Land-cover Classification Using ASTER Multi-band Combinations Based on Wavelet Fusion and SOM Neural Network

Hasi Bagan, Qinxue Wang, Masataka Watanabe, Satoshi Kameyama, and Yuhai Bao

Abstract

In this study, we developed a land-cover classification methodology using Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) visible near-infrared (VNIR), shortwave infrared (SWIR), and thermal infrared (TIR) band combinations based on wavelet fusion and the self-organizing map (SOM) neural network methods, and compared the classification accuracies of different combinations of ASTER multi-band data. A wavelet fusion concept named ARSIS (Amélioration de la Résolution Spatiale par Injection de Structures) was used to fuse ASTER data in the preprocessing stage. In order to apply the wavelet fusion method to ASTER data, the principal components of ASTER VNIR data were computed. The first principal component was used as the base image for wavelet fusion. In our experiments, the spatial resolution of ASTER VNIR, SWIR, and TIR data was adjusted to the same 15 m. SOM classification accuracy was increased from 83 percent to 93 percent by this fusion, and classification accuracy increased along with the increase of band numbers. Classification accuracy reaches the highest value when all 14 bands are used, but classification accuracy closely approached the highest value when three VNIR bands, three SWIR bands, and two TIR bands were used. A similar tendency was also obtained by the maximum likelihood classification (MLC) method, but the classification accuracies of MLC over all band combinations were considerably obviously lower than those obtained by the SOM method.

Introduction

The Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) on board the NASA Terra satellite has three visible near-infrared (VNIR) spectral bands in the 0.52 μm to 0.86 μm wavelength region, six shortwave infrared (SWIR) bands in the 1.6 μm to 2.43 μm region, and five thermal infrared (TIR) bands in the 8 μm to 11.65 μm region, with spatial resolutions of 15, 30, and 90 m, respectively. Because ASTER data have wide spectral coverage and relatively high spatial resolution, they can discriminate a variety of surface materials and reduce problems with some lower resolution data resulting from mixed pixels (Yamaguchi *et al.*, 1998). Thus, ASTER data are suitable for land-cover/ land-use classification.

Land-cover classification systems using only the VNIR and SWIR bands of ASTER data has been discussed in some recent papers. The most frequently used method is dividing ASTER data into two groups of images of 15 m and 30 m resolution, each having three and six bands, respectively. For each group, support vector machine (SVM)-based algorithms (Zhu and Blumberg, 2002) or segmentation algorithms (Marcal *et al.*, 2005) are used for classification processing. Other studies using principal component analysis (PCA) were applied to the nine VNIR and SWIR bands. From the previously obtained principal components, a supervised MLC approach was implemented (Gomez *et al.*, 2005). In addition, an approach based on the wavelet fusion method has been proposed (Hasi *et al.*, 2004). However, most of the methods mentioned above have not adopted the TIR band data in classification processing.

ASTER data can be used to perform land-cover classification effectively and accurately, but three problems need to be solved. The first is the difference in spatial resolutions; namely, the different spatial resolutions of VNIR, SWIR, and TIR bands must be converted to the same spatial resolution. Wavelet fusion has proved to be a highly efficient method to deal with data with different spatial resolutions. Ranchin and Wald (2000) launched the ARSIS (Ame'lioration de la Re'solution Spatiale par Injection de Structures) wavelet fusion concept for image fusion applications, and showed that the ARSIS concept achieved the best fusion result when compared with the Brovey transform, Intensity-Hue-Saturation (IHS) method, and principal component analysis (PCA) method. This concept has recently been successfully used for remote sensing image fusion applications (Mertens *et al.*, 2004; Pajares and Cruz, 2004).

Hasi Bagan is with the Institute of Industrial Science, University of Tokyo, Tokyo 153-8505, Japan.

Qinxue Wang: is with the Asia Water Environment Section, National Institute for Environmental Studies, 16-2 Onogawa, Tsukuba, Ibaraki, 305-8506, Japan (wangqx@nies.go.jp).

Masataka Watanabe is with the Faculty of Environmental Information, Keio University, 5522 Endo, Fujisawa, Kanagawa 252-8520, Japan.

Satoshi Kameyama is with the Watershed Ecosystem Section, National Institute for Environmental Studies, 16-2 Onogawa, Tsukuba, Ibaraki, 305-8506, Japan.

Yuhai Bao is with the College of Geographical Science, Inner Mongolia Normal University, Inner Mongolia 010022, China.

Photogrammetric Engineering & Remote Sensing
Vol. 74, No. 3, March 2008, pp. 333–342.

0099-1112/08/7403–0333/$3.00/0

Secondly, a highly accurate classification approach is necessary to classify the multi-dimensional ASTER data. A successful approach to multi-spectral data classification based on artificial neural networks has been addressed in many studies (e.g., multilayer perception (MLP) (Foody, 1999), ARTMAP (Carpenter *et al.*, 1997), radial basis function (Bruzzone and Fernandez-Prieto, 1999), and the SOM algorithm with learning vector quantization (LVQ) (Ito and Omatu, 1999; Ji, 2000). Compared with other classification methods, artificial neural networks have several advantages. They are error-tolerant and relatively insensitive to background noise. They can learn relationships from examples, without making assumptions about data distribution or the nature of the relationship between inputs and outputs (Paola and Showengerdt, 1995). The self-organizing map (SOM) neural network developed by Kohonen (1982) is a prominent unsupervised neural network model that projects a topology-preserving mapping of a high-dimensional input space onto converts a topology-preserving mapping of a high-dimensional input space to a low-dimensional map space. It has been reported that a SOM neural network achieves higher classification accuracy when classifying remote sensing data than the maximum likelihood classification (MLC) method and back-propagation neural network (Ji, 2000).

Finally, the contributions of multiple bands of VNIR, SWIR, and TIR data to classification accuracy need to be analyzed. In this study, in order to solve these problems, we chose the Hetao Irrigation District, Inner Mongolia Autonomous Region, China, and an ASTER scene acquired on 23 August 2003 was employed. The Hetao Irrigation District is located in an arid and semi-arid area, where inappropriate irrigation has caused severe soil salinization, and the soil salinization is likely to result in desertification in future. In order to solve these problems, land-cover types, including soil salinization and desert, need to be identified.

Our procedures in this study were as follows. First, the ARSIS wavelet fusion concept was applied. To solve the problem of spatial resolution, PCA was carried out on the data of three VNIR bands, and the first principal component (PC1) was used as a base image in the wavelet fusion of SWIR and TIR band data. Then, the SOM classification method was carried out on various combinations of the 14 fused bands with a 15 m spatial resolution to show changes of classification accuracy along with the variation of band combinations. At the same time, a traditional MLC method was applied for comparison. Finally, the classification results were analyzed and discussed through statistics, SOM topology preservation maps, and classification result maps.

Methods

Principal Component Analysis

PCA is a traditional method for analyzing multi-spectral remote sensing data, especially for image enhancement and data compression. It conducts a linear transformation of the multi-spectral space of the data to eigen vector space. The result of the principal component is a set of uncorrelated images whose energy variances are ordered by amplitude. The PC1 image contains information that is highly correlated to the three VNIR bands used as the input to PCA (Chavez and Kwarteng, 1989). Then, we normalized the PC1 to float data (0 to 255) for use as a base image in the wavelet fusion step.

