

Neuro-fuzzy Based Analysis of Hyperspectral Imagery

Fang Qiu

Abstract

A neuro-fuzzy system, namely Gaussian Fuzzy Learning Vector Quantization (GFLVQ), was developed based on the synergy of a neural network and a fuzzy system. GFLVQ is both a fuzzy neural network and a neural fuzzy system with supervised learning and unsupervised self-organizing capabilities. In this paper, GFLVQ was further improved to efficiently and effectively process hyperspectral data through training data informed initialization and a simplified fuzzy learning algorithm. A geovisualization tool was developed to facilitate knowledge discovery and understanding of the hyperspectral image. A case study was conducted using a Hyperion image. The results obtained from the improved neuro-fuzzy system were found to be significantly better than those from conventional statistics-based and endmember-based classifiers. The fuzzy spectral profiles produced from the geovisualization tool provided an extra insight into the neuro-fuzzy learning process, further opening up the black box of the neural network.

Introduction

The advent of hyperspectral data (e.g., airborne AVIRIS and spaceborne Hyperion imagery) provides hundreds of relatively narrow (≤ 10 nm), contiguous bands that may be useful for extracting land-use information and identifying the constituents of ground materials in the image pixel. However, significant collinearity introduced by the huge number of bands vastly complicates multivariate statistics based image processing (Benediktsson *et al.*, 1994). The existence of singular matrices resulted from relatively smaller training samples compared to the high dimensionality of the image data additionally pushes the limits of the established multispectral classification methods such as maximum likelihood classifier (Benediktsson *et al.*, 1994). Moreover, the performance of these classifiers relies on the assumption that all classes should generally follow a multi-band normal distribution, which is extremely difficult to achieve with a myriad of channels in hyperspectral data (Benz, 1999), especially with the prevalent existence of mixed pixels that contain multiple classes within each pixel (Bastin *et al.*, 2002).

Consequently, hyperspectral image classifiers that are free of distribution assumptions, such as spectral angle mapping, linear spectral unmixing, matched filtering, and spectral feature fitting have been proposed. Most of these classifiers operate on the basis of reference endmembers, which are pure pixels representing spectral extremes and corresponding to homogeneous regions comprised of a single class in the image. However, an accurate identification of absolute endmembers from hyperspectral images is

extremely difficult if not impossible because the majority of image pixels may be mixed (Foody *et al.*, 1995). Alternatively, spectra from a spectral library or those derived from extensive field work can be used if an absolute radiometric calibration can be performed on the image. A perfect radiometric calibration is unlikely to be achieved without *a priori* knowledge of atmospheric and directional effects and pre-launch calibration coefficients of the sensor systems (Wooster *et al.*, 1995).

For these reasons, researchers have turned to artificial intelligence (AI) approaches, which have been previously applied to multispectral data (Jensen, 2005). Among various AI approaches, fuzzy logic systems and neural networks are the two most widely used techniques with the potential to handle a large volume of information. A fuzzy logic system (or simply fuzzy system) is an expert system that can represent image classification decisions explicitly in the form of declarative fuzzy “if-then” rules (Binaghi *et al.*, 1997). As an extension of classic expert systems, fuzzy systems often extract rules from knowledgeable experts and use them to simulate human reasoning in an algorithmically structured manner to answer questions about a restricted domain application. These rules utilize the concept of a fuzzy set (Zadeh, 1965), a generalization of the classic Boolean set. Fuzzy sets deliberately make use of vague, imprecise, or uncertain information to generate simpler models that are more familiar to human thinking (Klir and Yuan, 1995). Fuzzy sets allow partial membership assigned to a class and have also been used to model land-cover constituents of a mixed pixel, which records a mixture of radiant flux from more than one ground classes, by using fuzzy classifiers such as unsupervised fuzzy c-mean (FCM) (Bezdek, 1981), supervised fuzzy c-mean (Wang, 1990; Zhang and Foody, 2001), and fuzzy k-mean (Tapia *et al.*, 2005) classifiers.

However, not all fuzzy classifiers are purely fuzzy logic based. Some are in fact “soft” classifiers that soften the “crisp” output of conventional “hard” classifiers to derive a fuzzy land-cover representation (Woodcock *et al.*, 2000; Zhang and Foody, 2001), for instance, from the *a posteriori* probabilities of a maximum likelihood classification (Foody *et al.*, 1992). Most fuzzy classifiers utilize fuzzy set concepts and/or fuzzy logic operations in certain stages of the classification, but they are not necessarily fuzzy expert systems in a strict sense, except the work of Penaloza and Welch (1996) and Binaghi *et al.* (1997). This is primarily due to the limitations of fuzzy systems in acquiring expert knowledge in the form of symbolic fuzzy “if-then” rules. Currently, most fuzzy systems solely rely on the

Photogrammetric Engineering & Remote Sensing
Vol. 74, No. 10, October 2008, pp. 1235–1247.

0099-1112/08/7410-1235/\$3.00/0

© 2008 American Society for Photogrammetry
and Remote Sensing

University of Texas at Dallas, Richardson, TX, 75083
(ffqiu@utdallas.edu).

encapsulation of human expertise to derive fuzzy “if-then” rules and to specify fuzzy set parameters. The lack of a mechanism to learn from examples makes the acquisition of knowledge a tedious and subjective process (Openshaw and Openshaw, 1997). In complex fuzzy systems, manual determination and optimization of fuzzy membership parameters is virtually impossible (Abe and Lan, 1996). It is desirable that knowledge automation capabilities be incorporated into existing fuzzy systems so that the benefits promised by fuzzy set theory can be made available to the image classification task.

Neural networks have been successfully employed to process multispectral and hyperspectral remote sensing images (Civco, 1993; Benediktsson *et al.*, 1994; Foody *et al.*, 1995; Gong, 1996; Filippi and Jensen, 2006). These successes are underpinned by many salient characteristics of neural networks. A single neuron simulates the computation of a multivariate linear regression model (Hewitson and Crane, 1994; Jensen *et al.*, 1999), making no *a priori* assumptions of normal and linear data distribution, due to its operation in a non-parametric fashion (Haykin, 1994; Foody *et al.*, 1995). Neural networks are able to adaptively learn from existing examples, making the classification process objective (Hagan *et al.*, 1996). Unlike statistical classifiers that require a sample size that is at least larger than the number of bands for each class, neural networks can learn effectively from small training samples and are not limited by singular covariance matrices (Benediktsson *et al.*, 1994). Additionally, neural networks are often bestowed with a good generalization ability that can provide robust prediction for previously unseen samples, incomplete or imprecise patterns (Fausett, 1994).

Despite the excellent performance in image classification, a neural network is often accused of being a black box, which hides the relation between inputs and outputs in the weights of the neurons behind its *hidden* layers (Benitez *et al.*, 1997). It is difficult to interpret these weights due to their complex nature. As a result, we cannot gain any understanding of the classification process due to the lack of an explanatory capability to provide insight into classification. For the same reason, it is also impossible to incorporate human expertise to simplify, accelerate, or improve the performance of image classification. A neural network always has to learn from scratch (Nauck *et al.*, 1997). For neural networks to be widely applied in complex remote sensing image classification tasks, an explanatory capability should be incorporated as part of a trained neural network.

Attempting to bring “human intelligence” into data processing tasks, these two technologies approach their goal from different perspectives, and both exhibit unique features and fundamental limitations as standalone systems. Each technology has the capability to solve one aspect of the problem, but neither by itself can provide a total solution. The examination of the assets and liabilities of neural networks and fuzzy systems reveals that the two technologies are functionally complementary with each other. The benefits possessed by one technology happen to be shortcomings of the other (Kulkarni and Lulla, 1999). If these two approaches are combined, one technology can provide the capabilities not available in the other.

The integration of neural networks and fuzzy systems was therefore suggested and is often referred to as neuro-fuzzy systems. The synergies of these two technologies can take different approaches, either as neural fuzzy systems, or as fuzzy neural networks, or as both. A neural fuzzy system is a fuzzy expert system that uses a learning algorithm derived from or inspired by neural network theory to determine its fuzzy membership parameters based on sample data. A fuzzy neural network is a fuzzified neural network

that uses fuzzy logic to adapt the weights of neurons during the learning process (Joshi *et al.*, 1996; Nauck *et al.*, 1997). In many cases, these two terms are often used interchangeably. In this paper, we use the term neuro-fuzzy system when distinguishing between the two is not needed. A neuro-fuzzy system is believed to deliver a more powerful solution than its individual components. A variety of neuro-fuzzy systems have been proposed, and some of the early systems were reviewed by Buckley and Hayashi (1994). Considerable efforts were made to synthesize the recent achievements of the two technologies since then, and the field remains an active research area with significant applications to remote sensing image classification.

In this paper, a Gaussian Fuzzy Learning Vector Quantization (GFLVQ) classifier is proposed to analyze hyperspectral remotely sensed images based on modifying and extending the neuro-fuzzy system that we previously designed for multispectral image classification (Qiu and Jensen, 2004). While retaining the strengths of the original system, the improved GFLVQ now delivers better efficiency and effectiveness in classifying hyperspectral images. Modifications to the original systems were made to specifically speed up the learning process in order to efficiently process large volumes of hyperspectral data, first to the system initialization, and second to the weight updating of the network. In order to further open the black box of the neural network and gain understanding of its learning and classifying process, a geovisualization tool was developed to produce fuzzy spectral profiles. This is in addition to the symbolic fuzzy “if-then” rules derived from the original system. The next section reviews recent developments in neuro-fuzzy systems and their application to remote sensing image classification. The methodology section explains the original GFLVQ system and the improvements to the system. A case study is then conducted by applying GFLVQ to a hyperspectral image. The results are compared to those from a statistics based maximum likelihood classifier (MLC) and several endmember-based hyperspectral classifiers, such as spectral angle mapping, linear spectral unmixing, matched filter, and spectral feature fitting. The final section addresses conclusions and future studies.

