CLASSIFICATION OF HYPERSPECTRAL IMAGES USING WAVELET NETWORKS

Pai-Hui Hsu, Associate Researcher
National Science & Technology Center for Disaster Reduction
9F., No.200, Sec. 3, Beisin Rd., Sindian City
Taipei County 231, TAIWAN
hsuph@ntu.edu.tw

ABSTRACT

In our previous works, the wavelet-based feature extraction algorithms have been developed to explore the useful information for the hyperspectral image classification. On the other hand, the idea of using artificial neural network (ANNs) has also proved useful for hyperspectral image classification. To combine the advantages of ANNs with wavelet-based feature extraction methods, the wavelet network (WN) has been proposed for data identification and classification. The value of wavelet networks lies in their capabilities of extracting essential features in time-frequency plane. Furthermore, both the position and the dilation of the wavelets are optimized besides the weights of the network during the training phase. In this paper, the basic concept of wavelet-based feature extraction is firstly described. Then the theory of wavelet networks is introduced for the hyperspectral image classification. Finally an AVIRIS image was used to test the feasibility and performance of classification using the wavelet networks. The experiment results showed that the wavelet networks exactly an effective tool for classification of hyperspectral images, and have better classification results than the traditional feed-forward multi-layer neural networks.

INTRODUCTION

Multispectral sensors have been developed and widely used for the observation of the earth surface since 1960’s. However, due to the limitation of the sensor design, sensors of this kind are composed of discrete detectors or line arrays of detector elements which can only collect spectral data less than 10 bands. In the mid 1980s, the new technology of imaging spectrometer with two-dimensional area arrays was developed to collect spectral data with a large number of bands simultaneously (Goetz et al., 1985; Vane et al., 1993). Because of the large number of spectral bands, the images acquired with imaging spectrometers are also referred to as hyperspectral images which are distinguished from the multispectral images with only three to ten bands. The detailed and subtle spectral signatures provided by the hyperspectral images can provide sufficient information to identify and quantify a large range of surface materials which cannot be identified by broadband or low spectral resolution imaging systems, such as the Landsat TM and SPOT systems. By means of the characteristics of the solar reflected spectrum measured by imaging spectrometers, a wide range of scientific researches and applications have being proposed based on the spectral analysis (Lillesand and Kiffer, 2000).

The high dimensionality of hyperspectral data should seemingly increase the abilities and effectiveness in classifying land use/cover types. However, the classification methods that have been successfully applied to multispectral data in the past are not as effective as to hyperspectral data. The major cause is that the size of training data set does not adapt to the increasing dimensionality of hyperspectral data. If the training samples are insufficient for the needs, which is a very common case in the hyperspectral images, the estimation of statistical parameters becomes inaccurate and unreliable. As the dimensionality increases with the number of bands, the number of training samples needed for training a specific classifier should be increased exponentially as well. The rapid increase in training samples size for density estimation has been termed the “curse of dimensionality” by Bellman (1961), which leads to the “peaking phenomenon” or “Hughes phenomenon” in classifier design (Hughes, 1968). The consequence is that the classification accuracy first grows and then declines as the number of spectral bands increases while training samples are kept the same.

For a given classifier, the “curse of dimensionality” can only be avoided by providing a sufficiently large sample size. The more complex the classifier, the larger should the ratio of sample size to dimensionality be to avoid the curse of dimensionality. However, in practice, the number of training samples is limited in most of the hyperspectral applications. Furthermore, the high dimensionality of hyperspectral data makes it necessary to seek new analytic methods to avoid a vast increase in the computational time. A simpler, but sometimes very effective way of dealing with high-dimensional data is to reduce the number of dimensions (Lee and Landgrebe, 1993; Benediktsson et al., 2000).
The dimensionality reduction can be implemented by feature extraction that a small number of salient features are extracted from the hyperspectral data when confronted with a limited set of training samples. Since the dimensionality is reduced before classification, the curse of dimensionality can be avoided. Thus most of the traditional classification methods such as the Maximum Likelihood Classifier (MLC) can be directly applied to the extracted features after the feature extraction.