Wavelet Fusion

The wavelet fusion approach is appropriate for performing fusion tasks because the multi-resolution analysis approach is well suited to management of different image resolutions (Pajares and Cruz, 2004). The wavelet transform decomposes a signal into a set of basic functions. The base is generated by dilations and translations of single function ψ called the wavelet: i.e.,

$$\psi_{a,b}(t) = \frac{1}{\sqrt{a}}\psi\left(\frac{t-b}{a}\right) \tag{1}$$

where a is the scaling parameter, b is the shifting parameter, $a, b \in R$, and $a > 0$.

In the one-dimensional (1D) case, the 1D continuous wavelet transform of a function f is defined as:

$$W_f(a,b) = \langle f,\psi_{a,b}\rangle = \frac{1}{\sqrt{a}}\int_{-\infty}^{+\infty} f(t)\overline{\psi\left(\frac{t-b}{a}\right)}dt \tag{2}$$

where $\overline{\psi\left(\frac{t-b}{a}\right)}$ is the complex conjugate of ψ. If ψ is such that

$$C_\Psi = \int_{-\infty}^{+\infty} \frac{|\Psi(\omega)|^2}{\omega}\, d\omega < +\infty, \tag{3}$$

f can be reconstructed by an inverse wavelet transform:

$$f(t) = \frac{1}{C_\Psi}\int_0^{+\infty}\int_{-\infty}^{+\infty} W_f(a,b)\psi_{a,b}(t)db\frac{da}{a^2}. \tag{4}$$

The binary partition discrete wavelets given by $a = 2^j$, and $b = 2^j k$ are:

$$\psi_{j,k} = 2^{-\frac{j}{2}}\psi(2^{-j}t - k), \quad j,k \in Z. \tag{5}$$

The multi-resolution formulation needs two closely related basic functions. In addition to the wavelet $\psi(x)$, we will need another basic function, called the scaling function $\varphi(x)$. $\varphi(x)$ can be expressed in terms of a weighted sum of shifted $\varphi(2x)$ as, for example, such as Daubechies four-tap filter (Daubechies, 1992) (Table 1):

$$\varphi(x) = \sqrt{2}\sum_{n=0}^{N-1} h(n)\varphi(2x-n). \tag{6}$$

The corresponding wavelet can be generated by the following equation:

$$\psi(x) = \sqrt{2}\sum_{n=0}^{N-1} g(n)\varphi(2x-n). \tag{7}$$

In general, we consider $h(n)$ as a low-pass filter (scaling coefficients) and $g(n)$ as a high-pass filter (wavelet coefficients) where:

$$g(n) = (-1)^n h(N-1-n), \tag{8}$$

and N is the wavelet index.

The mother wavelet $\psi(x)$ is good at representing the detail and high-frequency parts of a signal. The scaling function $\varphi(x)$ is good at representing the smooth and

TABLE 1. COEFFICIENTS OF H (LOW-PASS) FILTER USED TO GENERATE FOUR-TAP DAUBECHIES WAVELET (DAUBECHIES, 1992)

$H(0)$	$H(1)$	$H(2)$	$H(3)$
$\frac{1+\sqrt{3}}{4\sqrt{2}}$	$\frac{3+\sqrt{3}}{4\sqrt{2}}$	$\frac{3-\sqrt{3}}{4\sqrt{2}}$	$\frac{1-\sqrt{3}}{4\sqrt{2}}$

low-frequency parts of the signal. In forward wavelet analysis, a J-level discrete decomposition can be written as:

$$f(x) = \sum_n c_n^0 \varphi_n^0(x - n) = f_1 + g_1 = f_J + g_J + \cdots + g_2 + g_1$$
$$= f_J + \sum_{j=1}^{J} g_j \tag{9}$$

where $f_J = \sum_k c_k^J \varphi_k^J(x)$, $g_j = \sum_k d_k^j \psi_k^j(x)$, and coefficients c_k^j and d_k^j at resolution j are related to the coefficients c_k^{j-1} at level $j - 1$ by the following recursive equations:

$$c_k^j = \sum_{n \in Z} h(n - 2k) c_k^{j-1} \quad \text{and} \quad d_k^j = \sum_{n \in Z} g(n - 2k) c_k^{j-1} \tag{10}$$

for $j = 1, 2, \ldots, J$.

In an inverse wavelet transform, a reconstruction of the original fine scale coefficients of the signal can be made from a combination of the scaling coefficients and wavelet coefficients at a coarse resolution. Because all of these functions are orthonormal, we have

$$c_k^j = \sum_{n \in Z} h(k - 2n) c_n^{j+1} + \sum_{n \in Z} g(k - 2n) d_n^{j+1}. \tag{11}$$

The analysis and synthesis procedures lead to the pyramid-structured wavelet decomposition (Mallat, 1989). The 1D multi-resolution wavelet decomposition can be easily extended to two-dimensional (2D) by introducing separable 2D scaling and wavelet functions as the tensor products of their 1D complements.

When an $M \times N$ discrete image $F(m, n)$ is observed, that is, the 2D function is sampled. The 2D discrete wavelet transform (DWT) maps the image $F(m, n)$ to an $M \times N$ matrix of wavelet coefficients. For computer implementation, it is desirable that the dimensions M and N be a power of 2.

The 2D DWT wavelet analysis operation consists of filtering and down-sampling using 1D low- and high-pass filters L (with impulse responses $h(i)$) and H (with impulse responses $g(i)$), first horizontally (by rows) and then vertically (by columns). So, applying the 2D DWT to the original image $F(m, n)$ once, we get four lower resolution sub-images consisting of wavelet coefficients: $F_1^{LL}(m,n)$, $F_1^{LH}(m,n)$, $F_1^{HL}(m,n)$, and $F_1^{HH}(m,n)$ for $m = 1, \ldots, M/2$, $n = 1, \ldots, N/2$. The second transform is just to split $F_1^{LL}(m,n)$ in the same way, which produces the decomposition at the second level: $F_2^{LL}(m,n)$, $F_2^{LH}(m,n)$, $F_2^{HL}(m,n)$, and $F_2^{HH}(m,n)$ for $m = 1, \ldots, M/4$, $n = 1, \ldots, N/4$. General expressions for the higher levels can easily be derived.

The ARSIS concept, based on discrete wavelet transform, is used in this study. The goal of the ARSIS concept is to achieve high spatial resolution together with high-quality spectral content from two kinds of remote sensing images: (a) images with high quality spectral content but low quality spatial resolution, and (b) images with high quality spatial resolution but with a unique spectral band. Figure 1 presents the application of the ARSIS concept to fusion of ASTER SWIR and TIR imagery. Many filters are possible, but in this study the Daubechies four-tap filter is applied. Table 1 gives the coefficients of the four-tap filter designed by Daubechies (1992).

In Figure 1, P indicates the PC1 image at the spatial resolution of 15 m and S indicates the SWIR band. The wavelet transform is applied to each SWIR band separately. Two iterations of the multi-resolution analysis using the wavelet transform are applied to the original P image, and one iteration to the original S image.