Neuro-fuzzy Systems

In the field of neuro-fuzzy systems, there are several different approaches that attempt to integrate neural networks and fuzzy systems. These neuro-fuzzy systems can be generally classified into three major categories according to their underlying neural networks, including the most widely used multi-layer perceptron (MLP), adaptive resonance theory (ART), and self-organizing map (SOM) and learning vector quantization (LVQ) networks.

MLP-based

MLP with error back propagation (BP) learning algorithm is by far the most frequently used neural network for classifying remotely sensed imagery. MLP-based neuro-fuzzy systems are based on the interpretation of weights and activation functions of the neurons, usually in a specially configured topological network to extract fuzzy “if-then” rules by introducing new fuzzy logic operators (Benitez *et al.*, 1997). MLP neural networks are also used to derive and refine the fuzzy set parameters of an inference system (Mao *et al.*, 2002; Qu *et al.*, 2003). The development of MLP neuro-fuzzy systems is straightforward and can be readily implemented by employing an existing MLP neural network with a specified topological structure and activation functions. However, there are several fundamental difficulties with MLP neuro-fuzzy systems. Like many error-based neural networks, BP

learning often suffers a “stability-plasticity” dilemma, concerning a network to remain plastic enough to learn while remaining stable enough to recall stored patterns (Simpson, 1992). Whenever a new pattern arrives in a MLP network, it is used to adapt the weights of all neurons, causing successive training of a network to interfere with previously acquired knowledge. Thus, MLP requires a complete retraining of the whole network, which may modify or even erase previous learning (Gopal *et al.*, 1999; Liu *et al.*, 2004). To avoid this side effect of catastrophic forgetting of past memories, computational expensive relearning is often required, placing huge demands on the processor and memory, and leading to increasingly longer training times for even small size problems (Simpson, 1992; Carpenter *et al.*, 1997). Additionally, rare events may be averaged with other patterns and lose their identity, or the networks may become trapped in a local minimum without a guaranty to converge at the optimal solution (Lin and Lee, 1996). When used to construct neural fuzzy systems, MLP-based approaches often introduce new fuzzy logic aggregation operators that are complicated in nature (e.g., interactive-or operator), making the interpretation of the extracted rules semantically complex. Benitez *et al.* (1997) also noted that the fuzzy rules generated by a MLP network may contain fuzzy numbers out of the valid range of the input data.

ART-based

The ART-based neuro-fuzzy system was proposed by Carpenter *et al.* (1992) in an attempt to achieve synthesis between fuzzy logic and another successful neural network family that is based on adaptive resonance theory. The resulting Fuzzy ARTMAP and their modifications (Chen and Hoberock, 1996; Williamson, 1996; Liu *et al.*, 2004) exploit a formal similarity between the computations of fuzzy subsethood and ART category choice, resonance, and learning (Mannan *et al.*, 1998; Gopal *et al.*, 1999). ARTMAP can realize incremental supervised learning, which allows new patterns to be added into the neural network without a complete retraining. Using a winner-takes-all (WTA) strategy, fuzzy ARTMAP adapts only the weights of the neurons encoding the category that best matches the input pattern rather than all the neurons. This achieves learning stability by keeping previously learned patterns while simultaneously being flexible, or plastic, enough to master new patterns (Liu *et al.*, 2004). In contrast to MLP neural networks, fuzzy ARTMAP needs only a few training epochs to solve large scale problems with its fast learning property. Superior results from fuzzy ARTMAP were reported when applied to remotely sensed data (Mannan *et al.*, 1998; Gopal *et al.*, 1999; Gamba and Dell’Acqua, 2003). However, fuzzy ARTMAP is sensitive to noise and outliers, causing data overfitting and category proliferation that may lead to increased misclassification (Williamson, 1996; Liu *et al.*, 2004). The complex hierarchical network structure makes a straightforward interpretation of fuzzy ARTMAP difficult (Ju *et al.*, 2003) despite the fact that attempts have been made to deal with its “black box” nature to some degree through visualization (Liu *et al.*, 2001). Consequently, most ARTMAP-based neuro-fuzzy systems reported in the literature are only fuzzy neural networks rather than neural fuzzy systems.

SOM-based and LVQ-based

SOM-based and LVQ-based neuro-fuzzy systems are built upon the fuzzification of the well known Kohonen self-organizing map (SOM) (Kohonen, 1989) and learning vector quantization (LVQ) (Kohonen *et al.*, 1996) neural networks. Similar to ARTMAP, SOM adopts the WTA strategy where the winning cluster (prototype) determined by the closest Euclidean distance is moved closer to the matching input pattern (vector). This means that the weights of the SOM neural network

eventually become the center of the clusters. LVQ further improves SOM by pushing the center of the winning cluster away from the input pattern if they fail to match. This delivers faster and more effective learning. Original SOM/LVQ was an unsupervised learning paradigm (Kohonen, 1989), but the later versions of LVQ (LVQ1, LVQ2, LVQ3) incorporate a supervised learning design (Kohonen *et al.*, 1996). An unsupervised fuzzy LVQ (FLVQ) was proposed by Bezdek and Pal (1995) who used fuzzy c-mean algorithms as membership functions to determine the closeness of each cluster from the input pattern. Fuzzy c-mean based unsupervised FLVQ is primarily a fuzzy neural network and has been used to classify various remotely sensed images. Filippi and Jensen (2006) successfully employed FLVQ to classify AVIRIS hyperspectral images without using endmember spectra, and Baraldi *et al.* (1998) and Benz (1999) also provided an excellent review on the applications of FLVQ to multispectral and SAR imagery. SOM/LVQ-based neural fuzzy systems have also been devised. Vuorimaa (1994) built a FSOM system on the basis of the fuzzification of SOM to simulate continuously valued functions with a set of fuzzy “if-then” rules defined by triangular fuzzy membership functions. The fuzzy set parameters were fine-tuned using a modified version of the learning vector quantization (LVQ) algorithm. However, when applied to image classification problems, the algorithm did not converge because the output of a classification problem produces discrete categories rather than continuous values. The system also yielded many unclassified pixels because there was no guarantee that two adjacent triangular membership functions would overlap each other (Qiu and Jensen, 2004). Nomura and Miyoshi (1995) designed a fuzzy inference network (FIN), also based on a fuzzified SOM, using Gaussian membership functions rather than simple triangle functions. Gaussian membership functions are better for image classification because any two such functions will always have overlap with each other and therefore eliminate the problem of unclassified pixels. However, the classification crashed when used with remotely sensed data in the original range and was extremely slow when the data was scaled. It also generated unreasonable overall membership grades as output (Qiu and Jensen, 2004).

Methodology

Qiu and Jensen (2004) proposed a new neuro-fuzzy system with an attempt to open the black box of neural networks. The neuro-fuzzy system was developed on the basis of the fuzzification of the Kohonen’s LVQ2 neural network using Gaussian membership functions, and therefore is a supervised fuzzy neural network. To differentiate it from the aforementioned fuzzy c-mean based FLVQ neural networks that are unsupervised in nature, we refer to it as GFLVQ. While most neuro-fuzzy systems are either supervised or unsupervised classifiers, GFLVQ is externally a supervised image classification but internally takes advantage of unsupervised learning algorithms to form natural clusters within each class when the system is configured with multiple neurons for each class. Being a supervised system, GFLVQ utilizes *a priori* expert knowledge from the training data to effectively separate various land-cover types that exhibit interclass spectral similarity, which are difficult to discriminate using unsupervised systems. GFLVQ, when configured to encode multiple categories for each class, allows the unsupervised self-organizing clustering and competitive learning capability of the SOM/LVQ come into play within each class, so that the intraclass spectral variability often observed in remotely sensed data can also be captured. We selected the SOM/LVQ-based learning algorithm because a GFLVQ with embedded WTA strategy is able to achieve stable and plastic learning by only adapting winner neurons that match the

input pattern, similar to ART-based fuzzy neural networks. Moreover, by incorporating LVQ learning to push the center of a winning cluster away from the unmatched pattern, GFLVQ further speeds up the training process and improves the effectiveness of the classification. The LVQ2 algorithm that the system was actually based on also introduces an additional update rule for adapting the “first runner-up” in cases where one of the first two winners (nearest neighbors) matches the input pattern. This may slow down the training process, but it optimizes the relative distances of the input patterns from each class (Kohonen *et al.*, 1996).

Unlike most neuro-fuzzy systems that are either a fuzzy neural network or a neural fuzzy system, GFLVQ is both a fuzzified neural network and a neural network driven fuzzy expert system. It utilizes adaptive neural network learning algorithms to automatically fine-tune the associated parameters of the fuzzy “if-then” rules while using fuzzy set and fuzzy logic operators to update the weights of the neural network. As a neural fuzzy system, GFLVQ utilizes fuzzy “if-then” rules comprising of fuzzy numbers, special numerical fuzzy sets defined by the Gaussian functions and centered at the clusters that each competitive neuron resides. The simple topological structure of GFLVQ provides a straightforward interpretation of neural network based image classification in the form of symbolic rules. Unlike unsupervised fuzzy c-mean-based FLVQ (Bezdek and Pal, 1995) that only offer a geometric interpretation of the cluster means, GFLVQ provides insight not only into the centers of the clusters but also into their sizes and variances. These symbolic rules also facilitate the opening up of the “black box” in neural networks.

