Feature selection based on max-min-associated indices for classification of remotely sensed imagery

Bo Wu [a] , Xiaoqin Wang [a] , Huanfeng Shen [b] & Xiaocheng Zhou [a]

[a] Key Laboratory of Spatial Data Mining and Information Sharing of
the Ministry of Education, Fuzhou University, Fuzhou, PR China

[b] School of Resource and Environmental Science, Wuhan
University, Wuhan, Hubei, 430079, PR China

Version of record first published: 01 Mar 2012.

PLEASE SCROLL DOWN FOR ARTICLE

Taylor & Francis
Taylor & Francis Group

Feature selection based on max–min-associated indices for classification of remotely sensed imagery

BO WU†*, XIAOQIN WANG†, HUANFENG SHEN‡
and XIAOCHENG ZHOU†

†Key Laboratory of Spatial Data Mining and Information Sharing of the Ministry of Education, Fuzhou University, Fuzhou, PR China
‡School of Resource and Environmental Science, Wuhan University, Wuhan, Hubei 430079, PR China

This article proposes two novel feature selection methods for dimension reduction according to max–min-associated indices derived from Cramer's V-test coefficient. The proposed methods incrementally select features simultaneously satisfying the criteria of a statistically maximal association (A) between target labels and features and a minimal association (R) among selected features with respect to Cramer's V-test value. Two indices are developed by different combinations of the A and R conditions. One index is to maximize A/R and the other is to maximize $A-\lambda R$, which are referred to as the MMAIQ and MMAIS methods, respectively. Since the proposed feature selection algorithms are feature filter methods, how to determine the best number of features is another important issue. This article adopts an information lost criterion by measuring the variation between χ^2 and β statistics to optimize the number of features selected associated with the Gaussian maximal likelihood classifier (GMLC). To validate the proposed methods, experiments are conducted with both a hyperspectral image data set and a high spatial resolution image data set. The results demonstrate that the proposed methods can provide an effective tool for feature selection and improve classification accuracy significantly. Furthermore, the proposed methods with well-known feature selection methods, i.e. mutual information-based max-dependency criterion (mRMR) and sequential forward selection (SFS), are evaluated and compared. The experiments demonstrate that the proposed methods can offer better results in terms of kappa coefficient and overall classification accuracy measurements.

1. Introduction

Remote-sensing research focusing on feature extraction and selection has long attracted the attention of the remote-sensing community because feature extraction and selection are prerequisites for successful image processing and various applications. Tremendous efforts have been dedicated to developing various feature extraction and selection methods to improve image-processing effectiveness and classification

---

*Corresponding author. Email: wavelet778@sohu.com

accuracy in recent years (Foody 1996, Stuckens *et al.* 2000, Segl *et al.* 2003, Zhang *et al.* 2008). Previous research generally suggests that effective extraction and use of potential multiple features of remotely sensed data, such as spectral signatures, various induced indices, and textural or contextual information, can significantly improve classification accuracy (Melgani and Sebastiano 2003, Platt and Goetz 2004, Xu 2006, Lu and Weng 2007). However, not all extracted features are equivalent in their contribution to classification tasks; some of them are perhaps superfluous and useless because they have high correlations or are trivial. Accordingly, the use of all possible features in a classification procedure may add unnecessary information redundancy and significantly decrease image classification accuracy (Hughes 1968, Price *et al.* 2002). As a consequence, employing feature selection techniques to obtain the most useful subset from candidate features is critical to a successful classification of remotely sensed data into a thematic map (Lu and Weng 2007).

Given input data with N samples (total samples) and m features (number of features) $\mathbf{X} = \{\mathbf{x}_1, \ldots, \mathbf{x}_m\}$, and the target classification variable c, the feature selection problem is to find a subset S of n features from m ($n \leq m$) features that optimally characterizes c. Conceptually, feature selection in general requires a search strategy and criterion functions (Webb 2002, Tan *et al.* 2006). The search algorithm generates and compares possible feature selection solutions by calculating their criterion function values as a measure of the effectiveness of each considered feature subset. One classic feature selection method is a sequential search technique, where the best feature subset with the prefixed number of features is achieved by adding to or removing from the current feature subset one feature at a time, such as sequential forward selection (SFS) and sequential backward selection (SBS) (Aha and Bankert 1996, Jain and Zongker 1997). Recently, stochastic search algorithms, such as simulated annealing (Siedlecki and Sklansky 1988), genetic algorithms (Raymer *et al.* 2000) and clonal selection algorithms (Zhang *et al.* 2007), have also been used for feature selection. Besides the search strategies, an optimal subset is always relative to a certain evaluation function. Various optimal criteria, such as distance-based (Jensen 1996), entropy-based (Peng *et al.* 2005) and dependence-based (Liu and Setiono 1997) measures, etc., have been fully investigated.

In this article, we focus on dependence-based techniques. These techniques have the advantages that they can easily scale to very high dimensional data sets, they are computationally simple and fast, and they are independent of the classification algorithm. However, it has been recognized that a direct combination of individually good features in terms of certain criteria do not necessarily lead to the best integrated performance. As an alternative, some researchers have studied indirect or direct means to reduce redundancy among features, and select features with minimal redundancy (Yu and Liu 2004). For instance, Pudil *et al.* (1994) presented a method to maximize the joint dependency of features on the target class with a sequential forward floating search such that redundant features can be removed. Peng *et al.* (2005) proposed a heuristic mutual information-based max-dependency criterion (mRMR) method to minimize redundancy, which uses a series of intuitive measures of relevance to select promising features. The mRMR method was validated as an effective technique for feature selection in a remote-sensing image (Wu *et al.* 2009). In this article, we propose two feature selection indices based on maximal association and minimal redundancy derived from Cramer's V-test.

2. Max–min-associated indices for feature selection

2.1 *Cramer's* V-*test*

The χ^2 test is one of the most widely used measures to define the dependence of variables (Weiss 1995) and has been demonstrated to be effective in feature selection (Liu and Setiono 1997). However, it is known that the χ^2 test of dependence is very sensitive to sample size (Agresti and Finlay 1997). Cramer's V is the most popular nominal association used to measure the strength of the relationship between variables regardless of table size (Reynolds 1984). It has the advantage of not being affected by sample size and is therefore very useful in situations where one suspects that a statistically significant χ^2 was the result of a large sample size rather than any substantive relationship between the variables (Reynolds 1984). Therefore, Cramer's V-test is employed to measure the association between target and variables. Given an *r*-row by *s*-column cross-tabulation, Cramer's V can be directly derived from the χ^2 statistic:

$$V = \sqrt{\frac{\chi^2}{N \min\{(r-1),(s-1)\}}} \tag{1}$$

The value of Cramer's V varies between 0 and 1. If its value is large, it means that there is a tendency for particular categories of the first variable to be associated with particular categories of the second variable. It has been suggested in practice that a Cramer's V of 0.1 provides a good minimum threshold for suggesting there is a substantive relationship between two variables (Martínez-Casasnovas *et al.* 2008).

2.2 *Max–min-associated indices*

Our study explored the possibility that a combination of Cramer's V coefficients can be further exploited for optimal feature selection. In this article, two max–min-associated indices derived from the Cramer's V-test coefficient were developed.