Feature extraction is generally considered a data mapping procedure which determines an appropriate subspace of dimensionality $M$ from the original feature space of dimensionality $N$ ($M : N$) (Fukunaga, 1990; Jain et al., 2000). The way of feature extraction can be a linear or nonlinear data transformation. Regardless of how the data transformation is implemented, the feature extraction algorithm must be designed to preserve the information of interest for a special problem such as compression, denoising, or classification. For example, in hyperspectral image classification, effective features are those which are most capable of preserving class separability. Several different approaches proposed for feature extraction, such as Principle Component Analysis (PCA), Discriminant Analysis Feature Extraction (DAFE) using the linear discriminant analysis, and Decision Boundary Feature Extraction (DBFE), were experimentally compared in Hsu (2003). Although most of these methods provide effective and practical algorithm for deriving effective features in many circumstances, there are several drawbacks in these methods (Hsu, 2003; Tadjudin and Landgrebe, 1998). Since these methods depend on how well the training samples approximate the classification separability or decision boundaries, the number of training samples required could be much more for high dimensional data because it computes the class statistical parameters at full dimensionality. For hyperspectral images, the number of training samples is usually not enough to prevent singularity or to yield a good covariance estimate.

In order to avoid the problems caused by the limited training samples, several feature extraction methods based on the wavelet transform have been proposed for hyperspectral images (Hsu and Tseng, 2000; Hsu, 2003). In the past two decades, wavelet transform (WT) has been developed as a powerful analysis tool for signal processing, and also has been successfully applied in applications such as image processing, data compression and pattern recognition (Mallat, 1999). Due to the time-frequency localization properties, discrete wavelet and wavelet packet transforms have proven to be appropriate starting point for the classification of the measured signals (Pittner and Kamarthi, 1999). The WT decomposes a signal into a series of shifted and scaled versions of the mother wavelet function. The local energy variation of a hyperspectral signal in different spectral bands at each scale (or frequency) can be detected automatically and provide useful information for hyperspectral image classification.

In addition to the dimensionality reduction for the statistics-based classifier, nonparametric classifiers such as the artificial neural networks (ANNs) are also proposed to avoid the “curse of dimensionality” for hyperspectral images and have been applied successfully in various applications. The use of neural networks is motivated by their power in pattern recognition and classification due to the ultimately fine distribution and non-linearity of the process. Gong et al. (1997) had made an assessment of the capability of a neural network algorithm for conifer species identification. Their experiment results showed that the neural networks are sensitive to subtle spectral details and can be trained to separate samples from the same species at different sites. Plaza et al. (2004) combined the Hopfield neural network (HNN) and multi-layer perceptron (MLP) for estimating the abundance of endmember materials in hyperspectral images. The proposed neural model integrates the concepts of linear and nonlinear unmixing. Aguilar et al. (2000) developed a self-organizing neural network for the abundance extraction from the AVIRIS images.

Neural networks are distribution free since no prior knowledge of the statistics of the data is needed (Benediktsson et al., 1990). In the neural network approach to pattern recognition, the network behavior like a black box which is trained iteratively by presenting the training data and the class numbers at the inputs. Most of the neural processing algorithms are computationally intensive and involve many iterative calculations, especially for hyperspectral images. A characteristic of neural networks is that the networks need a long training time but are relatively fast data classifiers. For very-high-dimensional data, the training time of a neural network can be very long and the resulting neural network can be very complex. This is a serious drawback, especially when the dimensionality and the sample size of training data are large (Benediktsson et al., 1995).