GFLVQ has been used to classify multispectral remotely sensed image, and it significantly outperformed traditional statistical and standalone MLP based neural network approaches (Qiu and Jensen, 2004). However, when applied to hyperspectral imagery, GFLVQ in its original form takes a long time to converge and may become less robust with extended learning periods due to the huge volume of data processed by the system. The fuzzy “if-then” rules generated, while providing insight into the understanding of multispectral image classifications, are not easy to comprehend for hyperspectral imagery, because the rules involve hundreds of fuzzy sets due to the high dimensionality of the data. Therefore, modifications to GFLVQ are needed in order to face the challenges presented by hyperspectral data. Two algorithmic improvements are proposed in order to deliver an efficient and effective solution to hyperspectral image classification problems. Additionally, an interactive geovisualization tool is developed to facilitate the understanding of the hyperspectral image analysis process. The modifications and extension of the system are explained below in detail on the basis of a synoptic review of the original GFLVQ system. Readers are encouraged to refer to Qiu and Jensen (2004) for technique details.

The GFLVQ neuro-fuzzy system consists of three layers: an input layer, a competitive layer, and an output layer (Figure 1). The neurons in the input layer correspond to the input pixel DN values of all hyperspectral bands. The number of neurons in the competitive layer can be equal to or greater than the number of output classes. When the number of competitive neurons is greater than the number of classes, a class is allowed to have multiple clusters so that possible multimodal distributions in the data can be captured. The number of neurons in the output layer equals that of land-cover types. As shown in Figure 1, the output and competitive layers are not fully connected. The neuro-fuzzy system contains the fundamental ingredients of both a neural network and a fuzzy system, and consists of the following components: (a) fuzzification interface, (b) fuzzy inference engine and knowledge base, (c) neuro-fuzzy

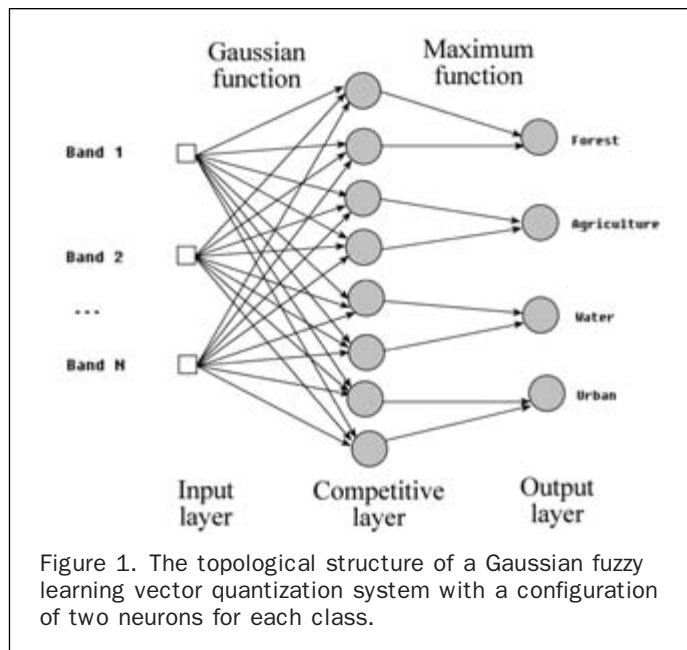


Figure 1. The topological structure of a Gaussian fuzzy learning vector quantization system with a configuration of two neurons for each class.

learning, (d) geovisualization for knowledge discovery and understanding, and (e) neuro-fuzzy classification and defuzzification.

Fuzzification Interface

Inputs into the neuro-fuzzy system are usually not fuzzy numbers. Therefore, the remote sensing input data need to be fuzzified into a set of fuzzy numbers when entered into the neuro-fuzzy system. The fuzzification of a single pixel DN value is accomplished by applying a Gaussian fuzzy membership function:

$$uA_{ij} = \exp\left(-1/2 \frac{(c_{ij} - x_j)^2}{\sigma_{ij}^2}\right) \quad (1)$$

where uA_{ij} is the membership grade for fuzzy number A_{ij} , c_{ij} is the mean parameter of the Gaussian function corresponding to the center of the i^{th} cluster of the j^{th} band, x_j is the j^{th} input variable (i.e., pixel DN value for j^{th} band), and σ_{ij} represents the Gaussian standard deviation parameter characterizing the size of the cluster and determining its fuzzy boundary. The system then uses the derived fuzzy membership grade rather than the Euclidean distance as a relative distance to determine the closeness of the input pixel to a cluster. It takes into consideration the data distribution of the clusters by using the mean and standard deviation parameters learned by the neuro-fuzzy algorithm as described below, which increases accuracy in pixel assignment to clusters with different dispersion sizes (Qiu and Jensen, 2004). The use of fuzzy membership grades also reduces the algorithm’s sensitivity to noise and outliers and prevents overfitting and category proliferation.

Fuzzy Inference Engine and Knowledge Base

A single Gaussian membership function determines the fuzzy membership grade of a pixel based on a single band. To assign a pixel to a particular class, pixel DN values of all bands are considered by a fuzzy “if-then” rule to derive an overall membership grade for the pixel. A fuzzy “if-then” rule is a combination of fuzzy sets (or fuzzy numbers) defined for all bands connected by a specific type of fuzzy logic operator \cap :

If band 1 is fuzzy number $A_{i1} \Omega$, band 2 is fuzzy number $A_{i2} \Omega$, ... band j is fuzzy number $A_{ij} \dots \Omega$ band, and n is a fuzzy number A_{in} , then the pixel y is class C_i

where n is the total number of input bands, C_i is the i^{th} target information class, and j is the band number. For a hyperspectral image, this rule will be very long with hundreds of antecedents. The fuzzy logic operator plays an important role in determining the overall membership grade for the pixel and the degree of the rule being activated (Nomura and Miyoshi, 1995; Benitez *et al.*, 1997). We use a new fuzzy *and-or* operator in the form of a geometric mean:

$$\text{and-or}(a_1, a_2, \dots, a_n) = (a_1 * a_2 \dots a_n)^{1/n} \quad (2)$$

where a is the membership grade for an input band and n is the number of bands. By geometrically averaging the membership grades of all the bands, the overall membership grade α_i of a fuzzy rule concerning i^{th} target information class becomes¹:

$$\alpha_i = \left(\prod_{j=1}^n \exp\left(-1/2 \frac{(c_{ij} - x_j)^2}{\sigma_{ij}^2}\right) \right)^{1/n} \quad (3)$$

where i, j, n, c , and σ are the same as previously defined. The *and-or* operator is a fuzzy operator sitting in-between the fuzzy *and* (intersection) and fuzzy *or* (union) operators. As an averaging operator, it allows a low membership grade of one band to be compensated by a high membership grade of another band so that a missing or noisy value in one band will not heavily affect the classification output of the entire pixel. In addition, the *and-or* operator is an "idempotent" function such that *and-or*(a, a, a) = a delivers a more reasonable final membership grade. The rules derived for all classes are used to construct the knowledge/rule base, which can be used to perform image classification without employing the actual neural network. One of the great features of GFLVQ is that you can configure the number of neurons in competitive layers to be greater than that of neurons in the output layers, allowing multiple clusters within a single output class. Also, multiple fuzzy if-then rules can be extracted for these multiple cluster classes. For example, for a network with twice as many competitive neurons as output neurons, two fuzzy "if-then" rules consisting of fuzzy sets connected by *and-or* operators can be extracted for each class. These two rules are then joined with a standard fuzzy union (*or*) operator (i.e., the maximum function) to determine the final membership grade.

Neuro-fuzzy Learning

The topological structure of the underlying neural network predetermines the number and form of fuzzy "if-then" rules used in the fuzzy inference engine. Each rule consists of several fuzzy sets (fuzzy numbers) defined by Gaussian functions with the two parameters of each Gaussian function (mean c and standard deviation σ) being the two weights of each neuron. The two weights of each neuron are initialized with random numbers within the range of valid inputs and updated using the following neuro-fuzzy learning algorithm. The updating of the mean parameter is accomplished by using Kohonen's supervised LVQ learning algorithm, utilizing the true target class information provided by the training data:

$$\Delta c_{ij} = \eta(x_j - c_{ij}), \text{ if } x \text{ and } c \text{ belong to same class} \quad (4)$$

$$\Delta c_{ij} = -\eta(x_j - c_{ij}), \text{ otherwise}$$

where η is the learning rate, a small positive number that decreases with training time. The supervised LVQ learning method has the effect of moving the mean parameter c_{ij} (the cluster center) towards the matched (correctly classified) input pattern and away from the unmatched (incorrectly classified) input pattern, simultaneously increasing both learning speed and classification accuracy. We devised a new learning rule to adjust the parameter σ , which updates the standard deviation of the Gaussian function (Qiu and Jensen, 2004):

$$\Delta \sigma_{ij} = (|x_j - c_{ij}| - \sigma_{ij}), \text{ if } x \text{ and } c \text{ belong to same class} \quad (5)$$

$$\Delta \sigma_{ij} = 0, \text{ otherwise.}$$

The geometric interpretation of this updated rule is that when the absolute deviation of the input pattern x_j from the center of the matched cluster (centered at c_{ij}) is larger than the current standard deviation σ_{ij} of the cluster, the standard deviation will be increased by η fraction of the difference. If the absolute deviation is smaller than the standard deviation, then the standard deviation will be decreased by a small fraction of the difference. This ensures that the size of the cluster shrinks or enlarges adaptively based on the deviation of the matched input patterns from the cluster center. If the input pattern x_j belongs to a different class than the winner cluster, no update of σ_{ij} is needed because the cluster has nothing to do with the input pattern.