Intuitively, selected features must have maximal target class-associated ability. Therefore, a max-associated criterion is used to search for features satisfying (2) with Cramer's V-test measurement between individual features x_i and class c (*A* condition),

$$\max\left[A(\mathbf{S}, c) = \frac{1}{|\mathrm{S}|} \sum_{x_i \in S} V(x_i, c) \right], \tag{2}$$

where x_i is the *i*th feature, $\mathbf{S}$ the selected subset and $|\mathrm{S}|$ the number of elements of subset $\mathbf{S}$. It is likely that features selected according to the max-associated condition (2) will result in rich redundancy, that is the dependency among these features could be larger. When two features depend highly on each other, the respective class-discriminated power would not change much if one of them was removed. Therefore, the following minimal-associated condition (*R* condition) among selected features could be added to select mutually exclusive features:

$$\min\left[R(S) \quad R = \frac{1}{|\mathrm{S}|^2} \sum_{x_i, x_j \in S} V(x_i, x_j) \right], \tag{3}$$

where $V(x_i, x_j)$ is the Cramer's V-test between x_i and x_j. The max–min-associated indices for feature selection are derived directly from the above two criteria. Two combined methods, referred to as Max–Min-Associated Indices Quotient (MMAIQ) and Substraction (MMAIS) respectively, are designed. These combinations are expressed in equations (4) and (5).

$$\max\left[\phi(A, R) = A/R\right], \tag{4}$$

$$\max\left[\phi(A, R) = A - \lambda R\right]. \tag{5}$$

It is apparent that both (4) and (5) simultaneously satisfy the constraints on A and R. That is, a good feature should be one with maximal target class-associated ability, and at the same time with minimal association among the selected features. In equation (5), there is a regularization parameter λ, whose function is to balance the functions of the two constraints in (2) and (3).

2.3 *Feature selection algorithm*

To select the candidate feature set, an incremental method is used to find the suboptimal features defined by equations (4) and (5). Although this search strategy does not allow the features to be reselected once they have been selected, it can usually ensure that the selected features with relevance and redundancy constraints are the most prominent features not to be removed (Yu and Liu 2004). In addition, the incremental search method is rather fast. Suppose we already have $\mathbf{S}_{p-1}$, the set with p–1 features, the task is to select the pth feature from set $\{\mathbf{X}\text{–}\mathbf{S}_{p-1}\}$, where $\mathbf{X}$ is the feature set, such that the feature maximizes equations (4) and (5). The incremental algorithm optimizes the following conditions:

$$\max_{x_j \in X-S_{p-1}}\left[V(x_j, c) - \frac{\lambda}{p-1}\sum_{x_i, x_j \in S}V(x_j, x_i)\right], \tag{6}$$

$$\max_{x_j \in X-S_{p-1}}\left[V(x_j, c)\bigg/\left(\frac{1}{p-1}\sum_{x_i, x_j \in S}V(x_j, x_i)\right)\right], \tag{7}$$

where $V(x_j, c)$ is the Cramer's V-test between x_j and target c. These optimizations can be computed efficiently in $O(|S| \cdot m)$ complexity. As a result, we can obtain the ranked features rapidly even if the dimension of features is possibly very high.

3. Implementation

Two important issues must be solved before the classification process. One is how to obtain cross-tabulation, such that Cramer's V can be calculated if the concerned features contain continuous variables. In this case, a preprocessing step of discretization is required to obtain cross-tabulation. Discretization can transform a continuous feature variable into a finite number of intervals, where each interval is associated with a numerical discrete value. The simplest discretization technique with equal-frequency was considered in this study, such that continuous ranges were divided into sub-ranges by a pre-specified frequency.

Another critical problem is how to optimize the best number of feature subsets. The best number of features is usually estimated by the K folds cross-validation of the correct classification rate (CCR). Let $\text{CCR}_{K,\text{cross}}$ be K cross-validation repetitions. The $\text{CCR}_{K,\text{cross}}$ is given by

$$\text{CCR}_{K,\text{cross}} = \frac{1}{K} \sum_{k=1}^{K} \text{CCR}_k = \frac{1}{K} \frac{1}{N_K} \sum_{k=1}^{K} \sum_{i=1}^{N_K} L\left(c_i, \hat{c}_i\right), \tag{8}$$

where N_K is the sample number of the kth cross-validation, $L(c_i, \hat{c}_i)$ denotes the 0–1 loss function of the ith sample between the ground truth label c_i and the predicted label $\hat{c}_i$, and $N_K = (K-1) \times N/K$ is the number of test sets during the kth iteration. It is apparent that $\text{CCR}_{K,\text{cross}}$ is actually the average over KCCRs of cross-validation repetitions. Because the CCR is not an accurate indicator with limited samples (Ververidis and Kotropoulos 2009), the lower CCR limit index $\text{CCR}_{K,\text{cross}}^{\text{Lower}}$ with compensatory information is thereby adopted to improve the CCR. The $\text{CCR}_{K,\text{cross}}^{\text{Lower}}$ criterion measures the variation between χ^2 and β to optimize the number of selected features associated with the Gaussian maximal likelihood classifier (GMLC) (Ververidis and Kotropoulos 2009):

$$\text{CCR}_{K,\text{cross}}^{\text{Lower}} = \text{CCR}_{K,\text{cross}} - L_{\text{cross}}\left(N_K, m\right) \times \left[\text{CCR}_{K,\text{cross}} - 1/C\right], \tag{9}$$

where C is the number of classes and $\text{CCR}_{K,\text{cross}}^{\text{Lower}}$ the lower CCR limit index with K cross-validation, and the loss information is given by

$$L_{\text{cross}}\left(N_K, m\right) = F_{\chi 2(m)}(t_1) - I_{\frac{1}{1+\frac{N_K^2-1}{N_K * t_1}}}\left(\frac{m}{2}, \frac{N_K - m}{2}\right), \tag{10}$$

where m denotes the number of features and t_1 is defined as

$$t_1 = -N_K W_{-1}\left(-\left[\frac{\Gamma(N_K/2)}{\Gamma\left(\frac{N_K-m}{2}\right)}\right]^{\frac{2}{N}} \frac{2^{\frac{m}{N_K}} \times N_K^{\frac{m}{N_K}-2}}{(N_K^2-1)^{\frac{m}{N_K}-1}} \times \exp\left(\frac{1-N_K^2}{N_K^2}\right)\right) - N_K + \frac{1}{N_K^2}, \tag{11}$$

where $\Gamma(\cdot)$ denotes γ function with respective parameter, $F_{x^2(m)}(x)$ is the continuous density function (cdf) of $\chi^2(m)$, $I_x(a,b)$ is the incomplete β function with parameters a and b, and $W_j(x)$ is the jth branch of Lambert's W function (Corless *et al.* 1996).

4. Experimental studies

Experiments have been conducted to test the performance of the proposed MMAIQ and MMAIS algorithms using a hyperspectral remote-sensing image (PHI) and a high spatial resolution image (QuickBird). The main concerns of selecting the two experimental data sets are: (1) there are adequate field-investigated samples to be used; (2) the two images collected by different sensors involve different data characteristics. Consistent comparisons between MMAIQ, MMAIS, SFS and mRMR algorithms were performed. One main reason for choosing the SFS and mRMR methods is that both are representative feature ranking techniques. The estimation of feature selection

quality for these algorithms was performed by means of the classification accuracy of the classified imagery acquired from pushbroom hyperspectral imagery (PHI) and QuickBird images.

4.1 *Experiment 1: Xiaqiao PHI*

The data set used in this experiment was collected in September 1999 at the Xiaqiao test site, a mixed agricultural area in Changzhou city, Jiangsu province, China, and was airborne PHI. The image was acquired at an approximate height of 2000 m with a 3 m spatial resolution. A sub-scene (346 × 350 pixels) of the PHI image with 80 bands was tested, and its spectral range was 417–854 nm. Figure 1 shows the experimental PHI image cube. The ground truth spectral data were collected in September 1999 by field spectrometer SE590. The observed image was expected to be classified into eight representative classes, that is, corn, vegetable – sweet potato, vegetable – cabbage, soil, float grass, road, water and puddle/polluted water.