ARSIS concepts are computed for the transformation of each P wavelet coefficient image P_2^{HL}, P_2^{LH}, and P_2^{HH} into each S wavelet coefficient image S_1^{HL}, S_1^{LH}, and S_1^{HH}. Then, these models are applied to the wavelet coefficient images P_1^{HL}, P_1^{LH}, and P_1^{HH} for the computation of the missing wavelet coefficient images S^{HL}, S^{LH}, and S^{HH}. The superscripts *HL, LH*, and *HH* indicate horizontal, vertical, and diagonal coefficient images, respectively. Finally, the synthesis step reconstructs the 15 m spatial resolution S image (S-HR). The wavelet coefficient images computed between 15 m and 30 m can be expressed as follows:

$$a^Z = \sigma^Z(S)/\sigma^Z(P),\ b^Z = m^Z(S) - a^Z m^Z(P); \tag{12}$$

for $Z = HL, LH$, and HH:

$$S^Z = a^Z P_1^Z + b^Z \quad \text{for } Z = HL, LH, \text{ or } HH \tag{13}$$

where $m^Z(S)$ and $m^Z(P)$ are the means of S_1^Z and P_2^Z, respectively, and $\sigma^Z(S)$ and $\sigma^Z(P)$ are the standard deviation of S_1^Z and P_2^Z, respectively.

To meet the demand of the 2D discrete wavelet transform (Pajares and Cruz, 2004), we used the nearest neighbor interpolation technique to resample TIR image pixel size from 90 m × 90 m to 60 m × 60 m. T is used to indicate each of the multi-spectral TIR image bands from 10 to 14 at a spatial resolution of 60 m. Here, two wavelet coefficients images need to be synthesized, i.e., between 60 and 30 m and between 30 and 15 m. Then, the equations for the ARSIS model are:

$$a^Z = \sigma^Z(T)/\sigma^Z(P),\ b^Z = m^Z(T) - a^Z m^Z(P); \tag{14}$$

for $Z = HL, LH$, or HH:

$$T^Z = a^Z P_2^Z + b^Z \quad \text{for } Z = HL, LH, \text{ or } HH \tag{15}$$

$$FT^Z = a^Z P_1^Z + b^Z \quad \text{for } Z = HL, LH, \text{ or } HH. \tag{16}$$

Similar to the notations above, $m^Z(T)$ and $m^Z(P)$ are the means of T_1^Z and P_3^Z, respectively; the $\sigma^Z(T)$ and $\sigma^Z(P)$ are the standard deviation of T_1^Z and P_3^Z, respectively.

SOM Neural Network Classification

A SOM is an unsupervised and nonparametric neural network approach. By assigning each input vector to the neuron with the nearest weight vector, a SOM is able to divide the input space into regions with common nearest weight vectors. Also, because the neighborhood relationship contributes to the inter-connections among neurons, a SOM exhibits another important property of topology preservation. That is, if two weight vectors are near each other in the input space, the corresponding neurons will also be close in the output space, and vice versa. Generally, the output neurons are arranged in a two-dimensional grid of map units.

Because a SOM defines a neighborhood relationship in the competitive layer, data in the input space with the same topological characters can be exported to special neurons or neighborhood neurons in the competitive layer. A cluster center is expressed as a weight vector. During the clustering process, not only the weight of the winning neuron is updated, but also the weights of all the neurons in the neighborhood of the winning neuron are updated. The update pulls the winning neuron and it neighbors closer to the input vector.

Provided that there are T neurons in the competitive layer, and neuron unit j is represented by a prototype vector, $w_j = [w_{j1}, \ldots, w_{jn}]$ where n is the input vector dimension. At each training step t, an input vector $x = [x_1, \ldots, x_n]$ is randomly chosen from the training set. Euclidean distance

$$\|x - w_j\| = \left(\sum_{i=1}^{n} (x_i - w_{ji})^2 \right)^{0.5} \tag{17}$$

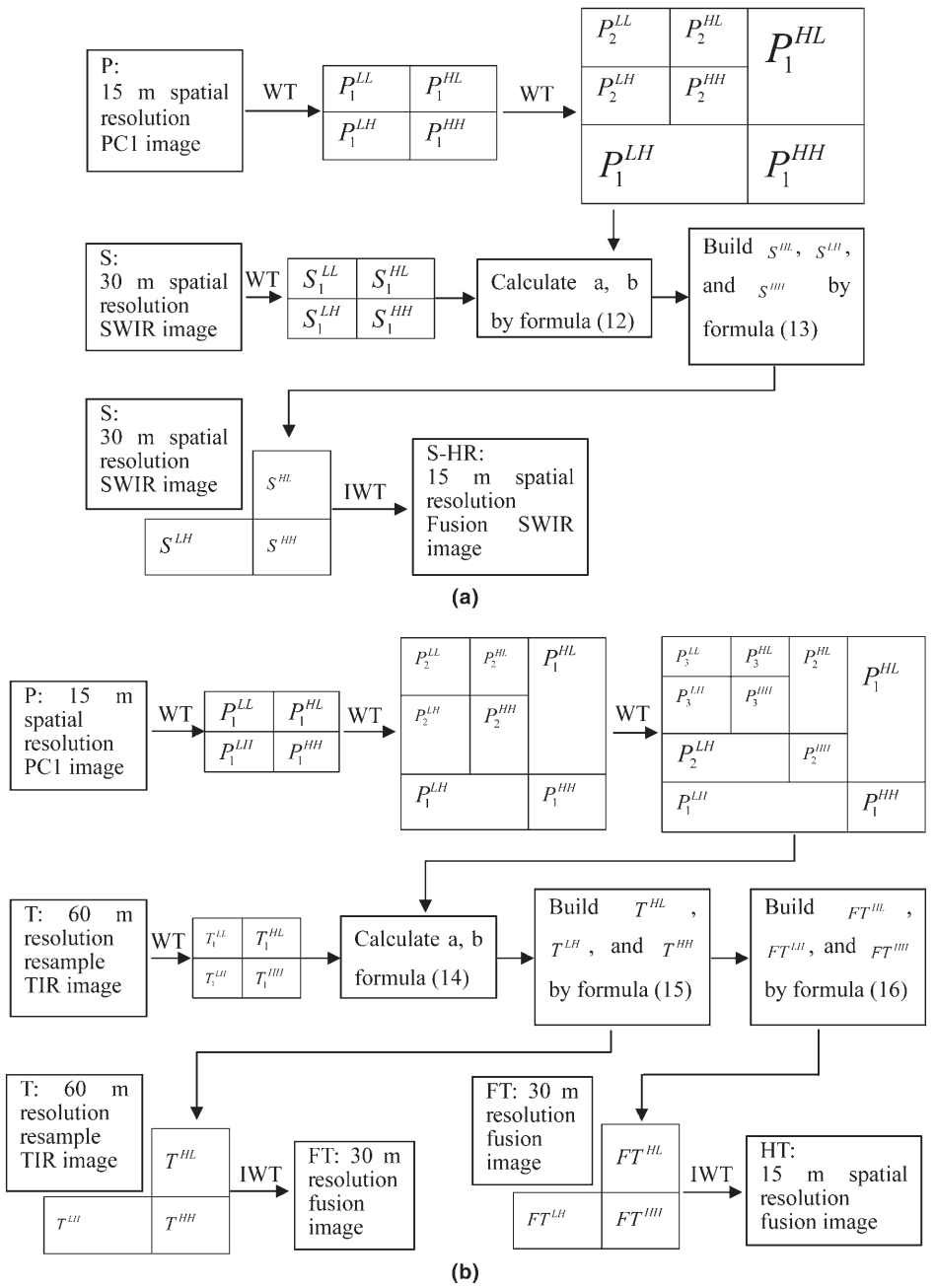

Figure 1. Flowchart of SWIR band fusion (a) and TIR band fusion (b) scheme based on ARSIS concept. WT denotes wavelet transform, and IWT denotes inverse wavelet transform.