Although the wavelet-based feature extraction methods perform well for dimensionality reduction and also effectively for classification, however, the relationship between the extracted features and the identified classes are not apparent. On the other hand, the advent of ANNs approaches in hyperspectral analysis is mainly due to their power in pattern recognition and classification. In order to combine the advantages of ANNs with wavelet-based feature extraction methods, the wavelet network (WN) has been proposed for data identification and classification (Dickhaus and Heinrich, 1996). The value of wavelet networks lies in their capabilities of extracting essential features in time-frequency plane. Furthermore, both the position and the dilation of the wavelets are optimized besides the weights of the network during the training phase. This hierarchical, multiresolution training can result in a more meaningful interpretation of the resulting mapping and adaptation of networks that are more efficient compared to
conventional methods. In addition, the wavelet theory provides useful guidelines for the construction and initialization of networks and, consequently, the training times are significantly reduced (Iyengar, 2002).

In this paper, the basic concept of wavelet-based feature extraction is firstly described. Then the theory of wavelet networks is introduced for the hyperspectral image classification. Finally an AVIRIS image was used to test the feasibility and performance of classification using the wavelet networks. The experiment results showed that the wavelet networks exactly an effective tool for classification of hyperspectral images, and better than the traditional feed-forward multi-layer neural networks.

WAVELET-BASED FEATURE EXTRACTION

The general process of the wavelet-based feature extraction methods is illustrated in Figure 1 (Hsu, 2003). Firstly, wavelet or wavelet packet transforms are implemented on the hyperspectral images and a sequence of wavelet coefficients is produced. Then, a simple feature selection procedure associated with a criterion is used to select the effective features for classification. The criterion of feature selection can be designed for signal representation or classification. In the stage of feature selection shown in Figure 1, some training data may be needed as samples to find the effective features for classification. Unlike the existing feature extraction methods such as DAFE or DBFE which need to estimate the statistic parameters at full dimensionality, the wavelet-based feature extraction optimizes the criterion in a lower dimensional space. Thus the problem caused by limited training samples can be avoided.

Wavelet and Wavelet Packets Decomposition

The orthogonal wavelet transform in terms of multi-resolution analysis (MRA) can decompose a signal into the low-frequency components that represent the optimal approximation, and the high-frequency components that represent the detailed information of the original signal (Mallat, 1989). The decomposition coefficients in a wavelet orthogonal basis can be computed with a fast algorithm that cascades discrete convolutions with conjugate mirror filters h and g, and subsamples the outputs. The decomposition formulas are described as following:

\begin{align*}
    a_{j,0}[p] &= \sum_{n=-\infty}^{\infty} h[n - 2p]a_j[n] = a_j * \tilde{h}[2p] \\
    d_{j,0}[p] &= \sum_{n=-\infty}^{\infty} g[n - 2p]a_j[n] = a_j * \tilde{g}[2p]
\end{align*}

where \( \tilde{h}[n] = h[-n] \) and \( \tilde{g}[n] = g[-n] \). \( a_j \) is the approximation coefficients at scale \( 2^j \), and \( a_{j+1} \) and \( d_{j+1} \) are respectively the approximation and detail components at scale \( 2^{j+1} \).

In the orthogonal wavelet decomposition algorithm, only the approximation coefficients are split iteratively into a vector of approximation coefficients and a vector of detail coefficients at a coarser scale. In other words, the successive details are never reanalyzed. The concept of wavelet packets which was introduced by Coifman et al. (1992) generalized the link between multi-resolution approximations and wavelets details (Mallat, 1999). In wavelet packets, each detail coefficients vector will be decomposed into two parts using the same approach as in approximation vector splitting. The consequently recursive splitting will produce a binary tree where each parent node is divided in two orthogonal subspaces and offer more detailed analysis of a hyperspectral images (Hsu and Tseng, 2002).