Figure 2 shows the reflectance curves of the above eight land-cover classes. As can be seen from figure 2, distinct spectral differences among these materials are at 700–840 nm, which is the best spectral range to discriminate between various vegetations. Therefore, it is expected that more bands will be selected in this range. The list of classes and the number of labelled samples for each class are given in table 1. The available labelled samples amount to 4308 pixels. A total of 1630 samples are randomly selected for training the classifier, and the remaining 2678 samples are employed for classification accuracy assessment. Prior study on the relationship between training sample size and data dimensionality has validated that the GMLC classifier can achieve a satisfactory classification accuracy if the number of training samples per class amounts to 2–4 times bands (Van Neil *et al.* 2005). Since spectral features are
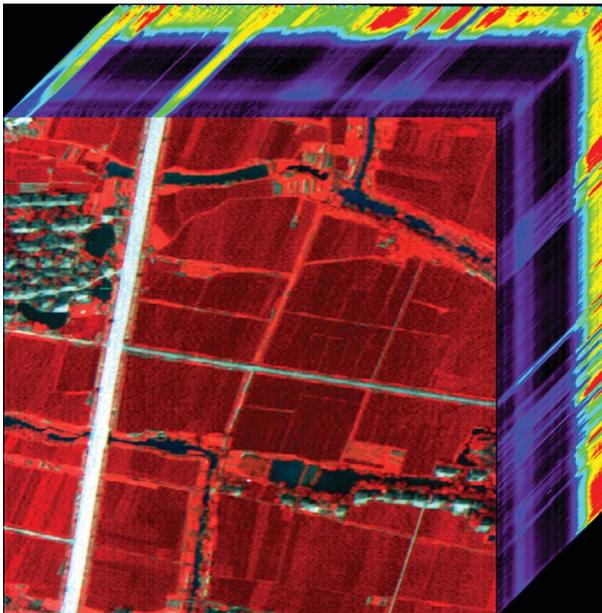
Figure 1.   Experimental hyperspectral image cube of Xiaqiao.
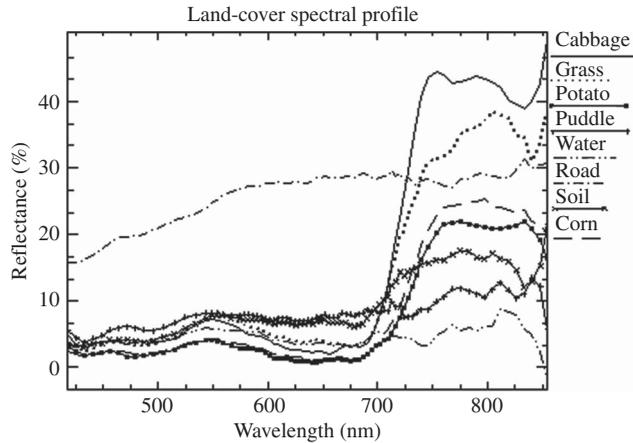
Figure 2. Reflectance curves of eight land-cover classes.

Table 1. List of classes and number of labelled training data-testing samples in each class for experiment 1.

Class name	Training samples	Testing samples	Number of labelled samples
Soil	186	329	515
Grass	154	290	444
Potato	190	330	520
Cabbage	162	309	471
Water	264	389	653
Puddle	178	285	463
Corn	188	382	570
Road	308	364	672
Total number of samples	1630	2678	4308

continuous attributes, the reflectance of all bands was first partitioned into six intervals using the equal-width and equal-frequency method.

The CCR is plotted versus the feature selection steps for cross-validation in figure 3. The optimal number of features predicted by the lower limit of the CCR is plotted with a red line. The vertical red line indicates the steps where the optimum selected feature subset is derived. It is clearly seen from figure 3 that the optimal feature number of PHI data is five with the $CCR_{K,cross}^{Lower}$ measurement. Consequently, only the first five ranked features obtained by the aforementioned methods were used for classification and comparisons.

Figures 4(*a*)–(*c*) illustrate the classification results using the GMLC classifier with full bands, SFS, and the mRMR feature selection algorithm, respectively, and figures 4(*d*) and (*e*) show the results with the proposed feature selection methods. To evaluate the classification accuracy, a test field map is provided in figure 4(*f*) based on field investigation.

A visual comparison of the five classification results in figure 4 suggests varying degrees of accuracy of pixel assignment. It can be found from figure 4 that five methods exhibit similar classification results, but classification images with feature selection
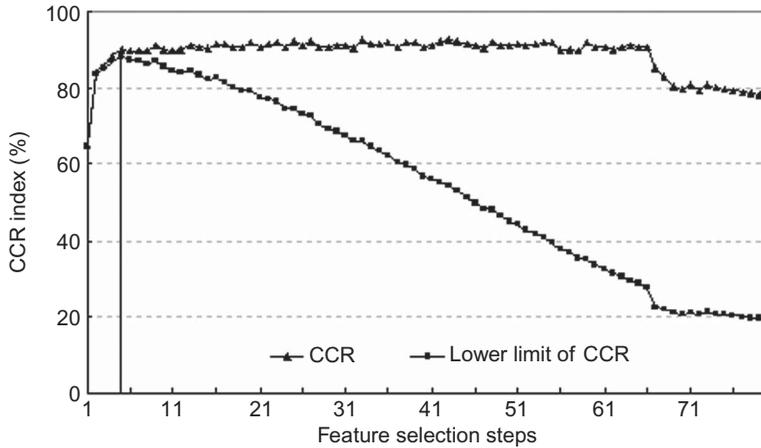
Figure 3. Lower limit of CCR estimated by cross-validation versus the feature selection steps for PHI data.

outperform the classification result with full bands, where cabbage, grass, corn and potato are confused to some degree.

For a more detailed verification of the classification results, we compared ground truth data with the classified images and assessed the accuracy of each method quantitatively using both overall accuracy and kappa coefficient. Tables 2 and 3 list the results of comparisons between ground truth data and classified images obtained by different methods. It is apparent from tables 2 and 3 that the classification accuracies with feature selection significantly outperform the original image, as well as greatly reduce the feature bands. The MMAIQ improved the overall accuracy from 79.1% to 89.2%, an increase of 10.1%, and the kappa coefficient from 0.76 to 0.87, improving by 0.11. Comparing SFS, mRMR, MMAIS and MMAIQ, MMAIQ obtains the best overall accuracy, i.e. the best percentage of correctly classified pixels among all the test pixels considered. The overall accuracy improved by 2.9% and 3.4% for the SFS and mRMR methods, respectively. MMAIS also achieves satisfactory accuracy; the overall accuracy and kappa coefficient are 87.1% and 0.852, respectively. Another comparison is to examine the five selected bands with different methods. It can be seen from table 3 that MMAIS and MMAIQ select more bands in the range of the near-infrared interval in accord with spectral curves shown in figure 2, where the best spectra to capture vegetable characteristics at 700–840 nm.

4.2 *Experiment 2: Fuzhou QuickBird*

To further validate the effectiveness and generalized ability of the proposed method, a subset image with 3859×2806 pixels (figure 5(*a*)) chosen from the QuickBird image of Fuzhou was also investigated. The image was acquired in June 2003, collecting panchromatic images with 0.6 m resolution and multispectral images with 2.4 m resolution. According to ground truth data, the test site contains typical parcels in urban areas of roads, vegetation patches, water, building areas, bare land and shadow regions.