In the original implementation of the GFLVQ system, the LVQ2 learning algorithm was actually employed, which update the first two winning clusters using Equation 4. It assumes that one winner belongs to the same class as the input pattern and the other does not. Having to update two clusters at a time, the LVQ2 algorithm slows down the training process a little in exchange for optimal relative distances to each class of a multispectral image. However, this delay can be substantial for hyperspectral data with hundreds of bands. Kohonen *et al.* (1996) also states that LVQ2 is less robust than LVQ1 and LVQ3. With only one cluster for each class in the network, if one of the first two winning clusters matches the input pattern the other will not. However, GFLVQ allows two or more clusters for one class, making it possible that both of the first two winning clusters match the input pattern. If that is the case, the centers of both winning clusters will be updated to move towards the input pattern, which will eventually merge two clusters and cause the system to become unstable. Therefore, to expedite the learning process and increase the robustness of the system, a modification was made in the improved GFLVQ so that it only considers the first winning cluster as in LVQ1. The new GFLVQ is now a fuzzified LVQ1 neural network. The consideration of the second winner is not necessary in GFLVQ because the optimization of relative distances is now achieved through Gaussian-based fuzzy membership instead of absolute Euclidean distance.

Another algorithmic modification brought into GFLVQ specifically for hyperspectral data analysis is the initialization of the network weight (i.e., c_{ij} and σ_{ij}). Instead of assigning random numbers for initialization like most unsupervised systems, we now make use of the training data for parameter initialization to reduce the elongated convergence time for classifying hyperspectral data. The improved GFLVQ first randomly subdivides the training data for each class into subsets according to the number of neurons (clusters) configured for that class. Then, the mean and standard deviation of each subset are calculated and assigned to corresponding neurons as initial weights. A similar approach has been used to derive class centroids by Zhang and Foody (2001) to build a one-step supervised

¹This equation for the original GFLVQ has a typo in Qiu and Jensen (2004). This is the corrected equation that changes the summation sign to a production sign.

fuzzy c-mean classifier. The differences are that GFLVQ initializes the mean and standard parameters for each cluster instead of each class, and fine-tunes the parameters by the fuzzy LVQ learning algorithm to derive a better fuzzy decision boundary rather than using them in a one-step solution.

Geovisualization for Knowledge Discovery and Understanding

Based on the interpretation of these fuzzy numbers, the knowledge used by the neuro-fuzzy classification system can be derived in the form of fuzzy “if-then” rules. These fuzzy “if-then” rules are composed of fuzzy numbers from different bands connected by the *and-or* logic operator. The fuzzy if-then rules derived from GFLVQ are helpful in understanding the neuro-fuzzy classification process. For example, a fuzzy if-then rule derived from a three-band multispectral image may be expressed like this, “If the NIR band is a fuzzy number 19.070 (with a fuzzy set boundary of 0.307) *and-or* the red band is a fuzzy number 43.256 (with a fuzzy set boundary of 0.653) *and-or* the green band is a fuzzy number 53.858 (with a fuzzy set boundary of 0.466), then the pixel is a water class.” For a hyperspectral image, this rule would be very long. Comprised of hundreds of fuzzy numbers, the rule would be difficult to comprehend, providing only limited insights into the classification process. The third enhancement to GFLVQ is therefore an extension to improve knowledge discovery and understanding of the classification process through a geovisualization tool. Based on the learned fuzzy membership parameters, fuzzy spectral profile plots are created for each target class with the x-axis representing the wavelength (or band number) and the y-axis representing the pixel DN number. This is similar to the fuzzy spectral signature plot proposed by Bastin *et al.* (2002) but with many improvements. Instead of displaying a spectral signature curve of a single endmember in a traditional spectral profile, the fuzzy spectral profile provides a visualization of the signature surface of the target class using different gray intensities on a gray ramp to depict the fuzzy membership grade of a pixel with a certain DN number (*x*) in a specific band (*y*). The points with maximum fuzzy membership in the fuzzy spectral profile correspond to the cluster center of each band, and are highlighted to form a continuous line equivalent to the signature curve of an endmember in a traditional spectral profile. The fuzzy spectral profile plot can be dynamically linked to an image viewer where dragging over a pixel with a “brush” will trigger the spectral curve of the pixel to appear on the fuzzy spectral surface as a highlighted curve line, visually illustrating the relative closeness of the pixel to a cluster center of the class. Human expertise can also be incorporated into the classification process by modifying the fuzzy rules and/or their membership parameters to improve classification performance. The geovisualization and fuzzy symbolic representation jointly help open the black box of the neural network and facilitate automated knowledge discovery and acquisition of the fuzzy system.

Neuro-fuzzy Classification and Defuzzification

With all the fuzzy “if-then” rules formed and their parameters automatically learned from data and further fine-tuned by experts, the rules can be used to directly classify pixels into land-use classes. In a hard classification, the result is a single classification map where a pixel is assigned to only one land-use type. However, the output of the neuro-fuzzy system is fuzzy membership grades for all classes, which helps provide information about various constituents found in a pixel if properly utilized (Jensen, 2005). If a hard classification map is desired, defuzzification, a reverse of the

fuzzification process, can be used to convert various fuzzy membership grades to a single classification type. Defuzzification is obtained by comparing all the membership grades of a pixel and then assigning it to the class with the maximum membership grade.

Case Study

A case study was conducted using a subset of a Hyperion Level 1 image acquired in Wuxi, China (centered at +31° 33' 10.80", +120° 9' 54.00") on 19 August 2004. Hyperion is a push broom hyperspectral imaging spectrometer onboard the Earth-Observing 1 (EO-1) spacecraft. Unlike widely used airborne hyperspectral instruments such as AVIRIS, Hyperion is a satellite sensor system with a coarse spatial resolution of 30 m by 30 m. The Hyperion Level 1 product was preprocessed by using image artifacts correction (for SWIR echo and smear), subtraction of dark frame, as well as an on-orbit absolute radiometric calibration to most of the bands (Barry, 2001). There were 242 unique spectral channels with a band width of about 10 nm for each band in the visible/near infrared (VNIR) spectral region (from 400 to 1,000 nm) and short-wave infrared region (from 900 to 2,500 nm). Spectral channels 1 to 70 were collected from the VNIR, and channels 71 to 242 were collected from the SWIR. However, some channels did not have a sufficient signal and were therefore not calibrated, resulting in their pixel value to be set to zero. Some other channels fell into the overlap region (from 900 to 1,000 nm) between VNIR and SWIR, resulting in 196 unique wavelengths (Barry, 2001). To alleviate the collinearity problem and ease the effort of collecting enough training samples in each class for the MLC to be performed, we selected most of the calibrated VNIR bands and discarded the overlapped SWIR bands. For the remaining bands, a uniform feature design (UFD) was applied as suggested by Filippi and Jensen (2006) to reduce the dimensionality of the data set while retaining as much spectral shape information as possible in the SWIR range. This resulted in a total of 112 bands used in the subsequent analysis. In this spectral subset of the original image, bands 1 to 21 were visible, bands 22 to 71 were near infrared, and bands 72 to 112 were middle infrared. In an effort to avoid a loss of data precision, a normalization process was performed scaling the pixel reflectance values to the range of 0 to 255. This was done because if original reflectance numbers are used, very small decimal numbers may be derived as a result of product and geometric mean operations, introducing rounding error. A VNIR color composite map of the study area is shown in Figure 2a with a band combination corresponding to 993 nm, 603 nm, and 711 nm. Four land-cover types at Level I of the USGS Classification System (Anderson *et al.*, 1976) were identified in the image and used for classification purposes, including urban/built-up, agriculture, forest, and water. Airborne hyperspectral images with finer spatial resolutions allow the extraction of information at lower levels in the classification system by identifying targets such as vegetation species, rock or soil type, or even mineral types. The satellite based Hyperion image with a 30 m by 30 m spatial resolution only permits the extraction of details at a high level in the classification scheme. A total of 577 training samples were identified in the image through a stratified random selection, of which 169 were for urban/built-up, 137 for agriculture, 151 for forest, and 120 for water. Every class has slightly more training samples than the number of bands, so MLC can be performed without encountering singular variance-covariance matrices. For comparative