Wavelet Based Feature Extraction

Based on the characteristics of wavelet decomposition, the wavelet-based feature extraction can be roughly divided into linear and non-linear methods. The basic idea of linear wavelet approximation method is similar to the concept of the PCT and Fourier-based approximation. They are all based on the same criterion that the best approximation with the minimum error is used as a set of important features. Therefore the features extracted by the linear methods almost have the identical effectiveness. On the other hand, features extracted by the nonlinear wavelet-based methods are selected from not only the approximation but also the detail information with large wavelet coefficients. The non-linear features calculated from the M largest amplitude wavelet coefficients can be interpreted as an adaptive grid approximation, where the approximation scale is refined in the neighborhood of singularities (Mallat,
The experiment results showed that the nonlinear wavelet-based methods are more effective for classification than linear methods (Hsu and Tseng, 2002).

On the other hand, for a given orthogonal wavelet function, one may generate a wavelet packets with a large family of orthogonal bases that include different types of time-frequency atoms. Each basis in the library offers a particular way of coding signals, preserving global energy, and reconstructing exact features. For discrete signals of size \( N \), the number of wavelet packet bases is more than \( 2^{N/2} \) (Mallat, 1999). To optimize the non-linear signal approximation, one may adaptively choose the “best basis” depending on the signal from this large family of bases. This can be done by the Best-Basis Algorithm (BB), which minimizes a concave cost function \( C \) (e.g. entropy or norm) to select the most suitable decomposition of a given spectrum, or the Local Discriminant Bases (LDB) Algorithm, which maximizes the discriminating function \( D \) between the nodes of the tree to search for a best basis for classification (Hsu, 2003).

Although some wavelet-based methods such as the nonlinear WFE and best basis algorithm are based on the best approximation for data representation, they are still effective for classification. In some circumstances, the LDB methods which take into account the discriminant information from the training data have better results. In the LDB methods, the best basis is selected within the subspace of wavelet packets, thus the problem of limited training sample size is avoided (Hsu, 2003).

![Figure 1. The general flow chart of wavelet-based feature extraction.](image-url)
WAVELET NETWORKS

Wavelets networks (WN) were first proposed by Zhang and Benveniste (1992) in the context of non-parametric regression of functions in $L^2(\mathbb{R}^2)$. The basic idea of wavelet networks is to combine the localization property of wavelet decomposition and the optimization property of neural networks learning. Zhang and Benveniste (1992) have found a link between the wavelet decomposition and neural networks, and present a basic backpropagation wavelet network learning algorithm. Their wavelet networks preserve the universal approximation properties of the traditional feed-forward neural networks and present an explicit link between the network coefficients and some appropriate transform. Based on the wavelet transform theory, the wavelet networks were proposed as an alternative to feed-forward neural networks for approximating arbitrary nonlinear functions.

For the classification applications, the wavelet networks can also be derived from the background of the general pattern-recognition model, with its successive steps of feature extraction, selection, and classification (Dickhaus and Heinrich, 1996). The basic concept of wavelet networks (WN) for classification tries to combine aspects of the wavelet transformation for purposes of feature extraction and selection with the characteristics decision capabilities of neural-network training. A wavelet network for classification consists of two parts shown in figure 2: wavelet based feature extraction and neural networks classification. A wavelet node of the feature extraction part is parameterized by a shift parameter, $\tau_k$, and a scale parameter, $a_k$. The outputs of the wavelet nodes, $q_k$, which can be interpreted as the correlation between the signal $x_i(t)$ and the modified wavelet $h_k(t)$, serve as input to the neural network classifier. The classifier of the right part can be any single-layer or multi-layer perceptrons. During the learning process, the wavelet node parameters are also updated to minimize the error, $E$.

\[ h(t) = \exp\left(j \omega_0 t - 0.5 \cdot t^2\right) \]

The wavelet nodes $h_k(t)$ in figure 2 are scaled and dilated versions of this wavelet mother function.