was chosen as the distance measure to determine the winning neuron. Pick any input vector x and compute:

$$\|x - w_c\| = \min\{\|x - w_j\|\},\ j \in \{1 \cdots\cdots T\} \tag{18}$$

where c is the winning neuron.

The winning neuron and its topological neighbors are moved closer to the input vector using:

$$w_{ji}(t+1) = w_{ji}(t) + \alpha(t)\lfloor x_i - w_{ji}(t)\rfloor \quad \text{if } j \in N_c(t) \tag{19}$$

$$w_{ji}(t+1) = w_{ji}(t) \quad \text{if } j \notin N_c(t) \tag{20}$$

where $Nc(t)$ is the set of neighborhood nodes of the winning neuron c at time t, $\alpha(t)$ is the learning rate, and its initial value is set as $0 < \alpha(t) < 1$.

Afterwards, a supervised learning technique, learning vector quantization (LVQ), is applied to the results of the SOM to fine-tune the weight vectors using input vectors x of known classification. Finally, the distribution of weight vectors of the neurons in the competitive layer becomes the best match for vectors of input samples of known classification. The details of the process can be described as follows (Ji, 2000).

At each training step t, initially, a sample data vector x is randomly chosen from the training set, and the winning neuron c is calculated according to Equation 18. If x and neuron c are in the same class, vector w_c of neuron c is updated using Equation 21:

$$w_{ci}(t+1) = w_{ci}(t) + \alpha(t)[x_i - w_{ci}(t)], \quad i \in \{1 \ldots\ldots T\}. \tag{21}$$

Otherwise, using Equation 22:

$$w_{ci}(t+1) = w_{ci}(t) - \alpha(t)[x_i - w_{ci}(t)], \quad i \in \{1 \ldots\ldots T\}. \tag{22}$$

Study Area and Data Used

The study area is located in the Hetao Irrigation District, Inner Mongolia Autonomous Region, China. The cloud-free ASTER scene used in this study was acquired on 23 August 2003 (Figure 2). The coordinates are in the WGS84 UTM Zone 49. The size of study area is 4,096 pixels × 4,096 pixels (15 m spatial resolution), which was clipped from the ASTER Level 1B data. All 14 bands of images of the VNIR, SWIR, and TIR instruments were used. Because the atmospheric effects are nearly homogeneous in an arid region at the scale of an ASTER scene, we used the ASTER Level 1B data instead of the Level 2 product in this study.

The major part of the study area is the Hetao Irrigation District, which covers the Wuliangsuhai Lake (very shallow; areas of water 0.8 to 1.0 m deep occupy 80 percent of the total lake area), a part of the Kubuqi Desert, and the Wulashan Mountains, where forest and shrubs are very sparse. Because this area is located in the middle of dry grassland on the edge of the Gobi Desert in Inner Mongolia, agriculture is impossible without irrigation. However, due to high evapotranspiration, the alkalinity of the soil has become a serious problem. Especially in the area around the Wuliangsuhai Lake, soil is rapidly alkalized, and those lands are nearly bare of vegetation.

In order to understand the actual ground situation, we conducted field surveys several times in 2000, 2002, and 2003, when we took photos for actual ground data collection and field validation of sites located by GPS facilities. According to the results of field surveys, we realized that our study area is mainly covered by nine land-cover types. Table 2 shows the description of the land-cover classes.

Except for the field surveys, two land-cover digital maps were also referenced: one is the National Land-use/Land-cover Datasets (NLCD-2000) of China which was generated from extensive field surveys, Landsat TM/ETM data in 2000 (Liu *et al.*, 2003), and another is the land-use/land-cover vector digital map of Wulateqianqi County that was generated from ground surveys and interpretation of aerial photos taken in 2002 (done by Inner Mongolia Normal University, unpublished data, 2003). The same classification system of 25 land-cover types was used for classification in these two maps, with a mapping scale of 1:100 000. Our study area was fully covered by NLCD-2000, and the major part of our study area was covered by the Wulateqianqi County-level map. Besides these, 1:10 000-scale aerial photographs acquired in June 2004 in Wulateqianqi County were also referenced.

Finally, based on the above-mentioned actual ground collection and digital maps, we selected both training and

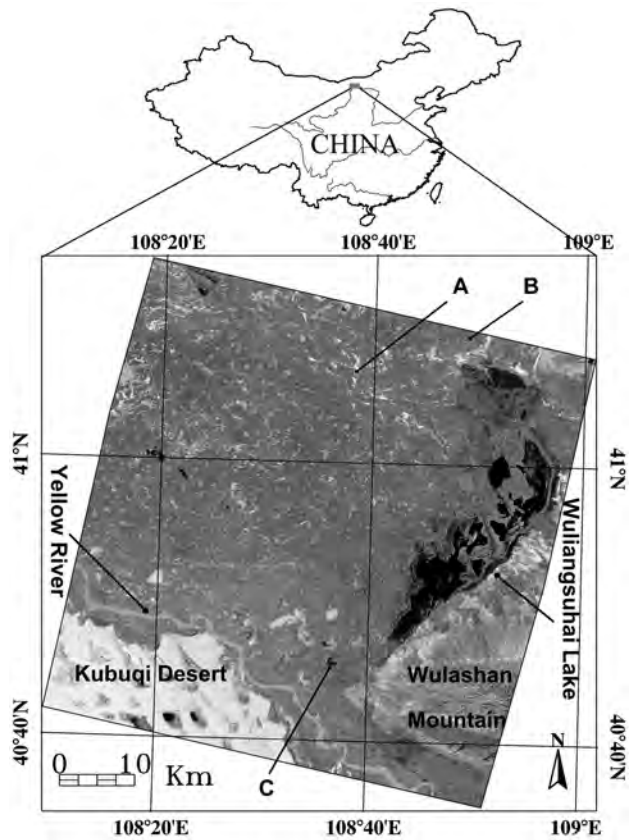

Figure 2. Location of false colored ASTER image (RGB = bands 3, 2, and 1) of study area in the Hetao Irrigation District, Inner Mongolia Autonomous Region, China, acquired on 23 August 2003. The letters A, B, and C indicate the image sections shown in Plate 2.