Because high spatial resolution data often create higher spectral variance within each class corresponding to land-cover units owing to fewer bands and lower spectral separable ability, pixel-based classification schemes employing only spectral
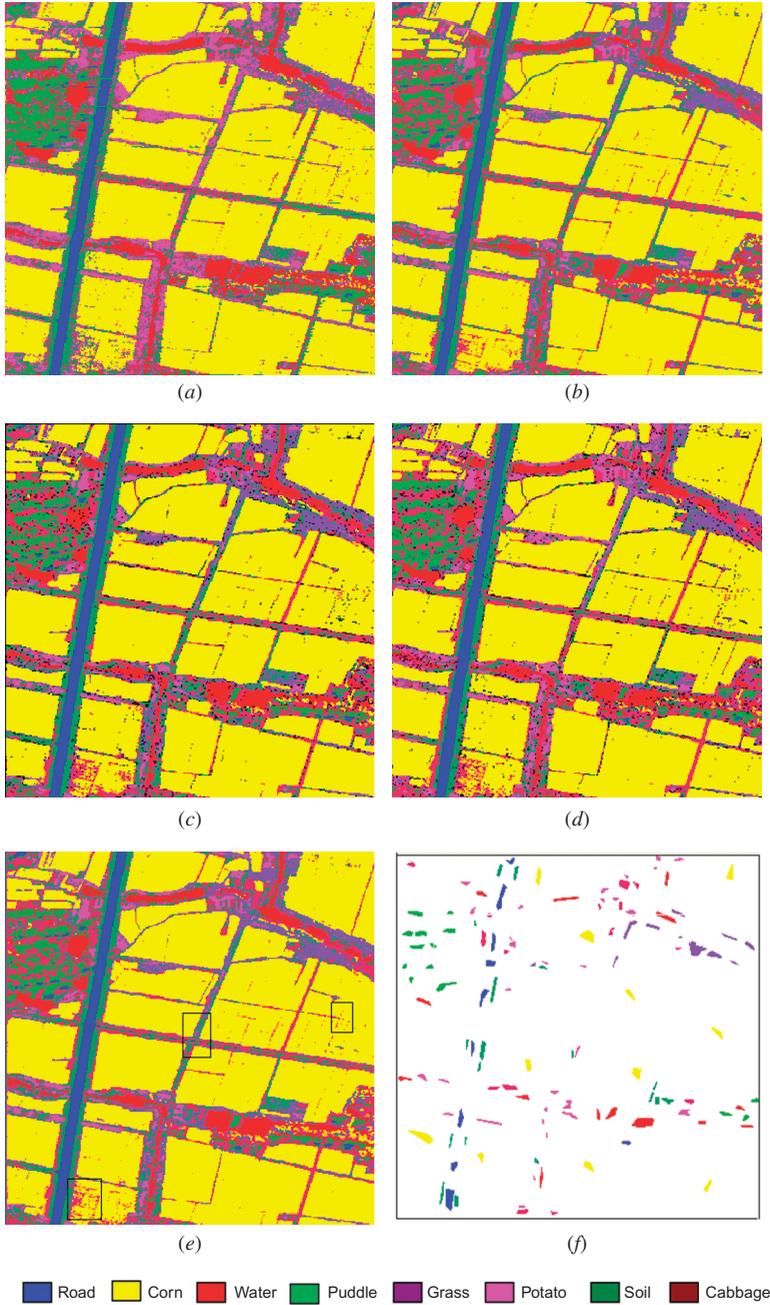
Figure 4.　Classification images with GMLC classifier for PHI data set. (*a*) full bands, (*b*) SFS, (*c*) mRMR, (*d*) MMAIS, (*e*) MMAIQ and (*f*) the image for test fields used in experiment 1. Some easily confused objects are highlighted with black boxes.

Table 2. Confusion matrix with GLMC classifier.

Methods	Soil	Grass	Potato	Cabbage	Water	Puddle	Corn	Road	Total
Full bands									
Soil	230	58	9	97	0	1	1	10	406
Grass	15	71	11	1	0	1	0	0	99
Potato	36	147	298	17	1	0	1	0	500
Cabbage	45	44	10	160	1	8	7	10	255
Water	2	0	0	6	382	13	0	0	403
Puddle	1	0	0	25	5	262	0	2	295
Corn	0	0	2	0	0	0	373	0	375
Road	0	0	0	3	0	0	0	342	345
Total	329	290	330	309	389	285	382	364	2678
SFS									
Soil	220	13	16	45	0	0	0	17	311
Grass	37	231	51	0	0	0	0	0	319
Potato	19	42	257	14	0	0	0	0	332
Cabbage	53	4	4	243	6	20	7	1	338
Water	0	0	0	0	383	8	0	0	391
Puddle	0	0	0	7	0	257	0	0	264
Corn	0	0	2	0	0	0	375	0	377
Road	0	0	0	0	0	0	0	346	346
Total	329	290	330	309	389	285	382	364	2678
mRMR									
Soil	240	35	3	43	0	0	1	18	340
Grass	24	195	47	1	0	0	0	0	267
Potato	10	60	271	10	0	0	0	0	351
Cabbage	55	0	3	249	7	22	11	1	333
Water	0	0	0	0	377	11	0	0	388
Puddle	0	0	0	6	5	252	0	0	263
Corn	0	0	6	0	0	0	370	0	376
Road	0	0	0	0	0	0	0	345	345
Total	329	290	330	309	389	285	382	364	2678
MMAIS									
Soil	238	6	3	36	0	0	0	33	316
Grass	24	225	28	1	0	0	4	0	382
Potato	15	59	297	9	0	0	0	0	380
Cabbage	50	0	0	240	13	19	4	1	327
Water	0	0	0	1	368	6	0	0	375
Puddle	0	0	0	22	8	260	0	0	290
Corn	0	0	2	0	0	0	374	0	376
Road	2	0	0	0	0	0	0	330	332
Total	329	290	330	309	389	285	382	364	2678
MMAIQ									
Soil	261	19	2	49	0	0	0	21	352
Grass	33	255	48	1	0	0	1	0	338
Potato	0	16	275	8	0	0	0	0	299
Cabbage	35	0	1	248	8	16	8	1	317
Water	0	0	0	0	377	11	0	0	388
Puddle	0	0	0	3	4	258	0	0	265
Corn	0	0	4	0	0	0	373	0	377
Road	0	0	0	0	0	0	0	342	342
Total	329	290	330	309	389	285	382	364	2678

Table 3. Comparison of classification accuracy using GMLC classifier with different feature selection methods.