![Figure 2. The general diagram of wavelet networks](image-url)

Wavelet Transform

A typical wavelet function used in the wavelet networks is the complex Morlet wavelet (Dickhaus and Heinrich, 1996):
\[ h_t(t) = \frac{1}{a} \exp \left[ j \omega_0 \frac{t - \tau_k}{a} - 0.5 \left( \frac{t - \tau_k}{a} \right)^2 \right] \] \hspace{1cm} (4)

The variable \( a \) is the scale parameter, and \( \tau \) is the dilation parameter of the wavelet function. If the scale \( a \) is large, the wavelet is a dilated low-frequency function, whereas for small values of \( a \), the wavelet is compact, corresponding to a high-frequency function. Formally, the node’s output, \( \varphi_{ik} \), is the result of the wavelet transform which is defined as the inner product of the node \( h_k \) and the signal \( x_i \), which is the input of the wavelet networks (the index \( i = 1, \ldots, N \) denotes the signal number).

\[ \varphi_{ik} = \langle h_k, x_i \rangle = \int h \left( \frac{t - \tau_k}{a_k} \right)x_i(t)dt \] \hspace{1cm} (5)

For the Morlet wavelet, \( \varphi_{ik} \) can be calculated for each wavelet node:

\[ \varphi_{ik} = \sqrt{o_{\cos ik}^2 + o_{\sin ik}^2} \] \hspace{1cm} (6)

\[ o_{\cos ik} = \int x_i(t)dt \cos \left( \omega_k \frac{t - \tau_k}{a_k} \right) \exp \left\{ -0.5 \left( \frac{t - \tau_k}{a_k} \right)^2 \right\} dt \] \hspace{1cm} (7)

\[ o_{\sin ik} = \int x_i(t)dt \sin \left( \omega_k \frac{t - \tau_k}{a_k} \right) \exp \left\{ -0.5 \left( \frac{t - \tau_k}{a_k} \right)^2 \right\} dt \] \hspace{1cm} (8)

**Training of the Neural Networks**

In figure 2, the neuron’s output, \( o_j \), is defined by the weighted sum of the outputs of the previous layer, \( o_k \), the neuron’s threshold, \( \theta_j \), and its activation function \( f \), that the sigmoidal function is used in this paper:

\[ o_j = f(\sum_k w_{kj} o_k + \theta_j) = \frac{1}{1 + \exp \left[ -\left( \sum_k w_{kj} o_k + \theta_j \right) \right]} \] \hspace{1cm} (9)

During the training phase, the ANN weights \( w_{ij} \) are adjusted to minimize the least-square error:

\[ E = \sum_{i=1}^N \sum_{j=1}^N (d_{ij} - y_{ij})^2 = \min \] \hspace{1cm} (10)

between the net’s desired output vector \( \mathbf{d} \) and its actual output \( \mathbf{y} \) for all input vectors \( \mathbf{x} \). The minimization problem can be solved by an iterative gradient technique. The partial derivative of the weights, \( w_{jk} \), in the classification part are calculated according to the generalized delta rule.
The partial derivatives of the weights to the neurons in the hidden layer of a two-layer perceptron are calculated as follows:

$$ \frac{\partial E}{\partial w_{hid ij}} = - \sum_{i=1}^{N} \delta_{hid} o_{hid i} = - \sum_{i=1}^{N} (d_{y} - y_{y}) y_{y} (1 - y_{y}) o_{hid i} $$ (11)

These two equations hold for neurons with a sigmoidal activation function.

### Training of the Wavelet Nodes

In the wavelet network, not only the weights are adjusted, but also the parameters of the wavelet nodes. The partial derivatives for a wavelet node’s scale parameter, $a_{k}$, and it shift parameter, $t_{k}$, depend on the wavelet basis chosen and are determined using the backpropagated error $E$:

$$ \frac{\partial E}{\partial \tau_{k}} = - \sum_{i=1}^{N} \sum_{j=1}^{H_{M}} \delta_{y} w_{y j} \frac{\partial \phi_{k}}{\partial \tau_{k}} $$

$$ \frac{\partial E}{\partial a_{k}} = - \sum_{i=1}^{N} \sum_{j=1}^{H_{M}} \delta_{y} w_{y j} \frac{\partial \phi_{k}}{\partial a_{k}} $$

