correlated feature of GMS due to the redundancy reduction property of PCA. Moghaddam et al. [15] decomposed the vector space \( \mathbb{R}^m \) into two exclusive and complementary subspaces: the principal subspace \( W \) feature spaces)

\[
\Phi^i = \{\Phi^j\}_{j=1}^W \text{ and orthogonal complement one } \Phi^{i'} = \{\Phi^{j'}\}_{j=W+1}^W,
\]

where \( \Phi^{j'} \) is the \( j \)-th eigenvector of \( \Phi^j \). For the class \( k \), the reordering CMPM after the GMST.

### 2.5. GMS filter-based classifier and stack filter

The positive Boolean function (PBF) has been successfully applied on stack filter. Each stack filter
corresponding to a PBF possesses the weak superposition property known as the threshold decomposition and the ordering stacking property [16–18]. Wendt et al. [16] defined the stack filters as the class of all filters and also pointed out the connection between stack filters and PBF. The optimal PBF is designed as a classifier by minimizing the mean absolute errors (MAE) among the training samples. The GMS filter-based classifiers are found from the summation of the absolute errors incurred at optimal PBF classifier. The MAE can be calculated as

\[ \text{MAE} = \frac{1}{L} \sum_{i=1}^{L} |d(z_{ij,k}) - S_{f}(z_{ij,k})| \]

where a stack filter \( S_{f} \) is defined as a filter whose decision property known as the threshold decomposition. Based on the threshold decomposition property, an occurrence of vector \( z_{ij,k} \) as the input of PBF can be decomposed into \( L - 1 \) levels binary vectors \( z_{ij,k}^{l} = (z_{ij,k,1}^{l}, z_{ij,k,2}^{l}, \ldots, z_{ij,k,L}^{l}), l \in \{1, \ldots, L - 1\} \). Let us consider two samples \( x_{ij}, \tilde{x}_{ij} \) and a specific class \( w_{k} \). According to the PBF criteria [7], an estimator Boolean function \( f \) is defined as an occurrence \( z_{ij,k}^{l} \) at level \( l \). If \( z_{ij,k}^{l} \leq z_{ij,k}^{l}, l \in \{1, \ldots, L - 1\}, \) then \( f(x) \leq f(y) \). The \( MN^{2} \) occurrences are treated as the training occurrences of the filter-based classifier. They are decomposed into \( MN^{2} \) \((L - 1)\) binary vectors of length \( n \). The desired value for each occurrence \( z_{ij,k} \) is determined by Eq. (4) and the MAE can be obtained by

\[ \text{MAE} = E[|d(z_{ij,k}) - S_{f}(z_{ij,k})|] \]

\[ = E \left[ \sum_{l=1}^{L-1} |d(z_{ij,k}^{l}) - f(T_{l}(z_{ij,k}^{l}))| \right] \]

\[ = E \left[ \sum_{l=1}^{N_{Q}} \sum_{l=1}^{M} \sum_{l=1}^{N} |d(z_{ij,k}^{l}) - f(T_{l}(z_{ij,k}^{l}))| \right] \]  

where a stack filter \( S_{f} \), a threshold function \( T_{l} \) and a Boolean function \( f \) are used at each level.

A non-linear filter-based classifier can be obtained by using the graphic searching method [5]. Finally, a thematic map is classified by this GMS filter-based multi-class classifier.

3. Experimental results

A plantation area in Taiwan was chosen for investigation. The image data was obtained by the MODIS/ASTER (MASTER) Airborne Simulator, an airborne imaging spectrometer developed for the Moderate Resolution Imaging Spectroradiometer (MODIS) and Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) [19], during the PacRim II projects. A ground truth survey was taken for the selected six land cover types on the same day. MASTER data is available in 50 contiguous bands over the 400–1300 nm wavelengths, with a spatial resolution of 10–30 m. The proposed filter-based GMS method was applied to 35 bands.
Table 1
The classification error matrix. A summary of classification error is appended below.