Feature selection	Full bands	SFS	mRMR	MMAIS	MMAIQ
Overall accuracy (%)	79.08	86.33	85.85	87.08	89.20
Kappa coefficient	0.7603	0.8436	0.8380	0.8521	0.8765
Selected bands	All	7, 18, 46, 63, 64	13, 37, 57, 65, 74	7, 59, 62, 67, 70	5, 54, 61, 76, 79
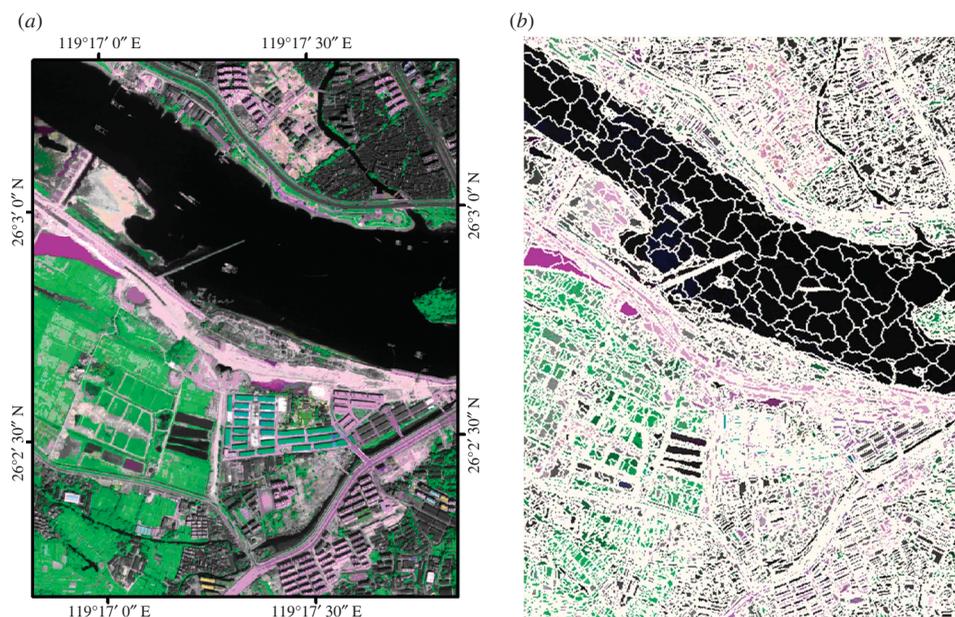
Figure 5. Experimental data of the QuickBird image in Fuzhou (*a*) fused image and (*b*) segmented results.

information often result in lower classification accuracy (Bruzzone and Carlin 2006). On the other hand, object-based methods can capture detailed geometrical, structural and textural feature information inherent in high spatial resolution images. All of these are potentially important features to discriminate between spectrally similar objects. Consequently, we adopt an object-based classification method to accomplish the classification of high spatial resolution data. In order to make full use of spectral and spatial information, panchromatic and multispectral images were first fused, and a 0.6 m multispectral image was obtained. Then the fused image was segmented into 7141 objects using the Fractal Network Evolution Approach (Baatz and Schape 2000) with the support of eCognition software (Definiens Imaging 2003). The scale, colour and shape parameters needed in eCognition were chosen as 80, 0.8, and 0.2, respectively, such that small objects, such as shadows and trees beside roads, etc., could be easily identified. Figure 5(*b*) shows the segmented objects. Various object

features, such as geometrical, structural and textural features, were extracted from the segmented image.

Generally, image spectral information provides the principal measure to identify ground objects in remotely sensed imagery; hence spectral features are first extracted from the segmented objects. Image characteristics such as shape, pattern and texture are possibly the most important features used in the visual interpretation of high spatial resolution imagery (Van Coillie *et al.* 2007). Measured textural parameters should therefore be extracted based on co-occurrence matrices of the image grey level. Object shape is obtained by calculating object length/width, proportion, shape, index and area, etc. Moreover, the normalized difference vegetation index (NDVI) and normalized difference water index (NDWI) are also extracted for better separation of vegetation and water from other materials (Xu 2006). Table 4 summarizes the object features into four categories, namely spectral, shape, texture and index features, which add up to 82 features.

A hybrid classification scheme (table 5) was employed according to the US Geological Survey land-use/land-cover classification system. The available labelled samples amount to 2981 objects. These labelled objects are randomly selected and validated by field investigation. Among them, 1055 samples (about 40%) are stratified and randomly selected for training the GMLC classifier, and the other 1926 samples are employed for classification accuracy assessment. The list of classes and the number of labelled samples for each class are given in table 6. From table 6, it can be seen that the training samples per class agree with the relationship between training sample size and data dimension (Van Neil *et al.* 2005).

The partition of intervals was carefully chosen as 3. Again, the optimal number of features was also determined by the lower correct classification rate ($\mathrm{CCR}_{K,\mathrm{cross}}^{\mathrm{Lower}}$) criterion. The CCR versus the feature selection steps for cross-validation was plotted in figure 6. It can be seen from figure 6 that the optimal feature subset was found with the $\mathrm{CCR}_{K,\mathrm{cross}}^{\mathrm{Lower}}$ measurement when the selected number reached 6. Consequently, the first six ranked features were used for classification and comparison.

The classification maps using different methods are shown in figure 7. For comparison purposes, a map identified by human interpretation is also displayed in figure 7(*f*), and some objects that are easily confused are highlighted with black boxes in figure 7(*f*). Figure 7(*a*) shows many uncertainties in the classification map that use whole features. These significant misclassifications include buildings, roads and bare land. In addition, water and shadows are confused to some extent. A possible reason is that some features exhibit high correlation, deteriorating the GMLC classification accuracy. It can be seen that figures 7(*b*)–(*e*) show a considerable improvement over figure 7(*a*), where buildings and bare land, and water and shadows are better recognized. Comparing figures 7(*b*)–(*f*), it is apparent that MMAIQ can better separate buildings from bare land and roads than SFS and mRMR (see top right box). Another observed improvement is that MMAIQ can discriminate polluted water from vegetation (see central box), whereas SFS and mRMR cannot. Comparing MMAIQ and MMAIS, both exhibit similar classification results, but some water near roads was inaccurately classed as shadow by MMAIS. These visual analyses are quantitatively illustrated by the confusion matrix in table 7.

Tables 7 and 8 list the results of comparisons between ground investigation data and classified thematic maps obtained by different methods. It is apparent from tables 7 and 8 that MMAIQ obtains the best overall accuracy again, that is, the best percentage of correctly classified pixels among all the test pixels considered. The overall accuracy

Table 4. Summarized object features extracted from segmented image.

Feature categories	Sub-class	Object features	Num. of features	Sequential number
Spectral features (22)	Mean	Brightness	1	5
		Max. diff.	1	56
		Means	4	59–62
	Std	Standard deviations	4	74–77
	Pixel based	Contrast with neighbouring pixels	3	8–10
		Max. pixel value	1	57
		Mean diff. to neighbours NIR band	1	58
		Mean of inner border NIR band	1	63
		Mean of outer border	1	64
		Min. pixel value	1	65
		Band ratio of red and NIR	2	69–70
		Std. dev. to neighbour pixels NIR band	2	78–79
Texture features (40)	GLCM	Ang. 2nd moment	5	13–17
		Contrast	5	18–22
		Correlation	5	23–27
		Dissimilarity	5	28–32
		Entropy	5	33–37
		Homogeneity	5	38–42
		Mean	5	43–47
		Std. dev.	5	48–52
Shape features (16)	Generic shape	Area	1	1
		Asymmetry	1	2
		Border length	1	4
		Compactness	1	6
		Compactness	1	7
		Density	1	11
		Elliptic fit	1	12
		Length	1	53
		Length/width	1	54
		Main direction	1	55
		Number of edges (polygon)	1	66
		Radius of largest enclosed ellipse	1	67
		Radius of smallest enclosing ellipse	1	68
		Rectangular fit	1	71
		roundness	1	72
		Width	1	80
Index features (4)	Shape based	Border index	1	3
		Shape index	1	73
	User defined	NDVI	1	81
		NDWI	1	82
Total			82	

and kappa coefficient are 89.9% and 0.876%, respectively, improving the overall accuracy of SFS and mRMR methods by 2.2 and 2.4%, respectively. The MMAIS also achieves higher accuracy. The overall accuracy and kappa coefficient are 89.7% and 0.874%, respectively.