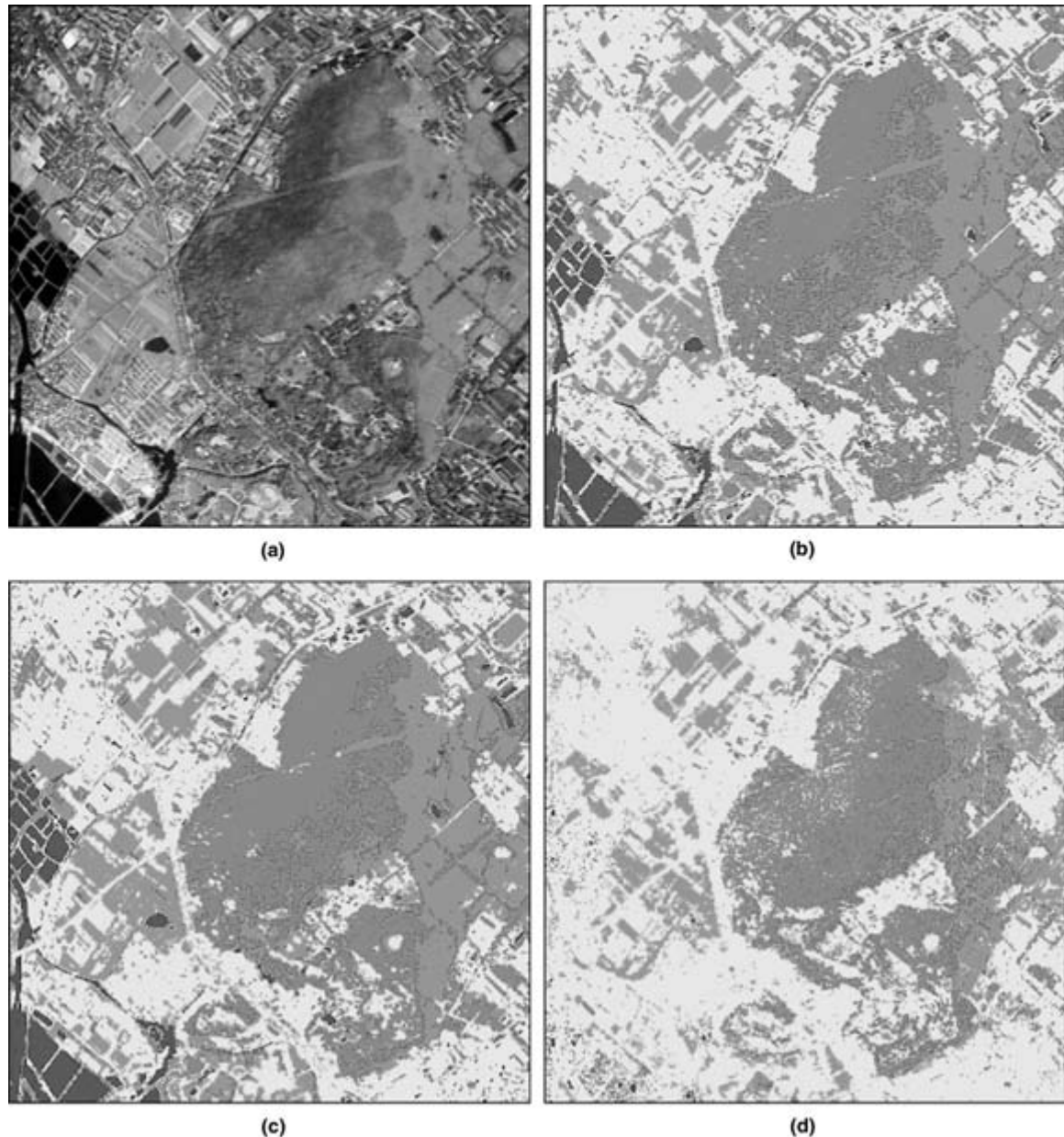


Figure 2. (a) Hyperion image VNIR color composite map (R = 993 nm, G = 603 nm, B = 711 nm); (b) through (h) Classification results from various image classifiers (red = forest, green = agriculture, yellow = urban, blue = water): (b) GFLVQ₁ classification, (c) GFLVQ₂ classification, (d) MLC classification,

purposes, the same samples were used for GFLVQ training and classification. Another 491 random samples (130 for urban, 118 for agriculture, 124 for forest, and 119 for water) were collected as ground reference for accuracy assessment purposes.

In order to perform endmember based hyperspectral image classification, an endmember spectrum representing a pure pixel needed to be identified for each class. However, endmembers are extremely difficult to locate in the Hyperion image because the majority of pixels are mixed, and pure pixels are relatively rare with the coarse spatial resolution. To achieve a best possible extraction of endmembers and to allow comparisons between the results from the endmember based

classifiers and those from MLC and GFLVQ, we adapted an endmember selection strategy suggested by Tompkins *et al.* (1997). This strategy combines the objective endmember selections grounded on purely statistical algorithms and the subjective endmember determinations incorporating *a priori* knowledge of the image and spectral properties of the scene to selection endmember. We started with a minimum noise fraction (MNF) transform (Green *et al.*, 1988) to remove noise components and reduce the dimensionality of the image to a manageable size by two consecutive principal component transforms. On the MNF image, we conducted a pixel purity index (PPI) analysis (Boardman *et al.*, 1995) to the training samples selected. In this way, we first excluded most of the

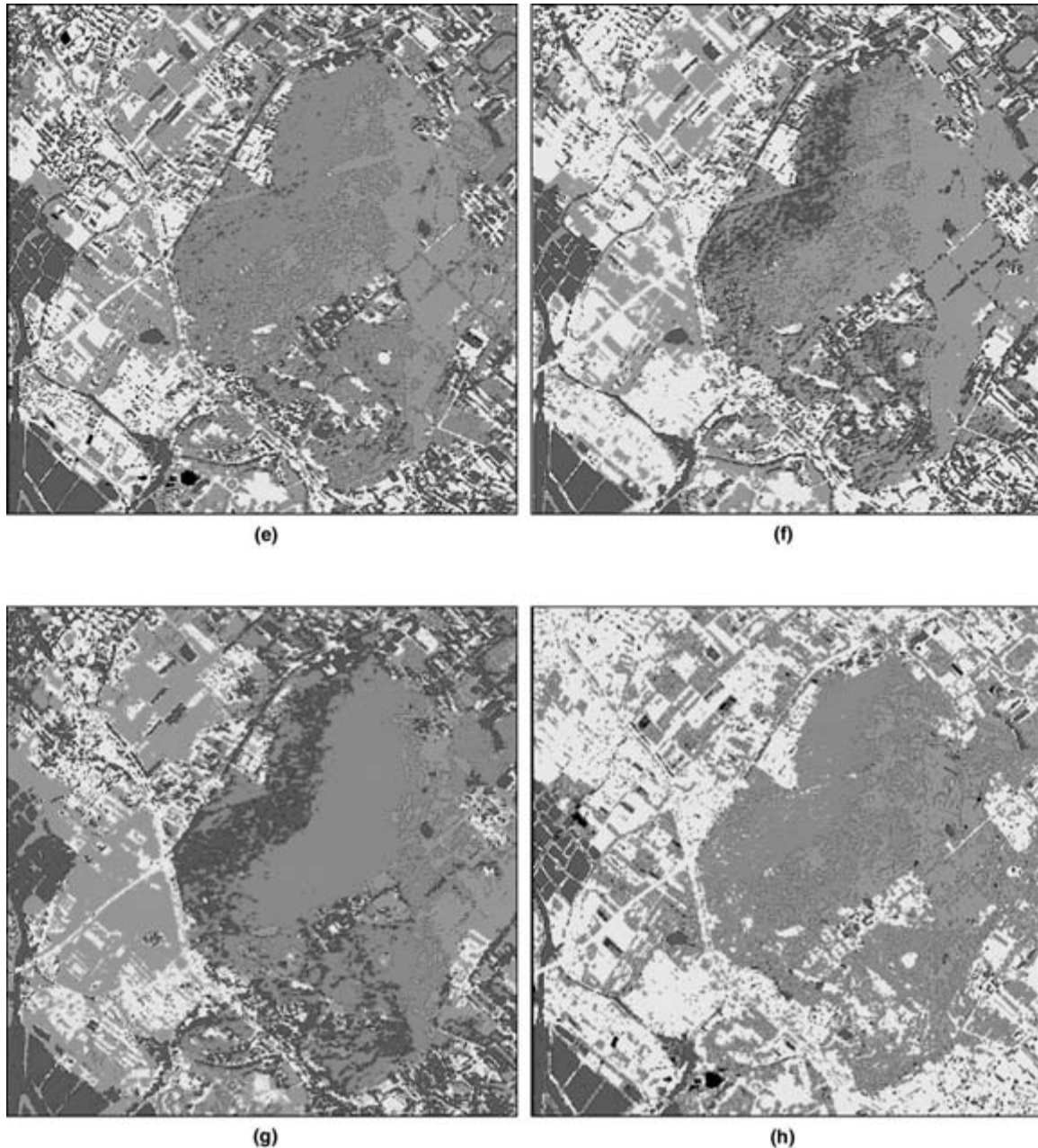
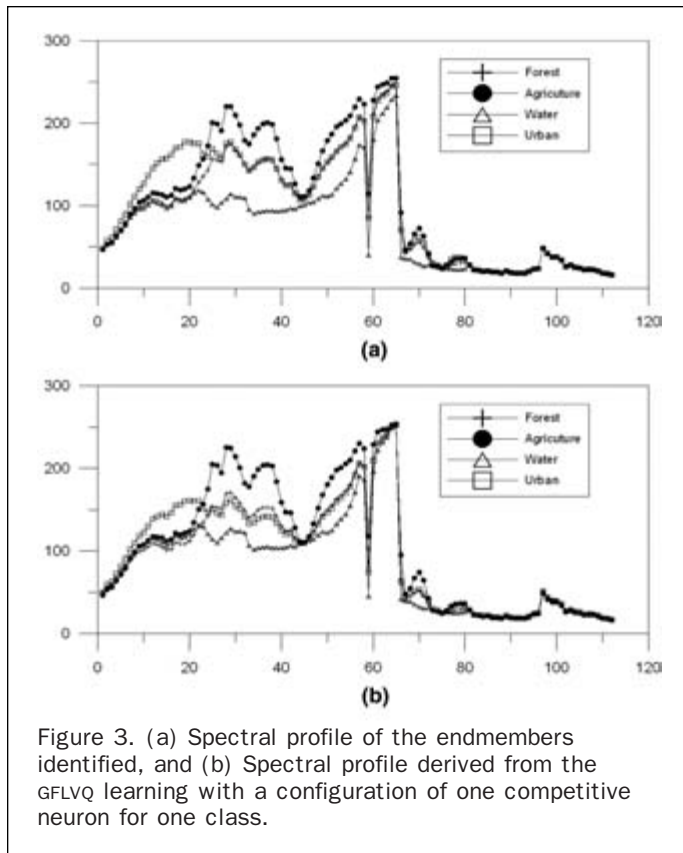


Figure 2., continued: (e) Spectral angle mapping (SAM), (f) Linear spectral unmixing (LSU), (g) Matched filter (MF), and (h) Spectral feature fitting (SFF). A color version of this figure is available at the ASPRS website: www.asprs.org.

pixels that are not pure, and then mathematically elect “extreme” pixels falling onto the ends of the units vectors based on a convex geometry argument. After applying a threshold of relative purity, we further refined the selection of spectrally pure endmembers and eliminated less “pure” pixels as candidate endmembers by performing a manual n -dimensional endmember visualization (Boardman and Kruse, 1994) through interactive rotation of the high-dimensional pixel clouds. The convex corners of the pixel clouds were identified and designated as the purest spectral endmembers.