Thus, in each iteration of the training cycle, the weights and the wavelet parameters are varied to reduce the error, $E$. This procedure is repeated until the net has settled down to a minimum.

$$ \frac{\partial E}{\partial \tau_{k}} = \sum_{i=1}^{N} \sum_{j=1}^{H_{M}} \delta_{y} w_{y j} \frac{1}{\phi_{k}} \sum_{i=1}^{I} \left( t_{i} \right) \exp \left( -0.5 \left( \frac{t_{i} - \tau_{k}}{a_{k}} \right)^{2} \right) \frac{1}{a_{k}} 

\left[ o_{\text{cosh}} \cdot \left( a_{k} \sinh \left( \frac{t_{i} - \tau_{k}}{a_{k}} \right) + \frac{t_{i} - \tau_{k}}{a_{k}} \cos \left( \frac{t_{i} - \tau_{k}}{a_{k}} \right) \right) \right] 

+ \left[ o_{\text{sinh}} \cdot \left( \frac{t_{i} - \tau_{k}}{a_{k}} \right) \frac{t_{i} - \tau_{k}}{a_{k}} \cos \left( \frac{t_{i} - \tau_{k}}{a_{k}} \right) \right] $$

$$ \frac{\partial E}{\partial a_{k}} = \frac{t_{i} - \tau_{k}}{a_{k}} \frac{\partial E}{\partial \tau_{k}} $$

$$ \frac{\partial E}{\partial \omega_{k}} = - \sum_{i=1}^{N} \sum_{j=1}^{H_{M}} \delta_{y} w_{y j} \frac{1}{\phi_{k}} \sum_{i=1}^{I} \left( t_{i} \right) \exp \left( -0.5 \left( \frac{t_{i} - b_{i}}{a_{k}} \right)^{2} \right) \frac{t_{i} - b_{i}}{a_{k}} 

\left[ - o_{\text{cosh}} \cdot \sin \left( \frac{t_{i} - b_{i}}{a_{k}} \right) + o_{\text{sinh}} \cdot \cos \left( \frac{t_{i} - b_{i}}{a_{k}} \right) \right] $$

ASPRS 2006 Annual Conference
Reno, Nevada • May 1-5, 2006
EXPERIMENTS

The purpose of this experiment is to test the feasibility and performance of wavelet networks for hyperspectral image classification. The test data sets (see figure 3) is an AVIRIS image derived from Purdue University. The image was taken over an agricultural portion on NW Indiana in 1992. The original data set has 224 spectral bands from 400nm to 2450nm with 10 nm spectral resolution. The number of bands is 220 after removing 4 noisy bands. The ground truth data shown in Figure 3(b) includes four different classes which are corn, grass, soybean-1 and soybean-2. The number of pixels of each known class is 400. In the wavelet networks, 100 samples of each class are randomly selected from the ground truth data for the network training. The others are used to assess the accuracy of the classification.

![Test images](image1.png) ![Ground truth data](image2.png)

**Figure 3.** The test data delivered by an AVIRIS image on NW Indiana in 1992

In order to demonstrate the effectiveness of the wavelet networks, a traditional feed-forward multi-layer neural network with one hidden layer was simultaneously used to classify the same data set. Table 1 lists the differences of the training parameters of these two networks. The number of the neurons in the hidden layer of the feed-forward network is 100. On the contrary, the number of the neurons in the hidden layer of the wavelet network was reduced to 10 because of the effect from the wavelet feature extraction which is implemented by 20 wavelet nodes before the hidden layer. The reduced number of neurons will reduce the network complexity. Furthermore, in the process of this experiment, we found that the feed-forward network is very time consuming when using 100 spectral samples with 220 bands for batch training. In order to reduce the computation time, only the average of the spectrum is used to train the feed-forward network. However, the network becomes unstable because of small number of inputs. Therefore, the spectral variance information of the 100 samples is especially added to generalize the feed-forward network. The experiment results showed that the classification accuracy using the wavelet network is 85.94% which is better than the accuracy of 44.44% using the feed-forward network. Because the spectral variance information is not used in the wavelet network, more iterations of training is required.