TABLE 2. DESCRIPTION OF LAND-COVER CLASSIFICATION SYSTEM AND PIXEL COUNTS BY LAND-COVER CLASSES

Land-cover Class	Class Description	Training Pixel Count	Test Pixel Count
1. Water	Water bodies	2,523	411
2. Reed (water)	Mixture of water and herbaceous (mainly reeds, etc.) or woody vegetation	2,315	416
3. Grasslands	Dominated by grass, occasional tree and shrub	2,447	402
4. Cropland	Land dedicated to the production of crops	2,181	402
5. Desert	Exposed sand	2,095	413
6. Urban/Built-up	Buildings and other human-made structures	2,057	406
7. Open area	Exposed soil and stone	2,108	410
8. Tidal flat	Muddy, sandy or mixed sediment land	2,345	413
9. Saline soil	Alkalized/Stalinized soils without recognizable plant life	1,185	413
Total		19,256	3,686

testing data (Congalton and Green, 1999; Lillesand *et al.*, 2004). Here, the testing data were chosen independent of training data and could represent the various land-cover types. Table 2 shows the counts of both training and testing samples.

Methodology

Data Fusion and Preprocessing

The ARSIS concept of the ASTER data wavelet fusion method was performed using a C/C++ program (Microsoft® Visual C++ 6.0). After fusion, the 14 bands of ASTER data have a common spatial resolution of 15 m. Under the same spatial resolution, the number of possible false-color composite image band combinations is increased from C_3^3 (=1) to C_{14}^3 (=728) through this process. Thus, through various arbitrary combinations of VNIR, SWIR, and TIR bands, images can be conveniently interpreted, and ambiguous ground objects can be easily detected and characterized. Figure 3 illustrates the R (2, VNIR) G (7, SWIR), B (14, TIR) false-color composite image. In Figure 3, the polluted water in the Yellow River and the alkaline lands distributed among croplands are more obvious than in Figure 2. Four false-color composite sub-images of R (3) G (2) B (1) and R (2) G (7) B (14) in Positions A and B of Figure 3 are illustrated in Figure 4, the before-fusion images on the left, and the after-fusion images on the right. The alkaline land on the left of Figure 4c is not too obviously different from either of its surroundings or the desert on the left of Figure 4a and the Kubuqi Desert in the lower left corner in Figure 2, which is the combination of R(3), G(2), B(1), while in Figure 4d, the alkaline land is yellow. The differences between its surroundings and the Kubuqi Desert in Figure 3 are both enhanced.

Classification and Performance Evaluation

The SOM neural network structure used for the ASTER image classification was set as follows: the input layer node number is equal to the number of ASTER bands used. The neurons in the competitive layer are in two dimensions of 20 × 20. Based on testing, the parameters of the network were set as the maximum iteration time $T = 3000$, the initial learning rate $\alpha(0) = 0.9$, and the descending learning rate $\beta = 0.0015$; the radius of $N(0)$, the initial neighborhood (rectangular lattice), was set to 12; the initial weight vector value ($w_i(0)$) of neuron i in the competitive layer was set randomly. The radius of the neighborhood, $N(t)$, decreased according to Equation 23 until the neighborhood $N(t) = 0$ meets:

$$N(t) = N(0) - [A \times (t + A \times t)], \quad \text{where } A = N(0)/T \tag{23}$$

where t is the training time, $N(t)$ is the neighborhood of the array points around the winning neuron, and [v] denotes a maximum integer that does not exceed v. The learning rate is defined as decreasing along with the increasing number of iterations (Equation 24) until learning rate $\alpha(t) = 0.0025$ meets:

$$\alpha(t) = \begin{cases} \alpha(t-1) - \beta/10, & 0 < t < T/4 \\ \alpha(t-1) - \beta/5, & T/4 \le t < \mathrm{T}/2. \\ \alpha(t-1) - \beta, & t \ge \mathrm{T}/2 \end{cases} \tag{24}$$

Neurons in the competitive layer were assigned to corresponding classes according to a majority voting principle while the coarse tuning process was finished.

A fine-tuning process was applied to the training results using the LVQ supervised method when the SOM learning process was finished. The maximum number of iterations was set to 1,000. The initial value of the learning rate was

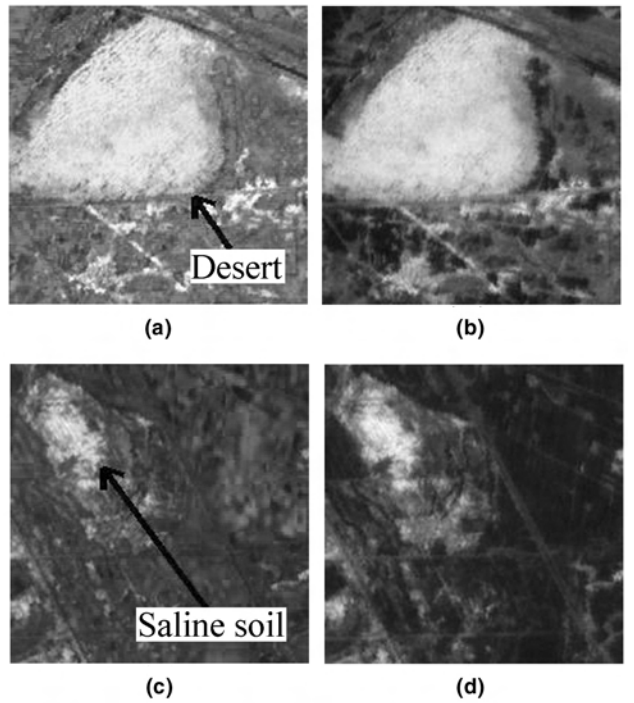

Figure 4. Comparison of false-color composite ASTER images indicated in Figure 3: (a) Image indicated by A, RGB = bands 3, 2, and 1, (b) Image indicated by A, RGB = bands 2, 7, and 14, (c) Image indicated by B, RGB = bands 3, 2, and 1, and (d) Image indicated by B, RGB = bands 2, 7, and 14.

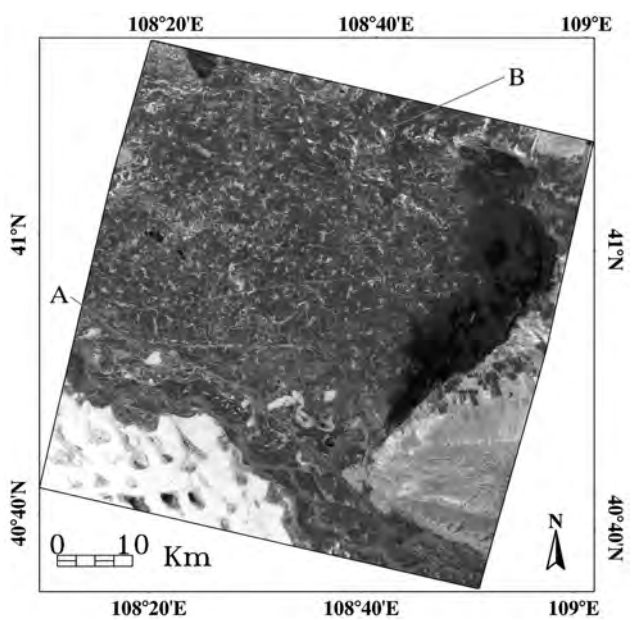

Figure 3. RGB = band 2 (VNIR), band 7 (SWIR), and band 14 (TIR) false-color composite ASTER image of the study area. The letters A and B indicate the image sections shown in Figure 4.