Figure 3a displays the spectral profile of the endmembers determined for each class. Omitting un-calibrated and overlapping bands from the original data, the scaled

reflectance values were presented relative to the new band numbers of the Hyperion spectral subset because using actual wavelengths would result in a discontinuity in the spectral profile. It is observed that the shape of the spectral curves for the forest and agriculture endmembers are similar but with a certain separability in their magnitudes. These endmembers were then subjected to endmember based hyperspectral image classifiers. The algorithmic details of these classifiers, including spectral angle mapping (Kruse *et al.*, 1993), linear spectral unmixing (Boardman, 1992), matched filtering (Harsanyi and Chang, 1994), and spectral feature fitting (Swayze and Clark, 1995), are not discussed in this paper as they have been presented in detail in their corresponding references.



Results and Discussions

The classification maps resulted from all the classification methods employed in this study are displayed in Figure 2b through 2h with red color standing for forest, green for agriculture, yellow for urban, and blue for water. Table 1 shows the corresponding overall accuracy and the Kappa coefficients (K_{hat} statistics) of the classification results based on the 491 random test points. K_{hat} is believed to be a better representation of the general quality of classification because it removes the effects caused by differences in sample size and also accounts for the off-diagonal elements in the error matrix (Rosenfield and Fitzpatrick-Lins, 1986). According to Fleiss (1981), Kappa coefficients larger than 0.75 suggest stronger agreement than mere chance. Landis and Koch (1977) suggested that Kappa coefficients larger than 0.81 are almost perfect. A high overall accuracy of 89 percent was achieved for the GFLVQ classification system with only one neuron for each class (referred to as GFLVQ₁), and its Kappa coefficient (K_{hat}) is 0.86. The classification with GFLVQ₁ (Figure 2b) produced accurate mapping for most classes, with a small portion of the forest class in the middle of the image misclassified as agriculture or urban. The overall accuracy of GFLVQ using two neurons for each

class (referred to as GFLVQ₂) is further increased to as high as 92 percent and the K_{hat} to 0.90. With the addition of one cluster for each class in the competitive layer, the misclassification of the forest class was greatly reduced (Figure 2c), leading to a much improved mapping of land-cover types. The MLC performed poorly in classifying the hyperspectral image, yielding an overall classification accuracy of 66 percent and a K_{hat} of 0.55. The water class, usually the easiest class to classify for MLC in multispectral image, is almost completely missing in the classification map of the hyperspectral image (Figure 2d). This is probably due to a relatively small training sample size for the water class (120), which implies sample size has a heavy impact on the success of MLC. The endmember based classifiers yielded inferior classification results compared to both of the GFLVQ classifiers. A relatively high overall accuracy of 81 percent and a high K_{hat} of 0.75 were achieved for both linear spectral unmixing (LSU) and spectral feature fitting (SFF). LSU greatly overestimated the water class and agricultural class, but vastly underestimated the forest class (Figure 2f). SFF exhibits remarkable confusion between forest and agriculture, with forest being overestimated, but agriculture land underestimated (Figure 2h). An overall accuracy of 0.77 and a K_{hat} of 0.69 were attained for spectral angle mapping (SAM) with water overestimated, urban underestimated, and some misclassification between forest and agriculture (Figure 2e). A overall accuracy of 62 percent and a K_{hat} of 0.49 were obtained for the matched filtering (MF) classifier, with urban tremendously underestimated, water massively overestimated, and forest and agriculture greatly confused with each other (Figure 2g). Kappa coefficients also offer another advantage in that different classifications can be compared statistically with each other (Rosenfield and Fitzpatrick-Lins, 1986). A significance Z-test was performed between GFLVQ₁ and GFLVQ₂, and all classifiers using K_{hat} values and their corresponding variances (Table 2). The results indicate that the classification accuracy of GFLVQ₁ was significantly better than those of MLC and all endmember based image classifiers at a 99 percent confidence level. GFLVQ₂ not only significantly outperformed all non-GFLVQ classifiers at the 99 percent confidence level, it was also better than GFLVQ₁ at the 90 percent probability level.

The fuzzy spectral profiles derived from the geovisualization tool of the two GFLVQ classifications are displayed in Figure 4a and 4b, which illustrate the actual spectral dispersion in each cluster of the classes. Figure 4a shows the fuzzy spectral profile plot of the system configured with one neuron for each class (GFLVQ₁). The maximum fuzzy membership points connect with each other to form a contour line (highlighted in light gray), characterizing the center spectral value of the clusters (i.e., the mean weight of the competitive neurons). These central spectral curves, when extracted and put together in a single plot (Figure 3b), are similar to the traditional spectral profiles displaying identified endmembers (Figure 3a). This suggests that GFLVQ is capable of capturing

TABLE 1. OVERALL ACCURACY AND KAPPA COEFFICIENT OF THE CLASSIFICATION RESULTS

	GFLVQ ₁	GFLVQ ₂	MLC	SAM	LSU	MF	SFF
Overall Accuracy	89%	92%	66%	77%	81%	62%	81%
Kappa Coefficient	0.86	0.90	0.55	0.69	0.75	0.49	0.75

TABLE 2. Z-TEST OF KAPPA COEFFICIENTS BETWEEN GFLVQ CLASSIFIERS AND OTHER CLASSIFIERS

	GFLVQ ₁	GFLVQ ₂	MLC	SAM	LSU	MF	SFF
GFLVQ ₁	N/A	-1.6585	9.5764**	5.3722**	3.6801**	10.7809**	3.6488**
GFLVQ ₂	1.6585*	N/A	11.3328**	7.0017**	5.3468**	12.4835**	5.287**

* - Significant at the 90% confidence level.
 ** - Significant at the 99% confidence level.

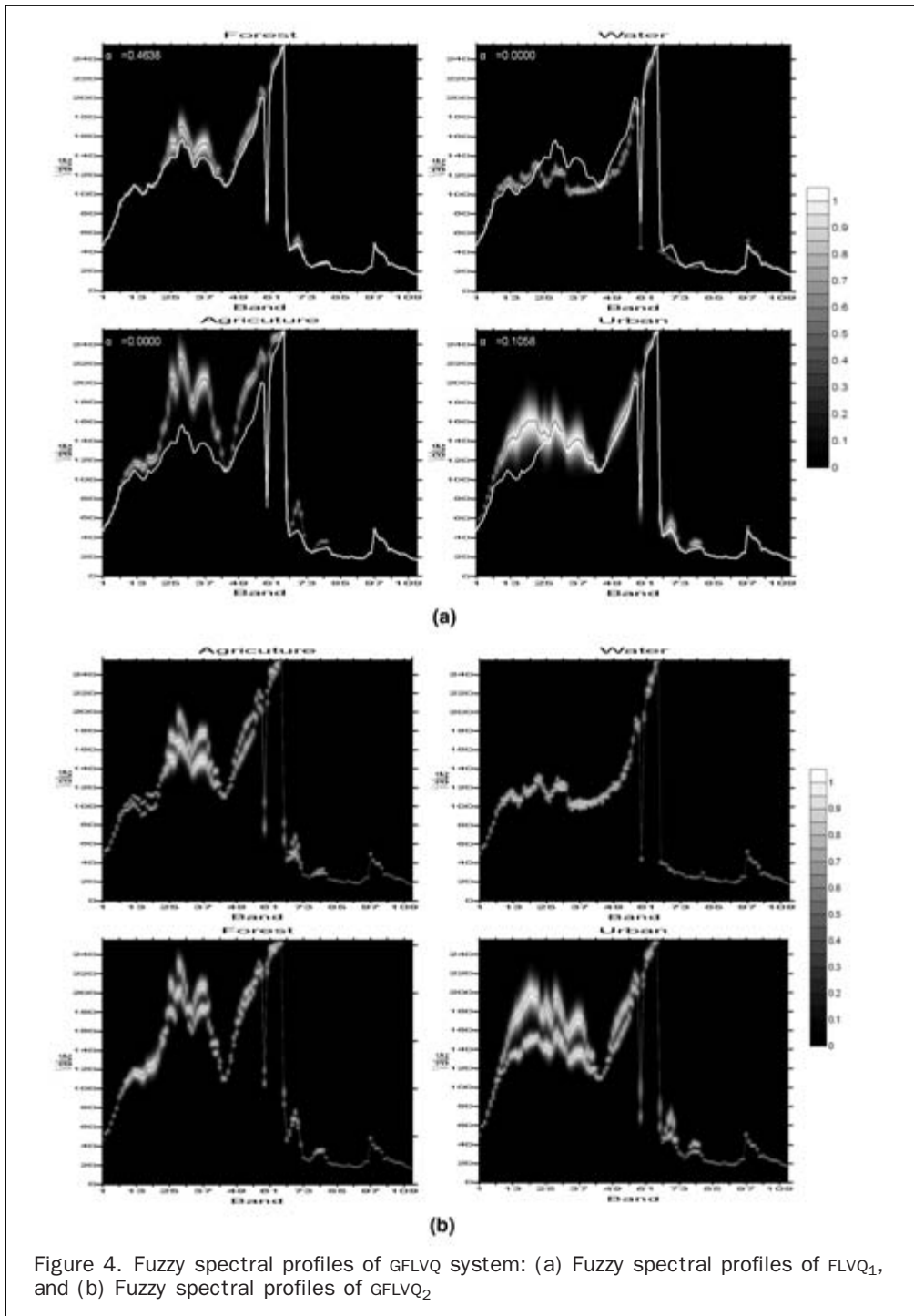


Figure 4. Fuzzy spectral profiles of GFLVQ system: (a) Fuzzy spectral profiles of FLVQ₁, and (b) Fuzzy spectral profiles of GFLVQ₂

characteristic patterns from sample data through its neuro-fuzzy self-organizing and competitive learning. Additionally, all the fuzzy membership grades, when displayed with different intensities in a gray ramp, suggesting the size of the cluster by providing a depiction of the spread of spectral clusters from their central values. The spreading may be broad or tight in different portions of the spectrum. For example, most of classes have a wider spreading in the VNIR region than in the SWIR region. The clusters of different classes may have a broad or tight spreading in general. For example, urban has a much broader spreading than water, suggesting a larger multidimensional spectral cluster for urban than for water. With a brushing tool applied to the image viewer dynamically linked to the fuzzy spectral profile plot, a user can drag over a pixel with the “brush” to display the spectral curve of the pixel on the fuzzy spectral surface. The user can then inspect the closeness of the pixel to the center curve of each cluster in Euclidean distance (which can also be provided by the traditional spectral profiles) and in fuzzy distance based on its relative location inside the spreading of each spectral cluster. In Figure 4a the spectral curve of the pixel is apparently a forest class because it is not only close to the center spectral curve of the forest cluster but also fully lies within the spread of the forest cluster while not completely within the coverage of any other cluster. The geovisualization tool also reports the overall fuzzy membership of the pixel in each class in the upper-left-hand corner of the fuzzy spectral profile plot. For example, the highest fuzzy membership of the pixel in 4a is for forest ($\alpha = 0.4638$), and the lowest are for water and agriculture ($\alpha = 0.000$, a very small number that can not be fully represented with four decimal places).