<table>
<thead>
<tr>
<th>Classification data</th>
<th>A</th>
<th>B</th>
<th>S</th>
<th>P</th>
<th>O</th>
<th>R</th>
<th>Row total</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>691</td>
<td>14</td>
<td>10</td>
<td>0</td>
<td>3</td>
<td>2</td>
<td>720</td>
</tr>
<tr>
<td>B</td>
<td>36</td>
<td>679</td>
<td>5</td>
<td>2</td>
<td>3</td>
<td>1</td>
<td>720</td>
</tr>
<tr>
<td>S</td>
<td>23</td>
<td>5</td>
<td>692</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>720</td>
</tr>
<tr>
<td>P</td>
<td>4</td>
<td>2</td>
<td>3</td>
<td>710</td>
<td>1</td>
<td>0</td>
<td>720</td>
</tr>
<tr>
<td>O</td>
<td>55</td>
<td>9</td>
<td>6</td>
<td>18</td>
<td>626</td>
<td>6</td>
<td>720</td>
</tr>
<tr>
<td>R</td>
<td>43</td>
<td>4</td>
<td>47</td>
<td>0</td>
<td>11</td>
<td>615</td>
<td>720</td>
</tr>
<tr>
<td>Column total</td>
<td>846</td>
<td>713</td>
<td>763</td>
<td>730</td>
<td>644</td>
<td>624</td>
<td>4320</td>
</tr>
</tbody>
</table>

Classes

<table>
<thead>
<tr>
<th>Classes</th>
<th>Number correct</th>
<th>Producer's accuracy (%)</th>
<th>User's accuracy (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A (sugar cane A)</td>
<td>691</td>
<td>81.68</td>
<td>96.00</td>
</tr>
<tr>
<td>B (sugar cane B)</td>
<td>679</td>
<td>95.23</td>
<td>94.32</td>
</tr>
<tr>
<td>S (seawater)</td>
<td>692</td>
<td>90.69</td>
<td>96.11</td>
</tr>
<tr>
<td>P (pond)</td>
<td>710</td>
<td>97.30</td>
<td>98.61</td>
</tr>
<tr>
<td>O (bare soil)</td>
<td>626</td>
<td>97.20</td>
<td>86.94</td>
</tr>
<tr>
<td>R (rice)</td>
<td>615</td>
<td>98.56</td>
<td>85.42</td>
</tr>
</tbody>
</table>

Kappa coefficient 0.915

Overall accuracy 92.90%

selected from the 50 contiguous bands (excluding the low signal-to-noise ratio MIR-infrared channels) of MASTER.

The criterion for calculating the classification accuracy of experiments was based on exhaustive test cases. There were 150 labeled samples for each class. They were randomly collected from ground survey data by iterating every fifth sample interval for each class. They were partitioned into 30 (20%) training samples (less than the number of bands used in this experiment) and 120 (80%) test samples for each test case. Six land cover classes were used in the experiment. The A, B, S, P, O and R stand for sugar cane A, sugar cane B, sea water, pond, bare soil and rice land cover classes, respectively. There were 720 test samples (NM for each class). Four windows of fixed length three, four, five and six (n = 3, 4, 5, 6) were chosen for each class. Three correlation coefficient threshold values, \( \kappa = 0.75, 0.80 \) and 0.85, were selected to carry out GMST. Finally, the accuracy was obtained by averaging all of the multiple combinations stated above. An example error matrix was given in Table 1 to illustrate the accuracy assessment of GMS method for the best case. A summary evaluation of classification accuracy in Table 2 also demonstrates the validity of GMS method for supervised high-dimensional classification.

Interestingly, we conducted two modified versions of GMS in our experiments. After finding the GMS sets \( \Phi_k \), where \( k \in \{1, \ldots, N\} \), a feature scale uniformity procedure was performed to unify the feature scales of these GMSs to an identical GMS set for all classes.