0.25, and the descending learning rate was 0.000275. The learning rate decreased along with the increasing number of iteration times until learning rate 0.00025 was met. All these parameters were predefined for the LVQ adjustment.

Table 3 shows the different band combinations that vary from very simply three VNIR bands to all 14 bands including VNIR, SWIR, and TIR bands. The each-band combinations listed in Table 3 is classified by the SOM method on the same training data. The test data is used to evaluate the significance of differences in classification accuracy by different band combinations.

To compare the effects of different methods, images of the different band combinations listed in Table 3 were also classified by the MLC method on the same training and testing data. SOM classification methods were performed using the C/C++ program (Microsoft® Visual C++ 6.0), and MLC classifications were performed by ENVI™ software, version 4.0. Table 4 shows the classification accuracies and Kappa coefficients achieved by both SOM and MLC methods for each band combination in Table 3.

Results and Discussion

Table 4 shows the classification accuracies and Kappa coefficients of both SOM and MLC, along with the increase in band numbers. With both SOM and MLC methods, the classification accuracy increases as the number of bands increases. But, in SOM classifications, when all VNIR bands and a part of the SWIR and TIR bands are used, the accuracy is higher than that using all VNIR and SWIR bands. For example, the classification accuracy of the band combination of three VNIR, three SWIR, and two TIR bands was higher than that of use all VNIR and SWIR bands and close to the accuracy of all 14 bands. Moreover, in all the band combinations, the classification accuracies and Kappa coefficients of the SOM are generally higher than those of the corresponding MLC classification.

The land-cover classification results using the SOM and MLC methods on 14 bands are presented in Plate 1a and Plate 1b, respectively.

TABLE 3. LAND-COVER CLASSIFICATION ANALYSIS FOR VARIOUS BAND COMBINATIONS OF ASTER

Number of Bands	Band Combination
3	1 2 3
6	1 2 3 4 5 8
7	1 2 3 4 5 8 13
8	1 2 3 4 5 8 10 13
9	1 2 3 4 5 6 7 8 9
10	1 2 3 4 5 6 7 8 9 13
12	1 2 3 4 5 6 7 8 9 10 12 14
14	1 2 3 4 5 6 7 8 9 10 11 12 13 14

Comparison between all classifications identified by the two kinds of classification methods are described as follows (Table 5):

A confusion matrix was computed to evaluate the ability of the SOM method to identify land-cover and to assess its accuracy compared with the MLC classifier. The confusion matrix is a commonly used tool for assessment of accuracy of land-cover classification. The matrix scores how the classification process has labeled a series of test sites or test pixels of which the correct land-cover label is known (Foody, 2004; Wilkinson, 2005). Tables 6 and 7 show the confusion matrixes for classifications by the SOM and MLC methods on 14 bands, respectively. The true class labels are displayed across rows, while the actual mapped classes are displayed in columns. The diagonal of the confusion matrix displays the number of pixels at which the true class and the mapped class agree with each other.

Table 6 shows the accuracies of land-cover classification using 14 bands of data with the SOM. Aside from grassland and open areas, the accuracies of the SOM are higher than those of the MLC. One possible explanation is that the MLC approach assumes that the probability distributions of the classes are in the form of multivariate normal models. This is an assumption rather than a demonstrable property of natural spectral or information classes, while the SOM does not need the assumption. For instance, the large spectral variance within the cropland land-cover classes makes accurate classification difficult. Nonetheless, the SOM is able to classify cropland territory with an accuracy of 98.01 percent. MLC, however, has an accuracy of only 80.1 percent for this class.

From Figure 2, Table 5, and Plate 1, we can see that the MLC magnified grassland area and reduced cropland, while the SOM correctly revealed these local conditions. Sub-images of ASTER combination R(3), G(2), B(1) at Positions A, B, and C in Figure 2, respectively, and sub-images of the SOM and MLC classification results at corresponding positions were extracted to evaluate partial classification effects in Plate 2. In Sub-image A of Plate 2, the SOM extracted the saline soil class correctly, while MLC magnified the saline soil area. In Sub-image B, the SOM extracted cropland class correctly, while MLC reduced the cropland region and magnified the grassland area. In Sub-image C, the SOM sketched the tidal flat range correctly, while the MLC had a relatively large classification error.

On the other hand, the SOM did not correctly identify the open area class in the Wulashan mountainous area (lower right in Figure 2), and it tended to assign exposed rock to the urban/built-up class, while the MLC performed better than the SOM. However, for small towns and villages, the classifications by both the SOM and MLC were not adequate. The MLC magnified open areas and reduced housing estates, while the SOM did just the opposite. One reason is probably because the roofs of the houses

TABLE 4. CLASSIFICATION ACCURACY AND KAPPA COEFFICIENTS FOR THE VARIOUS BAND COMBINATIONS

	Overall Accuracy							
	3 Band	6 Band	7 Band	8 Band	9 Band	10 Band	12 Band	14 Band
MLC	78.0521	82.3114	82.7998	84.0477	83.5865	84.5632	87.0863	87.3576
SOM	83.1796	89.2566	91.834	92.2952	90.9387	92.3766	92.675	93.1362
	Kappa Coefficient							
	3 Band	6 Band	7 Band	8 Band	9 Band	10 Band	12 Band	14 Band
MLC	0.7531	0.801	0.8065	0.8206	0.8154	0.8264	0.8547	0.8578
SOM	0.8108	0.8791	0.9081	0.9133	0.8981	0.9142	0.9176	0.9228