Figure 4b shows the fuzzy spectral profiles of the system with two competitive neurons for each class (GFLVQ₂) produced by taking the maximum fuzzy membership grades of the two clusters (i.e., fuzzy union operation). In addition to all features that are available to GFLVQ₁ fuzzy spectral profiles, this plot offers a means to explore the distribution of the training samples. It is observed that a bimodal distribution can be used to better characterize urban, forest, and agriculture classes. This is probably the reason why GFLVQ₂ outperformed GFLVQ₁. However, the water class is an exception due to the homogeneous nature of most water surfaces. The central spectral curves of the two water clusters are similar in pattern and share a great deal of overlap. With expert intervention, the water class could be configured with only one neuron as it was found that eliminating the redundant water cluster did not affect the final classification.

GFLVQ's two different configurations both produced significantly better classification results than statistics and endmember based classifiers. MLC is not actually an appropriate hyperspectral image classifier due to the potential singular matrices problem. With extra effort to collect more training samples than the number of bands for each class, singular matrices may be avoided, but yielding optimal results from MLC is not always guaranteed if the sample size is not substantially large. MLC is used in this paper as a reference classifier because both MLC and GFLVQ employed Gaussian functions and there appears to be a similarity between MLC and GFLVQ₁. However, MLC and GFLVQ are fundamentally different in their underlying philosophy. MLC relies on large training samples to derive statistically sound parameters. It is recommended that a sample of training data at a minimum of 10 times the number of bands is needed to calculate variance-covariance matrices (Jensen, 2005), which is very

difficult to achieve for hyperspectral data. This explains why MLC yields a poor classification even for the water class. Operating in a non-parametric manner, GFLVQ uses adaptive learning to reach and fine-tune optimal decision rules, resulting in superior classification with the same small training samples. MLC assumes a multivariate Gaussian error model and requires an approximately normal distribution for each class in all bands. This is not always possible even with a pure class. For example, due to differences in species uniformity, age, health condition, and canopy structure, etc, a forest class may follow a multimodal distribution. A possible solution could be to visually analyze the modes in the histograms of each band and divide the training samples of a class into small subsets (Ju *et al.*, 2003). Each subset would be treated and classified as an individual subclass (e.g., water 1 and water 2) in MLC and would be relabeled as a single class (e.g., water) after classification. However, the identification of subclasses by visually inspecting the histogram of each band in a hyperspectral image becomes extremely impractical if not completely impossible. GFLVQ, when configured such that each class is associated with multiple neurons, does not need to separate training samples of each class into different subclasses through visual histogram analysis. The internal unsupervised classification capability embedded in GFLVQ is able to adaptively model the multimodality of the training samples and automatically assign training samples into different clusters of the class with improved accuracy.

The operation of GFLVQ does not rely on endmembers, which are difficult to identify in satellite images with a coarse spatial resolution, such as the Hyperion image. Endmember based hyperspectral image classifiers may suffer from the “mixed pixels” problem and have produced inferior results, because the purity of the endmembers is critical to the success of these classifiers. Water is the only class that has a high producer's accuracy (above 90 percent, not shown) for all endmember based classifiers because most water pixels are pure. It is rare to find pure pixels for other classes in a Hyperion image. Some of the endmember based classifiers require the extraction of all endmembers for materials present in a given scene be extracted, which further troubled certain hyperspectral classifiers like linear spectral unmixing. Even if all endmembers can be determined with the highest purity, relying only on the endmembers, these conventional hyperspectral classifiers limit themselves to only the center vector of the spectral clusters, leaving the extra information about dispersions of the clusters provided by the training samples.

Conclusions

An improved Gaussian Fuzzy Learning Vector Quantization hyperspectral image classifier is proposed based on modifications and extensions to the original version designed for multispectral image classification. The improved system is a fuzzified LVQ1 neural network that can quickly reach classification convergence through training data, and a simplified fuzzy learning algorithm applied only to the first winning cluster. GFLVQ is a supervised image classifier that makes use of *a priori* knowledge in the training data to update a single neuron without complete retraining. When configured with multiple competitive neurons for each class, GFLVQ is able to self-organize the training samples into natural spectral cluster, taking advantage of the unsupervised competitive learning ability embedded in the system. A geovisualization tool is also developed to

help to visualize and understand the fuzzy expert system and its “if-then” rules through fuzzy spectral profile plots, allowing users to interact with individual pixels to explore sampling distributions. When applied to a Hyperion image, the two different configurations of GFLVQ both significantly outperformed the statistics based maximum likelihood classifier and the endmember based hyperspectral image classifiers at a 99 percent confidence level.

Future studies will be directed to further improve the system for subpixel image analysis and mapping, as well as high spatial resolution image classification and feature extraction. An automatic determination of the number of competitive neurons will also be explored in order to construct a fully adaptive, incremental learning driven neuro-fuzzy system.

Acknowledgments

The author is appreciative to the help received from the colleagues at the University of Texas at Dallas. Ms. Caiyun Zhang assisted in great deal in the implementation and application of the improved system. Mr. Shaofei Chen provided assistance in classification accuracy assessment, and Mr. Qi Li helped with the migration of the system from the UNIX system to the Windows® environment.