Table 5. Land-use classification scheme.

Class name	Description
Bare land	Vacant area, including bare soil, and unproductive surfaces, etc.
Building	Residential, industrial, commercial buildings
Road	Transportation, including roads, bridges, railways and their affiliations
Shadow	Shadow, including tree shadow, high building shadow and bridge shadow, etc.
Vegetable	Green vegetation, including parks, green belt, grass and trees, etc.
Water	Water, including rivers, reservoirs, ponds and streams, etc.

Table 6. List of classes and number of labelled training data-testing samples in each class for experiment 2.

Class name	Training samples	Testing samples	Number of labelled samples
Bare land	148	280	428
Building	267	514	781
Road	154	292	446
Shadow	192	345	537
Vegetable	198	373	571
Water	96	122	191
Total number of samples	1055	1926	2981
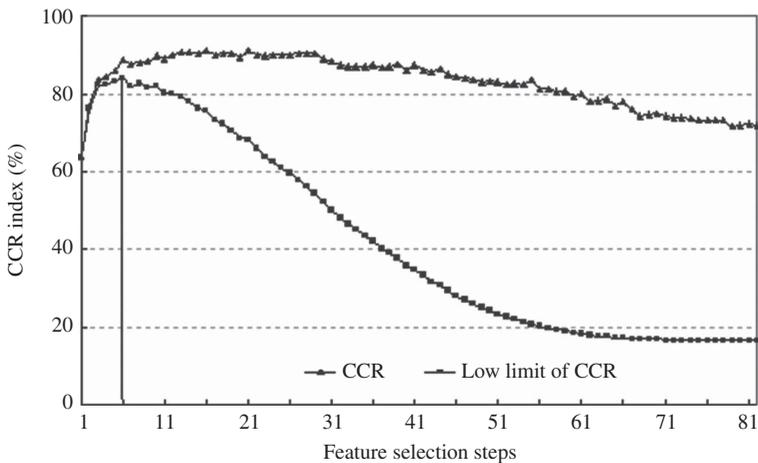
Figure 6. Lower limit of CCR estimated by cross validation versus the feature selection steps for QuickBird data.

From tables 8 and 4, the six features correspond to 5, 8, 34, 54, 70 and 81, which stand for brightness, contrast to neighbouring pixels, textural entropy, length/width, near-infrared band ratio, and NDVI, respectively. This result demonstrates that spectral, geometrical, structural, NDVI and textural features do make significant contributions to high spatial resolution image classification. This has also been validated by a good deal of earlier research. The selected features are shown in figure 8,
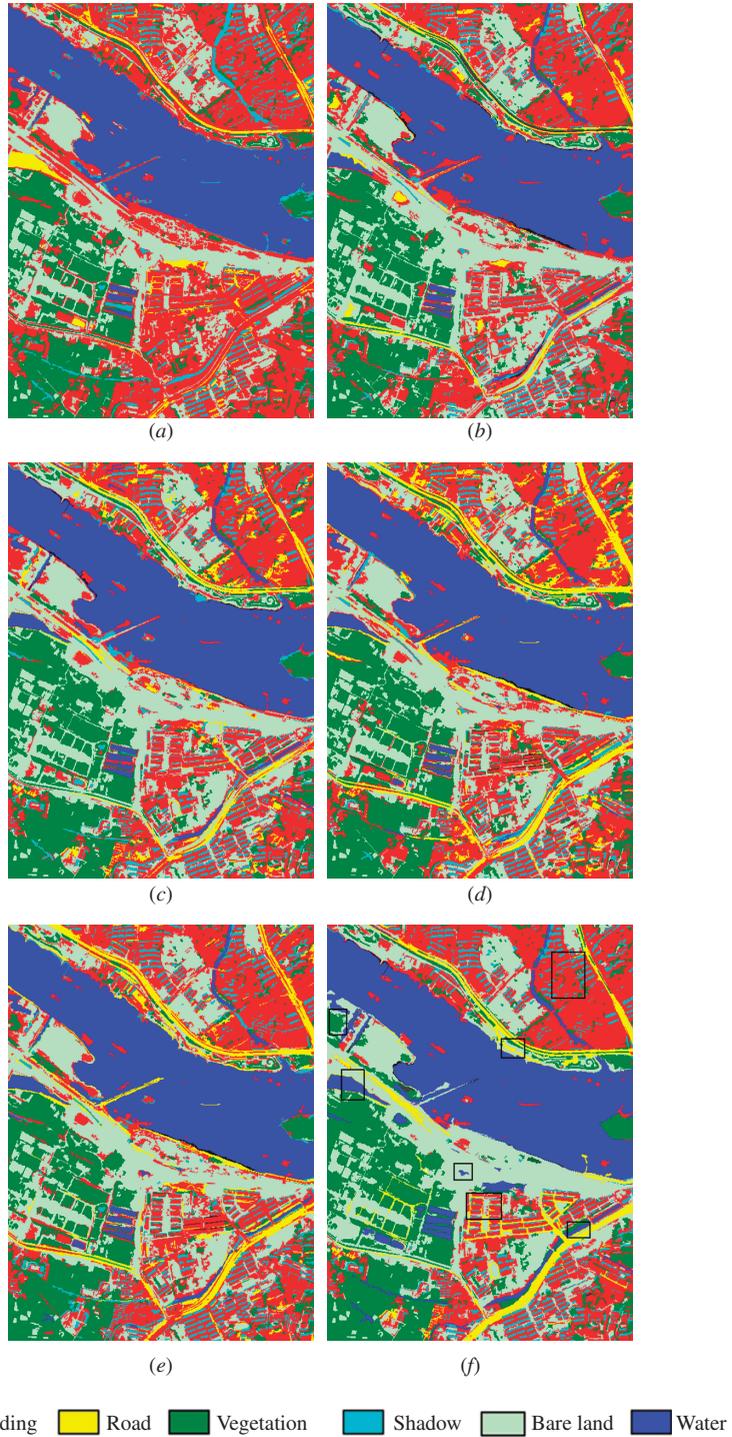
Figure 7.　Classification images with different methods. (*a*) Without feature selection, (*b*) SFS, (*c*) mRMR, (*d*) MMAIS, (*e*) MMAIQ and (*f*) the map of manual interpretation. Some easily confused objects are highlighted with black boxes.

Table 7. Confusion matrix of QuickBird data with GLMC classifier.

Methods	Bare land	Building	Road	Shadow	Vegetation	Water	Total
SFS							
Bare land	249	4	4	0	2	0	259
Building	11	466	88	5	0	3	573
Road	4	29	182	0	5	0	220
Shadow	0	7	5	323	0	13	348
Vegetation	16	4	13	15	363	0	411
Water	0	4	0	2	3	106	115
Total	280	514	292	345	373	122	1926
mRMR							
Bare land	252	8	4	0	2	0	266
Building	3	384	19	2	0	0	408
Road	9	110	246	3	0	0	368
Shadow	0	8	5	318	0	3	334
Vegetation	16	4	18	15	368	3	424
Water	0	0	0	7	3	116	126
Total	280	514	292	345	373	122	1926
MMAIS							
Bare land	252	0	4	0	2	0	258
Building	3	417	19	0	0	0	439
Road	11	80	259	0	5	0	355
Shadow	0	13	5	321	0	6	345
Vegetation	14	4	5	15	363	0	401
Water	0	0	0	9	3	116	128
Total	280	514	292	345	373	122	1926
MMAIQ							
Bare land	257	5	4	0	2	0	268
Building	3	410	14	2	0	0	429
Road	9	79	264	0	5	0	357
Shadow	0	16	5	321	0	6	348
Vegetation	11	4	5	19	363	0	402
Water	0	0	0	3	3	116	122
Total	280	514	292	345	373	122	1926

Table 8. Comparison of classification accuracy using GMLC classifier with different feature selection methods.