Table 2
Summary evaluation of classification accuracy for different number of classes and feature subspaces

<table>
<thead>
<tr>
<th>Number of classes</th>
<th>Number of feature subspaces</th>
<th>Overall accuracy (%)</th>
<th>Kappa coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td>5</td>
<td>92.90</td>
<td>0.915</td>
</tr>
<tr>
<td>6</td>
<td>4</td>
<td>91.46</td>
<td>0.897</td>
</tr>
<tr>
<td>6</td>
<td>3</td>
<td>89.17</td>
<td>0.869</td>
</tr>
<tr>
<td>5</td>
<td>5</td>
<td>94.93</td>
<td>0.935</td>
</tr>
<tr>
<td>5</td>
<td>4</td>
<td>93.93</td>
<td>0.922</td>
</tr>
<tr>
<td>5</td>
<td>3</td>
<td>92.18</td>
<td>0.901</td>
</tr>
<tr>
<td>4</td>
<td>5</td>
<td>93.91</td>
<td>0.924</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>93.30</td>
<td>0.915</td>
</tr>
<tr>
<td>4</td>
<td>3</td>
<td>91.65</td>
<td>0.896</td>
</tr>
<tr>
<td>3</td>
<td>5</td>
<td>88.26</td>
<td>0.864</td>
</tr>
<tr>
<td>3</td>
<td>4</td>
<td>88.55</td>
<td>0.866</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>88.83</td>
<td>0.867</td>
</tr>
</tbody>
</table>
Fig. 7. The accuracy comparison of five feature extraction methods, GMS, PCA, Euclidean, IGMS and UGMS, where GMS specifies the average accuracy rates of four different (3, 4, 5, 6) window lengths. EUB, EAB and ELB represent the average classification accuracy of the best upper bounds, the average bands and the worst lower bound feature bands by using Euclidean distance measures, respectively. PCA stands for the average classification rates of PCA with three different (85, 90, 95%) cumulative percents. IGMS and UGMS are the average accuracy rates of four different (3, 4, 5, 6) window lengths using IGMS and UGMS feature extraction methods.

Fig. 8. Classified map of the Aus-Ku test site using GMS Method.
one used union operations applied to the band numbers inside each GMS module to unify the feature scales of different classes produced by GMST and constructed an identical union GMS (UGMS) set. It built an identical UGMS set for all classes \( \omega_k \), \( k \in \{1, \ldots, N\} \). Every different class has the same UGMS set. The second one used intersection operations applied to the band numbers inside each GMS module of different classes to select an identical intersection GMS (IGMS) set for all classes. These two different modified versions of GMS methods and two different conventional feature extraction methods, minimum Euclidean distance (MED) and PCA, were applied to GMS filter-based classifiers for accuracy comparison. In Fig. 7, the plot shows that GMS is superior to IGMS, UGMS, MED and PCA in general. The GMS feature extraction is very suitable for the filter-based multi-class classifier.

The encouraging results have shown that an adequate classification rates, almost near 95% average accuracy for the best case, are archived with only few training samples. Fig. 8 shows an example of the classified map obtained by applying GMS filter-based classification schemes to the MASTER high-dimensional data sets.

4. Conclusions

In this paper, a sophisticated GMS filter-based multi-class classifier is proposed for supervised high-dimensional data classification. We first introduce the GMS which can be obtained by a quick greedy band reordering algorithm. It is efficient with little computational complexity. The GMS is built by grouping highly correlated bands into a small set of bands. Reordering the bands regardless of the original order of wavelengths in high-dimensional data is an important characteristic of GMS. The GMS can be treated as not only a preprocess of the filter-based classifier but also a unique feature set for high-dimensional data. It makes use of the potential separability of different classes to overcome the drawback of the common covariance bias problems encountered in conventional PCA.

The characteristic of GMS is suitable for the non-linear filter-based multi-class classifier. The proposed GMS filter-based classifier enhances the separable features of different classes to improve the classification accuracy significantly. The conducted experiments demonstrated the validity of our proposed GMS filter-based classification scheme. Furthermore, we can take the benefits of the parallel property in GMS filter-based classifiers to achieve real time computation.