References

- Abe, S., and M. Lan, 1996. Effective method for fuzzy rule extraction from numerical data, *Fuzzy Logic and Neural Network Handbook* (C.H. Chen, editor), New York, McGraw-Hill, pp. 7.1–7.33.
- Anderson, J.R., E. Hardy, J. Roach, and R. Witmer, 1976. *A Land Use and Land Cover Classification System for Use with Remote Sensor Data*, USGS Professional Paper 964, Washington, D.C., 28 p.
- Baraldi, A., P. Blonda, F. Parmiggiani, G. Pasquariello, and G. Satalino, 1998. Model transitions in descending FLVQ, *IEEE Transactions on Neural Networks*, 9:724–738.
- Barry, P., 2001. *EO-1/Hyperion Science Data User's Guide*, TRW Space, Defense & Information Systems, One Space Park, Redondo Beach, California.
- Bastin, L., P.F. Fisher, and J. Wood, 2002. Visualizing uncertainty in multi-spectral remotely sensed imagery, *Computers & Geosciences*, 28:337–350.
- Benediktsson, J.A., K. Amason, J.R. Sveinsson, and A. Eiriksson, 1994. Classification of AVIRIS data from Iceland-Geological applications, *International Geoscience and Remote Sensing Symposium, Surface and Atmospheric Remote Sensing: Technologies, Data Analysis and Interpretation*, 08–12 August, Pasadena, California, California Institute of Technology, IV, pp. 2351–2353.
- Benz, U., 1999. Supervised fuzzy analysis of single and multichannel SAR data, *IEEE Transactions on Geoscience and Remote Sensing*, 37:1023–1037.
- Beñítez, J.M., J.L. Castro, and I. Requena, 1997. Are artificial neural networks black boxes?, *IEEE Transactions on Neural Networks*, 8:1156–1164.
- Bezdek, J., 1981. *Pattern Recognition with Fuzzy Objective Function Algorithms*, Kluwer, Academic Publishers, Norwell, Massachusetts.
- Bezdek, J.C., and N.R. Pal, 1995. Two soft relatives of learning vector quantization, *Neural Networks*, 8:729–743.
- Binaghi-Madella, E.P., M.G. Montesano, and A. Rampini, 1997. Fuzzy contextual classification of multisource remote sensing images, *IEEE Transactions on Geoscience and Remote Sensing*, 35:326–340.
- Boardman, J.W., 1992. *Sedimentary Facies Analysis Using Imaging Spectrometry: A Geophysical Inverse Problem*, Ph.D. dissertation, University of Colorado, Boulder, Colorado.
- Boardman, J.W., and F.A. Kruse, 1994. Automatic spectral analysis: A geological example using AVIRIS data, North Grapevine Mountain, Nevada, *10th Thematic Conference on Geologic Remote Sensing*, Ann Arbor, Michigan, pp. 407–418.
- Boardman, J.W., F.A. Kruse, and R.O. Green, 1995. Mapping target signatures via partial unmixing of AVIRIS data, *Summaries, 5th JPL Airborne Earth Science Workshop*, pp. 23–26.
- Buckley, J.J., and Y. Hayashi, 1994. Fuzzy neural networks: A survey, *Fuzzy Sets and Systems*, 66:1–13.
- Carpenter, G., M. Gjaja, S. Gopal, and C. 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.
- Carpenter, G.A., S. Grossberg, N. Markuzon, J.H. Reynolds, and D.B. Rosen, 1992. Fuzzy ARTMAP: A neural network architecture for incremental supervised learning of analog multidimensional maps, *IEEE Transactions on Neural Networks*, 3:698–712.
- Chen, B., and L.L. Hoberock, 1996. A fuzzy neural network architecture for fuzzy control and classification, *Neural Networks, IEEE International Conference*, 03–06 June, 2:1168–1173.
- Civco, D.L., 1993. Artificial neural network for land cover classification and mapping, *International Journal of Geographical Information System*, 7:173–186.
- Fausett, L., 1994. *Fundamentals of Neural Networks: Architectures, Algorithms, and Applications*, Englewood Cliffs, New Jersey, Prentice Hall.
- Filippi, A.M., and J.R. Jensen, 2006. Fuzzy learning vector quantization for hyperspectral coastal vegetation classification, *Remote Sensing of Environment*, 100:512–530.
- Foody, G.M., M.B. McCulloch, and W.B. Yates, 1995. Classification of remotely sensed data by an artificial neural network: Issues related to training data characteristics, *Photogrammetric Engineering & Remote Sensing*, 61(3):391–401.
- Foody, G.M., N.A. Campbell, N.M. Trodd, and T.F. Wood, 1992. Derivation and applications of probabilistic measures of class membership from the maximum likelihood classification, *Photogrammetric Engineering & Remote Sensing*, 58(12):1335–1341.
- Gamba, P., and F. Dell'Acqua, 2003. Improved multiband urban classification using a neuro-fuzzy classifier, *International Journal of Remote Sensing*, 24:827–834.
- Gopal, S., C.E. Woodcock, and A.H. Strahler, 1999. Fuzzy neural network classification of global land cover from a 1° AVHRR data set, *Remote Sensing of Environment*, 67:230–243.
- Gong, P., 1996. Integrated analysis of spatial data from multiple sources: using evidential reasoning and artificial neural network techniques for geological mapping, *Photogrammetric Engineering & Remote Sensing*, 62(4):513–523.
- Green, A.A., M. Berman, P. Switzer, and M.D. Craig, 1988. A transformation for ordering multispectral data in terms of image quality with implications for noise removal, *IEEE Transactions on Geoscience and Remote Sensing*, 26:65–74.
- Hagan, M.T., H.B. Demuth, and M. Beale, 1996. *Neural Network Design*, Boston, Massachusetts, PWS.
- Haykin, S., 1994. *Neural Networks: A Comprehensive Foundation*, New York, Macmillan.
- Harsanyi, J.C., and C.I. Chang, 1994. Hyperspectral image classification and dimensionality reduction: An orthogonal subspace projection approach, *IEEE Transactions on Geoscience and Remote Sensing*, 32:779–785.
- Hewitson, B.C., and R.G. Crane, 1994. *Looks and Use In Neural Nets: Applications in Geography* (B.C. Hewitson and R.G. Crane, editors), Dordrecht/Boston/London, Kluwer Academic, pp. 1–9.
- Jensen, J.R., F. Qiu, and M. Ji, 1999. Predictive modeling of coniferous forest age using statistical and artificial neural network approaches applied to remote sensing data, *International Journal of Remote Sensing*, 20:2805–2822.
- Jensen, J.R., 2005. *Introductory Digital Image Processing*, Third edition, Upper Saddle River, New Jersey, Prentice Hall.

- Joshi, A., S. Weerawarana, N. Ramakrishnan, E.N. Houstis, and R.I. John, 1996. Neuro fuzzy support for problem-solving environments: A step toward automated solution of PDEs, *IEEE Computational Science and Engineering*, 3:44–56.
- Ju, J., E.D. Kolaczyk, and S. Gopal, 2003. Gaussian mixture discrimination analysis and sub-pixel land cover characterization in remote sensing, *Remote Sensing of Environment*, 84:550–560
- Klir, G.J., and B. Yuan, 1995. *Fuzzy Sets and Fuzzy Logic: Theory and Application*, Upper Saddle River, New Jersey, Prentice Hall.
- Kulkarni, A.D., and K. Lulla, 1999. Fuzzy neural network for supervised classification: Multispectral image analysis, *Geocarto International*, 14:41–49.
- Kohonen, T., 1989. *Self-Organization and Associate Memory*, Third edition, New York, Springer-Verlag.
- Kohonen, T., J. Hynninen, J. Kangas, J. Laaksonen, and K. Torkkola, 1996. *LVQ_PAK: The Learning Vector Quantization Program Package*, Technical Report of Helsinki University of Technology, Finland, 26 p.
- Kruse, F.A., A.B. Lefkoff, J.W. Boardman, K.B. Heidebrecht, A.T. Shapiro, P.J. Barloon, and A.F.H. Goetz, 1993. The spectral image processing system (SIPS) interactive visualization and analysis of imaging spectrometer data, *Remote Sensing of Environment*, 44:145–163.
- Landis, J., and G.G. Koch, 1977. The measurement of observer agreement for categorical data, *Biometrics*, 33:159–174.
- Liu, W., S. Gopal, and C. Woodcock, 2001. Spatial data mining for classification, visualization and interpretation with ARTMAP neural network, *Data mining for Scientific and Engineering Applications* (R. Grossman, Editor), Kluwer Academic Publishers, The Netherlands, pp. 205–222.
- Liu, W., K. Seto, E. Wu, S. Gopal, and C. Woodcock, 2004. ART-MMAP: A neural network approach to subpixel classification, *IEEE Transactions on Geoscience and Remote Sensing*, 42:1976–1983.
- Lin, C.T., and C.S.G. Lee, 1996. *Neural Fuzzy Systems: A Neuro-fuzzy Synergism to Intelligent Systems*, Upper Saddle River, New Jersey, Prentice Hall P T R., 797 p.
- Mannan, B., J. Roy, and A.K. Ray, 1998. Fuzzy ARTMAP supervised classification of multi-spectral remotely-sensed images, *International Journal of Remote Sensing*, 19:767–774.
- Mao, J., Y. Wang, and W. Sun, 2002. Remote sensing images classification using fuzzy B-spline function neural network, *Proceedings of the 4th World Congress on Intelligent Control and Automation*, 10–14 June, Shanghai, P.R. China.
- Nauck, D., F. Klawoon, and R. Kruse, 1997. *Foundation of Neuro-Fuzzy Systems*, Chichester, John Wiley & Sons.
- Nomura, T., and T. Miyoshi, 1995. An adaptive rule extraction with the fuzzy self-organizing map: A comparison with other methods, *Proceedings of the 3rd International Symposium on Uncertainty Modeling and Analysis and Annual Conference of the North American Fuzzy Information Processing Society*, IEEE Computer Society, September 17–20, College Park, Maryland, pp. 311–316.
- Openshaw, S., and C. Openshaw, 1997. *Artificial Intelligence in Geography*, Chichester, John Wiley & Sons.
- Penaloza, M.A., and R.W. Welch, 1996. Feature selection for classification of polar regions using a fuzzy expert system, *Remote Sensing Environment*, 58:81–100.
- Qiu, F., and J.R. Jensen, 2004. Opening the black box of neural networks for remote sensing image classification, *International Journal of Remote Sensing*, 25:1749–1768.
- Qu, J., C. Wang, and Z. Wang, 2003. Structure-context based fuzzy neural network approach for automatic target detection, *Proceedings of the IGARSS '03 Geoscience and Remote Sensing Symposium*, IEEE International, 2:767–769.
- Rosenfield, G.H., and K. Fitzpatrick-Lins, 1986. A coefficient of agreement as a measure of thematic classification accuracy, *Photogrammetric Engineering & Remote Sensing*, 52(2):223–227.
- Simpson, P.K., 1992. Fuzzy min-max neural networks – Part 1: Classification, *IEEE Transactions on Neural Networks*, 3:776–786.
- Swayze, G.A., and R.N. Clark, 1995. Spectral identification of minerals using imaging spectrometry data: Evaluating the effects of signal to noise and spectral resolution using the Tricorder algorithm, *Summaries of the 5th Annual JPL Airborne Earth Science Workshop*, JPL Publication 95–1, pp. 157–158.
- Tapia, R., A. Stein, and W. Bijker, 2005. Optimization of sampling schemes for vegetation mapping using fuzzy classification, *Remote Sensing of Environment*, 99:425–433.
- Tompkins, S., J.F. Mustard, C.M. Pieters, and D.W. Forsyth, 1997. Optimization of endmembers for spectral mixture analysis, *Remote Sensing of Environment*, 59:472–489.
- Vuorimaa, P., 1994. Fuzzy self-organizing map, *Fuzzy Sets and Systems*, 66:223–231.
- Wang, F., 1990. Improving remote sensing image analysis through fuzzy information representation, *Photogrammetric Engineering & Remote Sensing*, 56:1163–1169.
- Williamson, J.R., 1996. Gaussian ARTMAP: A neural network for fast incremental learning of noisy multidimensional maps, *Neural Networks*, 9:881–897.
- Woodcock, C.E., and S. Gopal, 2002. Fuzzy set theory and thematic maps: accuracy assessment and area estimation, *International Journal of Geographical Information Science*, 14:153–172.
- Wooster, M.J., T.S. Richards, and K. Kidwell, 1995. NOAA-11 AVHRR/2-thermal channel calibration update, *International Journal of Remote Sensing*, 16:359–363.
- Zadeh, L.A., 1965. Fuzzy sets, *Information and Control*, 8:338–353.
- Zhang, J., and G.M. Foody, 2001. Fully-fuzzy supervised classification of sub-urban land cover from remotely sensed imagery: Statistical and artificial neural network approaches, *International Journal of Remote Sensing*, 22:615–628.