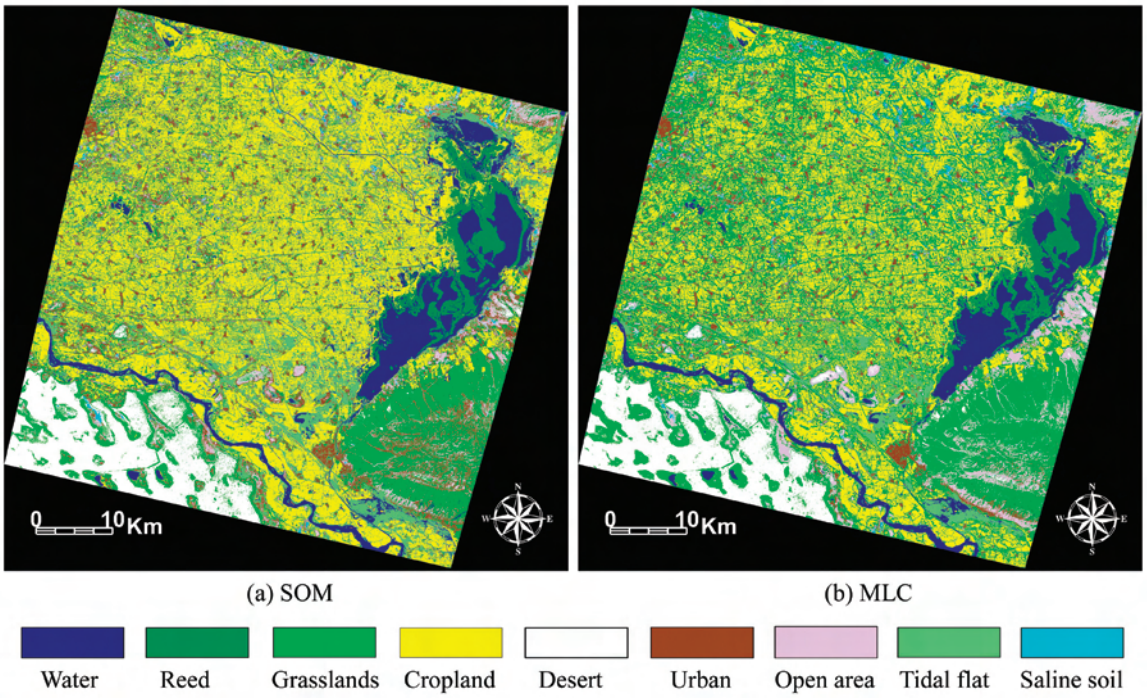

Plate 1. Classification results with ASTER 14-band data: (a) SOM neural network method, and (b) MLC method. The overall accuracies of test data using the SOM and MLC methods were 93 percent and 87 percent, respectively.

TABLE 5. COMPARISON OF SOM AND MLC CLASSIFIED LAND-COVER MAPS SUMMARIZED USING CLASS PERCENTAGES

Class	1.	2.	3.	4.	5.	6.	7.	8.	9.
MLC	6.34%	3.30%	32.48%	30.66%	7.06%	5.75%	5.53%	7.17%	1.70%
SOM	7.31%	4.97%	15.26%	37.48%	7.20%	9.27%	3.96%	14.12%	0.43%

TABLE 6. CONFUSION MATRIX FOR THE CLASSIFICATION PERFORMED BY THE SOM NEURAL NETWORK RESULTS WITH TEST DATA

Class	1.	2.	3.	4.	5.	6.	7.	8.	9.	Σ	User Acc.
1.	402	0	4	0	0	0	0	1	0	407	98.77
2.	0	416	0	0	0	0	0	0	0	416	100
3.	3	0	321	3	7	11	19	2	0	366	87.70
4.	0	0	5	394	0	0	0	3	0	402	98.01
5.	0	0	2	0	406	0	1	0	4	413	98.31
6.	0	0	22	0	0	357	55	0	3	437	81.69
7.	0	0	28	0	0	14	331	0	7	380	87.11
8.	6	0	20	5	0	23	0	407	0	461	88.29
9.	0	0	0	0	0	1	4	0	399	404	98.76
Σ	411	416	402	402	413	406	410	413	413	3686	
Prod. Acc.	97.81	100	79.85	98.01	98.31	87.93	80.73	98.55	96.61		

Overall Accuracy: (3433/3686) 93.1362%, Kappa Coefficient = 0.9228

are mainly made of concrete and adobe blocks in the urban/built-up class of our study area. Thus, urban/built-up and open areas are difficult to discriminate effectively. This fact can be proved by the 2D topology preservation map of 14-band SOM training results shown in Figure 5. After SOM training, each class was clustered into respective conjoint topology regions, but the distribution locations of urban/built-up and open areas were adjacent and partially intersecting in the 2D plane, which affected the classification accuracies of the two classes by the SOM. In addition, water was distributed into two distinct regions because the water spectrum of the Yellow River is different from that of other water bodies. Thus, Yellow River water and other waters were clustered into different clustering regions, although this did not affect the water classification accuracy by the SOM.

Table 7. Confusion Matrix for the Classification Performed by the Maximum Likelihood Classification (MLC) Results with Test Data

Class	1.	2.	3.	4.	5.	6.	7.	8.	9.	Σ	User Acc.
1.	337	2	0	0	0	0	0	86	0	425	79.29
2.	0	411	0	0	0	0	0	0	0	411	100
3.	65	1	341	80	8	5	36	22	0	558	61.11
4.	0	2	1	322	0	0	0	0	0	325	99.08
5.	0	0	4	0	405	0	7	0	0	416	97.36
6.	0	0	11	0	0	372	8	1	0	392	94.9
7.	0	0	45	0	0	19	358	0	42	464	77.16
8.	9	0	0	0	0	10	0	303	0	322	94.1
9.	0	0	0	0	0	0	1	1	371	373	99.46
Σ	411	416	402	402	413	406	410	413	413	3686	
Prod. Acc.	82	98.8	84.83	80.1	98.06	91.63	87.32	73.37	89.83		

Overall Accuracy = 87.3576% (3220/3686), Kappa Coefficient = 0.8578

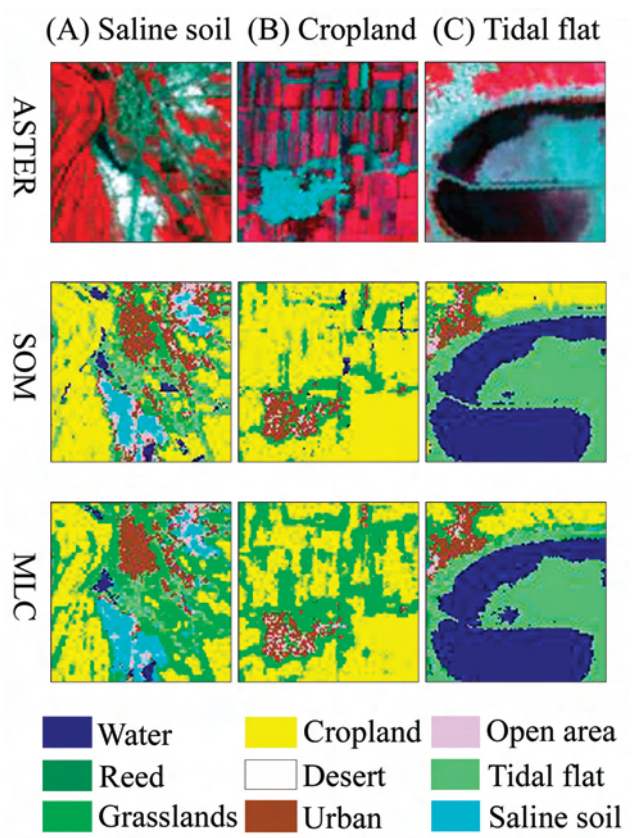

Plate 2. Three original ASTER image subsets (100 pixels × 100 pixels, RGB = 3,2,1) for the location indicated in Figure 2, and the corresponding classification results from SOM and MLC (top) original ASTER image subsets, (middle) classification maps obtained by SOM classification, and (bottom) classification maps obtained by MLC classification.