References

Yung-Lang Chang received his BS degree in electrical engineering from Chung Yuan Christian University, Taiwan, in 1987, MS degree in computer engineering from Syracuse University in 1993, and PhD degree in computer science and information engineering from National Central University, Taiwan, in 2003. From 1987 to 1999 he was a Teaching Assistant in the Department of Electrical Engineering, National Taiwan Institute of Technology. He worked for NDC IBM Taiwan as a Hardware Engineer from 1989 to 1991. From 1992 to 1993 he was a Teaching Assistant in the Department of Electrical and Computer Engineering, Syracuse University and IBM-Syracuse University affiliated off-campus education center. He was a Software Development Engineer with ALCATEL from 1993 to 1994 and is currently an Associate Professor of the Department of Computer Science and Information Engineering at St. John and St. Mary Institute of Technology, Taiwan. He is a member of SPIE. His research interests are in image processing, pattern recognition and remote sensing.

Kuo-Chin Fan received his BS degree in electrical engineering from National Tsing-Hua University, Taiwan, in 1981. In 1983, he worked for the Electronic Research and Service Organization (ERSO), Taiwan, as a Computer Engineer. He started his graduate studies in electrical engineering at the University of Florida in 1984 and received his MS and PhD degrees in 1985 and 1989, respectively. From 1984 to 1989 he was a Research Assistant in the Center for Information Research at University of Florida. In 1989, he joined the Institute of Computer Science and Information Engineering at National Central University where he became Professor in 1994. From 1994 to 1997 he was chairman of the department. Currently, he is the Director of the Computer Center, National Central University, Chung-Li, Taiwan. He is a member of SPIE and IEEE. His current research interests include image analysis, optical character recognition, and document analysis.

K-S. Chen received his BS degree from National Taiwan Institute of Technology, MS and PhD degrees from the University of Texas at Arlington in 1987 and 1990, respectively, all in electrical engineering. Since 1992, he has been with the Faculty of the Center for Space and Remote Sensing Research, Institute of Space Sciences, and Institute of Communication Engineering at the National Central University, Taiwan where he is now a Professor and Director. He is active in the national and international remote sensing community. He participated in several international remote sensing projects and SAR campaigns (GlobalSAR, PARM I & II). For many years, he has been a technical committee member or steering committee member in several international remote sensing conferences, e.g., IEEE Geoscience and Remote Sensing Symposium, Progress in Electromagnetic Research Symposium (PIERS), Asian Pacific Microwave Conference, Asian Remote Sensing Conference, etc. His research interests are in the areas of face recognition, biometrics, authentication, image analysis, computer vision, and pattern recognition.
was a Technical Chairman of 1999 PIERS held in Taipei. Currently, he is an Associate Editor (microwave remote sensing) of IEEE Transactions on Geoscience and Remote Sensing, and member of editorial board for Journal of Electromagnetic Waves and Applications. In addition, he is Editor-in-Chief of Chinese Journal of Photogrammetry and Remote Sensing. Since 1997, he has been the national correspondent of Commission I of International Society of Photogrammetry and Remote Sensing (ISPRS). He was recently appointed as Chairman of Commission F of URSI (Union of Radio Science International), China, Taipei. He is a senior member of IEEE and member of Electromagnetic Academy. He has authored three book chapters; over 60 referred journal papers, and 110 conference papers in the areas of remote sensing and wave scattering and propagation.

Jeng-Horng Chang received his BS degree in electrical engineering from National Taiwan Institute of Technology in 1991, the MS degree in electrical and computer engineering from University of Missouri-Columbia in 1993 and PhD degree in computer science and information engineering at National Central University, Taiwan, in 2001. He is currently an Associate Professor of the Department of Information Management at Takming College, Taiwan. His research interests include pattern recognition, image processing and artificial intelligence.