Table 2 and table 3 respectively show the accuracy assessment of these two classifications methods. The producer’s accuracy (error of omission) of class corn is increased when using the wavelet networks.

Table 1. Classification results using neural network and wavelet network

<table>
<thead>
<tr>
<th></th>
<th>Classification Result Using Neural Network</th>
<th>Classification Result Using Wavelet Network</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of Layers</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>Number of wavelet nodes</td>
<td>NA</td>
<td>10</td>
</tr>
<tr>
<td>Number of nodes of hidden layer</td>
<td>100</td>
<td>20</td>
</tr>
<tr>
<td>Number of Iterations (Using Means /Using means and Covariances)</td>
<td>157/243</td>
<td>1021/NA</td>
</tr>
<tr>
<td>Total Accuracy</td>
<td>44.44%</td>
<td>85.94%</td>
</tr>
</tbody>
</table>
Table 2. Accuracy assessment using the feedforward neural networks

<table>
<thead>
<tr>
<th>class</th>
<th>Corn</th>
<th>Grass</th>
<th>Soybean-1</th>
<th>Soybean-2</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Corn-notill</td>
<td>23(6%)</td>
<td>11</td>
<td>32</td>
<td>6</td>
<td>72</td>
</tr>
<tr>
<td>Grass</td>
<td>6</td>
<td>277(69%)</td>
<td>3</td>
<td>7</td>
<td>293</td>
</tr>
<tr>
<td>Soybean-notill</td>
<td>186</td>
<td>53</td>
<td>269(67%)</td>
<td>245</td>
<td>753</td>
</tr>
<tr>
<td>Soybean-min</td>
<td>185</td>
<td>59</td>
<td>96</td>
<td>142(36%)</td>
<td>482</td>
</tr>
<tr>
<td>Total</td>
<td>400</td>
<td>400</td>
<td>400</td>
<td>400</td>
<td>1600</td>
</tr>
</tbody>
</table>

Table 3. Accuracy assessment using the wavelet networks

<table>
<thead>
<tr>
<th>class</th>
<th>Corn</th>
<th>Grass</th>
<th>Soybean-1</th>
<th>Soybean-2</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Corn-notill</td>
<td>297(74%)</td>
<td>4</td>
<td>10</td>
<td>46</td>
<td>357</td>
</tr>
<tr>
<td>Grass</td>
<td>0</td>
<td>396(99%)</td>
<td>0</td>
<td>1</td>
<td>392</td>
</tr>
<tr>
<td>Soybean-notill</td>
<td>1</td>
<td>0</td>
<td>335(84%)</td>
<td>6</td>
<td>342</td>
</tr>
<tr>
<td>Soybean-min</td>
<td>102</td>
<td>0</td>
<td>55</td>
<td>347(87%)</td>
<td>504</td>
</tr>
<tr>
<td>Total</td>
<td>400</td>
<td>400</td>
<td>400</td>
<td>400</td>
<td>1600</td>
</tr>
</tbody>
</table>

CONCLUSIONS

In this paper, the wavelet network algorithm is used for the classification of hyperspectral images. In the wavelet networks, the task of feature extraction is performed with the wavelet decomposition, whereas the classification is carried out by the multi-layer neural networks. The advantages of wavelet networks include optimally adapted features as network inputs, improvement of the classification accuracy, and reduction of the calculated time. The experiment results showed that the wavelet networks exactly an effective tool for classification of hyperspectral images.

The experiment showed that large number of iterations in the training phase of wavelet networks is needed for classification. In the future, the training algorithm will be improved to reduce the number of iterations. Furthermore, because the results of classification using wavelet networks are strongly depend on the choice of wavelet basis, the classification accuracies of wavelet network using different wavelets function will be tested in the future.
REFERENCES