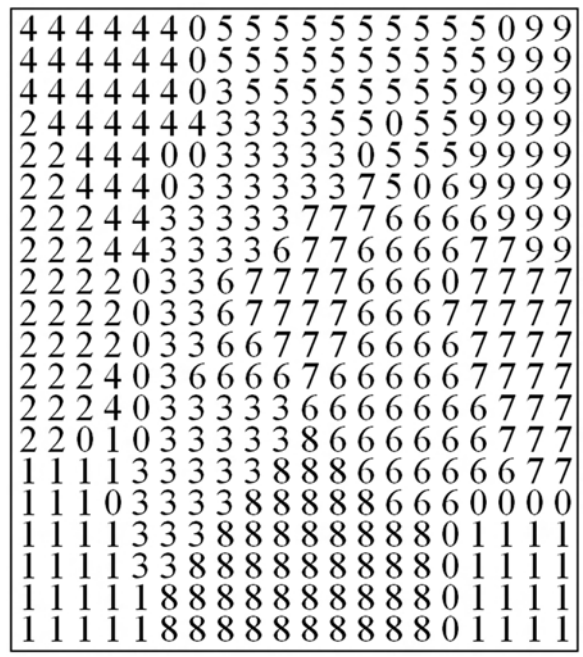

Figure 5. SOM training result shows nine land-cover types located in a 2D topology preservation map

Conclusions

Wide availability of high-quality multi-spectral and multi-spatial resolution satellite images such as ASTER data allow us to update land-cover maps more frequently at lower cost. To solve the problem of different pixel resolutions of the image bands, we need to develop algorithms for more accurate and sophisticated classification methods. In this study, we combined a SOM neural network classification method with ARSIS wavelet fusion. The ARSIS wavelet fusion was used to transform an image into a set of bands with the same spatial resolution, and the SOM classified the multi-bands data effectively. The results show that after ARSIS wavelet fusion, images are easier to interpret and distinguish. The SOM classification results indicate that the classification accuracy increases with the increase in the number of bands used. In particular, when some SWIR and TIR bands and all three VNIR bands were used, the accuracy exceeded that when all VNIR and SWIR bands were used, and the accuracy was close to that using all 14 bands. MLC classifications exhibited a similar trend, which confirmed the SOM result. Moreover, the classification accuracies of the SOM were generally higher than those of the corresponding MLC classifications, which was demonstrated by statistical results and sub-image comparisons. Finally, we concluded that the SOM neural network combined with the ARSIS wavelet fusion method improved the accuracy of ASTER data classification significantly.

Acknowledgments

This research was funded by the project "Establishment of an Early Detection Network for the Impacts of Global Warming" from the National Institute for Environmental Studies (NIES), Japan and the Japan Society for the Promotion of Science (JSPS). The authors would like to thank Professor Yoshifumi Yasuoka and Dr. Wataru Takeuchi of the Institute of Industrial Science, The University of Tokyo for the useful comments and suggestions. The authors also wish to thank three anonymous reviewers who provided thoughtful comments that greatly improved this manuscript from its original form.

References

Bruzzone, L., and D. Fernandez-Prieto, 1999. A technique for the selection of kernel-function parameters in RBF neural networks for classification of remote-sensing images, *IEEE Transactions on Geoscience and Remote Sensing*, 37:1179–1184.

Carpenter, G.A., M.N. Gjaja, S. Gopal, and C.E. Woodcock, 1997. ART neural networks for remote sensing: Vegetation classification from Landsat TM and terrain data, *IEEE Transactions on Geoscience and Remote Sensing*, 35:308–325.

Chavez, P.S. Jr., and A.W. Kwarteng, 1989. Extracting spectral contrast in Landsat Thematic Mapper image data using selective principal component analysis, *Photogrammetric Engineering & Remote Sensing*, 55(3):339–348.

Congalton, R.G., and K. Green, 1999. *Assessing the Accuracy of Remotely Sensed Data: Principles and Practices*, Lewis Publishers, Boca Raton, Florida, 137 p.

Daubechies, I., 1992. *Ten Lectures on Wavelets*, Society for Industrial and Applied Mathematics, Philadelphia, Pennsylvania, 357p.

Foody, G.M., 1999. The signification of border training patterns in classification by a feedforward neural network using back propagation learning, *International Journal of Remote Sensing*, 20:3549–3562.

Foody, G. M., 2004. Thematic map comparison evaluating the statistical significance of differences in classification accuracy, *Photogrammetric Engineering & Remote Sensing*, 70(6):627–633.

Gomez, C., C. Delacourt, P. Allemand, P. Ledru, and R. Wackerle, 2005. Using ASTER remote sensing data set for geological mapping, in Namibia, *Physics and Chemistry of the Earth*, 30:97–108.

Hasi, B., J. Ma, Q. Li, Z. Liu, and X. Han, 2004. Land-use classification from ASTER data using wavelet fusion and self-organized neural networks, *Science in China Series D-Earth Sciences*, 47:651–658.

Ito, Y., and S. Omatu, 1999. Extended LVQ neural network approach to land cover mapping, *IEEE Transactions on Geoscience and Remote Sensing*, 37:313–317.

Ji, C.Y., 2000. Land-use classification of remotely sensed data using Kohonen self-organizing feature map neural networks, *Photogrammetric Engineering & Remote Sensing*, 66(12):1451–1460.

Kohonen, T., 1982. Self-organized formation of topologically correct feature maps, *Biological Cybernetics*, 43:59–69.

Lillesand, T.M., R.W. Kiefer, and J.W. Chipman. 2004. *Remote Sensing and Image Interpretation*, Fifth edition, John Wiley and Sons, New York, 763 p.

Liu, J., M. Liu, D. Zhuang, Z. Zhang, and J. Deng, 2003. Study on spatial pattern of land-use change in China during 1995–2000, *Science in China Series D-Earth Sciences*, 46:373–384.

Mallat, S.G., 1989. A theory for multiresolution signal decomposition: The wavelet representation, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11:674–693.

Marcal, A.R.S., J.S. Borges, J.A. Gomes, and J.F. Pinto da Costa, 2005. Land cover update by supervised classification of segmented ASTER images, *International Journal of Remote Sensing*, 26:1347–1362.

Mertens, K.C., L.P.C. Verbeke, T. Westra, and R.R. De Wulf, 2004. Sub-pixel mapping and sub-pixel sharpening using neural network predicted wavelet coefficients, *Remote Sensing of Environment*, 91:225–236.

Pajares, G., and J.M. Cruz, 2004. A wavelet-based image fusion tutorial, *Pattern Recognition*, 37:1855–1872.

Paola, J.D., and R.A. Schowengerdt, 1995. A review and analysis of backpropagation neural networks for classification of remotely sensed multispectral imagery, *International Journal of Remote Sensing*, 16:3033–3058.

Ranchin, T., and L. Wald, 2000. Fusion of high spatial and spectral resolution images: The ARSIS concept and its implementation, *Photogrammetric Engineering & Remote Sensing*, 66(1):49–61.

Wilkinson, G.G., 2005. Results and implications of a study of fifteen years of satellite image classification experiments, *IEEE Transactions on Geoscience and Remote Sensing*, 43:433–440.

Yamaguchi, Y., A.B. Kahle, H. Tsu, T. Kawakami, and M. Pniel, 1998. Overview of Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER), *IEEE Transactions on Geoscience and Remote Sensing*, 36:1062–1071.

Zhu, G., and D.G. Blumberg, 2002. Classification using ASTER data and SVM algorithms; The case study of Beer Sheva, Israel, *Remote Sensing of Environment*, 80:233–240.

(Received 10 May 2006; accepted 10 July 2006; revised 18 August 2006)