Feature selection	SFS	mRMR	MMAIS	MMAIQ
Overall accuracy	0.8769	0.8744	0.8972	0.8988
Kappa coefficient	0.8474	0.8461	0.8739	0.8758
Selected features	82, 44, 4, 1, 3, 8	82, 4, 44, 37, 81, 10	70, 54, 61, 31, 10, 81	70, 54, 5, 34, 8, 81

which shows that they all correlate with our interpreted experience, and accordingly have good visual interpretation. First, brightness contains spectral energy providing the principal measure of identifying objects. Second, band ratio and contrast with neighbouring pixels are two important features to enhance image quality. In addition,
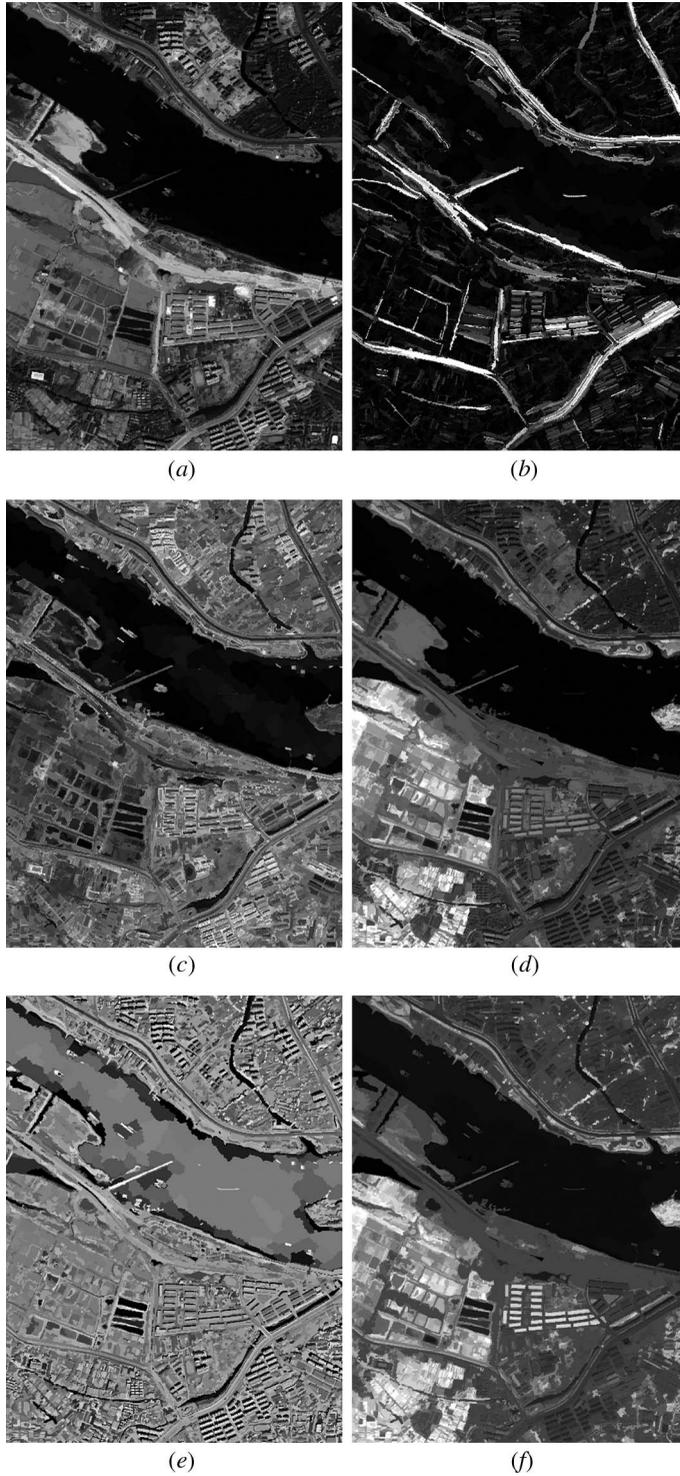
Figure 8. Selected features with the MMAIQ method are shown in the images. (*a*) Brightness, (*b*) length/width, (*c*) GLCM entropy, (*d*) band ratio, (*e*) contrast with neighbouring pixels and (*f*) NDVI.

length/width is a good shape indictor to separate roads from buildings, although they have similar spectral characteristics. Again, NDVI is a good vegetation descriptor to improve classification results, and textural entropy may be the most important clue in identifying homogeneous regions.

5. Discussions

A key parameter in the preprocessing step of discretization is the number of intervals used to partition each feature, which will be further discussed later. In general, a trade-off must be made between information quality and statistical equality that indicates good predictive accuracy and sufficient sample size in every interval. Figure 9 shows the CCR versus the number of partitioned intervals. It seems that there is no obvious rule with different data. When the partitioned interval increases from 2 to 10, the CCR of PHI data and QuickBird data varies from 84.7% to 89.2%, and 77.9% to 89.7%, respectively. From figure 9, the parameter seems insensitive; therefore, we can obtain a satisfied classification accuracy even if the parameter does not optimize.

As for the proposed MMAIQ and MMAIS, in general, MMAIQ imposes a greater penalty on redundancy. Empirically, MMAIQ often leads to better classification than MMAIS for candidate features. However, the joint effect of these features is less robust, especially for features containing noise. By contrast, MMAIS has an additional parameter to neatly adjust the function of A and R conditions. Generally, enlarging the parameter λ in the MMAIS model may stress a penalty on the redundancy, and result in similar effects to MMAIQ.

We are aware of several research areas yet to be considered regarding the proposed method. To begin with, discretization of continuous variables with equal-width and equal-frequency can be further refined by using instance class labels in the discretization process, such that features are optimally partitioned into a fixed number of unequal intervals. This optimal process has the potential to further improve feature selection. In addition, because of the computational complexity of joint association among features, the proposed scheme adopts an incremental forward selection method. A mechanism to remove potentially redundant features with a backward refining process from already selected features can be considered. Moreover, as with filter-based feature selection methods, they can be wrapped by other classifiers such as Support Vector Machines (SVMs) and k-nearest neighbour (k-NN) classifiers to demonstrate their generalized capability. Furthermore, although the method has been proven to be effective with remotely sensed data, other types of data, for instance
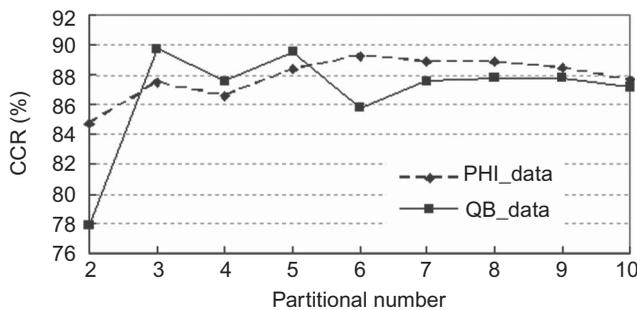
Figure 9. CCR versus the partitional number of features.

mixed type attribute data, need to be examined in order to strengthen its robustness. We will test the method with benchmark databases in the future.

6. Conclusions

In this article, new methods, MMAIQ and MMAIS, for feature selection in remote-sensing images and its derived data have been introduced. Our methods have been compared in terms of overall accuracy and kappa coefficient with the SFS and mRMR methods, which are known for their general abilities and good performance. The experimental results consistently show that the proposed feature selection methods can provide an effective tool for feature selection and improve classification accuracy significantly. When compared with SFS and mRMR, MMAIQ performs the best feature selection, and offers better or comparable classification accuracy in two experiments with different types of image. MMAIS also achieves satisfactory results in the same experiments. These results testify that MMAIQ and MMAIS provide new and effective options for feature selection.

References

AGRESTI, A. and FINLAY, B., 1997, *Statistical Methods for the Social Sciences*, 3rd ed. (Upper Saddle River, NJ: Prentice Hall).

AHA, D.W. and BANKERT, R.L., 1996, A comparative evaluation of sequential feature selection algorithms. In *Learning from Data: Artificial Intelligence and Statistics*, H.F. Douglas and L. Hans-Joachim (Eds.), pp. 199–206. Lecture Note in Statistics (New York, NY: Springer-Verlag).

BAATZ, M. and SCHAPE, A., 2000, Multi-resolution segmentation: an optimization approach for high quality multi-scale image segmentation. In *Angewandte Geographische Information Sverarbeitung XII*, J. Strobl and T. Blaschke (Eds.), pp. 12–23 (Heidelberg: Wichmann).

BRUZZONE, L. and CARLIN, L., 2006, A multilevel context-based system for classification of very high spatial resolution images. *IEEE Transactions on Geoscience and Remote Sensing*, **44**, pp. 2587–2600.

CORLESS, R., GONNET, G., HARE, D., JEFFREY, D. and KNUTH, D., 1996, On the Lambert W function. *Advances in Computational Mathematics*, **5**, pp. 329–359.

DEFINIENS IMAGING, 2003, *ECognition Professional User Guide 5*. (Munich, Germany). Available online at: http://www.gis.unbc.ca/help/software/ecognition4/ELuserguide.pdf (accessed 30 June 2011).

FOODY, G.M., 1996, Approaches for the production and evaluation of fuzzy land cover classification from remotely sensed data. *International Journal of Remote Sensing*, **17**, pp. 1317–1340.

HUGHES, G.F., 1968, On the mean accuracy of statistical pattern recognizers. *IEEE Transaction on Information Theory*, **14**, pp. 55–63.

JAIN, A. and ZONGKER, D., 1997, Feature selection: evaluation, application, and small sample performance. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **19**, pp. 153–158.

JENSEN, J.R., 1996, *Introduction to Digital Image Processing: A Remote Sensing Perspective*, 2nd ed. (Piscataway, NJ: Prentice Hall).

LIU, H. and SETIONO, R., 1997, Feature selection via discretization. *IEEE Transactions on Knowledge and Data Engineering*, **9**, pp. 642–645.

LU, D. and WENG, Q., 2007, A survey of image classification methods and techniques for improving classification performance. *International Journal of Remote Sensing*, **28**, pp. 823–870.

MARTÍNEZ-CASASNOVAS, J.A., KLAASSE, A., NOGUÉS, J. and RAMOS, M.C., 2008, Comparison between land suitability and actual crop distribution in an irrigation district of the Ebro valley (Spain). *Spanish Journal of Agricultural Research*, **6**, pp. 700–713.

MELGANI, F. and SEBASTIANO, B.S., 2003, A Markov random field approach to spatio-temporal contextual image classification. *IEEE Transactions on Geoscience and Remote Sensing*, **44**, pp. 2478–2487.

PENG, H., LONG, F. and DING, C., 2005, Feature selection based on mutual information: criteria of max-dependency, max-relevance, and min-redundancy. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **27**, pp. 1226–1238.

PLATT, R.V. and GOETZ, A.F.H., 2004, A comparison of AVIRIS and Landsat for land use classification at the urban fringe. *Photogrammetric Engineering and Remote Sensing*, **70**, pp. 813–819.

PRICE, K.P., GUO, X. and STILES, J.M., 2002, Optimal Landsat TM band combinations and vegetation indices for discrimination of six grassland types in eastern Kansas. *International Journal of Remote Sensing*, **23**, pp. 5031–5042.

PUDIL, P., NOVOVICAOVA, J. and KITTLER, J., 1994, Floating search method in feature selection. *Pattern Recognition Letters*, **15**, pp. 1119–1125.

RAYMER, M.L., PUNCH, W.F. and GOODMAN, E.D., 2000, Dimensionality reduction using genetic algorithms. *IEEE Transactions on Evolutionary Computation*, **4**, pp. 164–171.

REYNOLDS, H.T., 1984, *Analysis of Nominal Data, Series: Quantitative Applications in the Social Sciences*, 2nd ed., pp.15–60 (Newbury Park, CA: Sage Publications).

SEGL, K., ROESSNER, S., HEIDEN, U. and KAUFMAN, H., 2003, Fusion of spectral and shape features for identification of urban surface cover types using reflective and thermal hyperspectral data. *ISPRS Journal of Photogrammetry and Remote Sensing*, **58**, pp. 99–112.

SIEDLECKI, W. and SKLANSKY, J., 1988, On automatic feature selection. *International Journal of Pattern Recognition and Artificial Intelligence*, **2**, pp. 197–220.

STUCKENS, J., COPPIN, P.R. and BAUER, M.E., 2000, Integrating contextual information with per-pixel classification for improved land cover classification. *Remote Sensing of Environment*, **71**, pp. 282–296.

TAN, P.N., STEINBACH, M. and KUMAR, V., 2006, *Introduction to Data Mining* (Boston, MA: Pearson Education).

VAN COILLIE, F.M.B., LIEVEN, P.C. and ROBET, R., 2007, Feature selection by genetic algorithms in object-based classification of IKONOS imagery for forest mapping in Flanders Belgium. *Remote Sensing of Environment*, **110**, pp. 476–487.

VAN NEIL, T.G., MCVICAR, T.R. and DATT, B., 2005, On the relationship between training sample size and data dimensionality: Monte Carlo analysis of broadband multi-temporal classification. *Remote Sensing of Environment*, **98**, pp. 468–480.

VERVERIDIS, D. and KOTROPOULOS, C., 2009, Information loss of the Mahalanobis distance in high dimensions: application to feature selection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **31**, pp. 2275–2281.

WEBB, A.R., 2002, *Statistical Pattern Recognition*, 2nd ed. (Hoboken, NJ: Wiley).

WEISS, N.A., 1995, *Introductory Statistics*, 4th ed. (Reading, MA: Addison-Wesley Publishing Company).

Wu, B., Xiong, Z., Chen, Y. and Zhao, Y., 2009, Classification of quickbird image with maximal mutual information feature selection and support vector machine. *Procedia Earth and Planetary Science*, **1**, pp. 1165–1172.

Xu, H., 2006, Modification of Normalized Difference Water Index (NDWI) to enhance open water features in remotely sensed imagery. *International Journal of Remote Sensing*, **27**, pp. 3025–3033.

Yu, L. and Liu, H., 2004, Efficient feature selection via analysis of relevance and redundancy. *Journal of Machine Learning Research*, **5**, pp. 1205–1224.

Zhang, L., Zhao, Y., Huang, B. and Li, P., 2008, Texture feature fusion with neighborhood oscillating Tabu search for high resolution image. *Photogrammetric Engineering and Remote Sensing*, **74**, pp. 323–332.

Zhang, L., Zhong, Y., Huang, B., Gong, J. and Li, P., 2007, Dimensionality reduction based on clonal selection for hyperspectral imagery. *IEEE Transactions on Geoscience and Remote Sensing*, **45**, pp. 4172–4186.