FEATURE-BASED CLASSIFICATION OF LANDSAT - 7 ETM SATELLITE IMAGES

IBRAHIMOVA S.R., BADEA A.*

Azerbaijan National Aerospace Agency (ANASA)

** Romanian Space Agency (ROSA)*

Abstract. In this paper, the suitability of Back Propagation Neural Network (BPNN) for classification of remotely-sensed images is explored. Statistical features are extracted from the first order histogram measures of LANDSAT - 7 ETM images. They are based on the first-order distribution measures such as mean, standard deviation, skew-ness, kurtosis, energy, and entropy. Also, the texture features are extracted from the generalized concurrence matrix. The extracted features are fed to the network's input layer that consists of 120 neurons. Using the known features, the BPNN is trained for five classes of LANDSAT - 7 ETM image. Then, the trained network is used to classify the entire image. Both the proposed and the Maximum Likelihood Classification (MLC) methods are tested on LANDSAT - 7 ETM images.

Introduction. Statistical features representing pixel intensities and their variations are regarded as one of the basic feature groups in specifying image content characteristics which can help in increasing the classification accuracies.

Artificial neural network applications in various fields including signal processing, pattern recognition, medicine, and particularly, the number of reported applications on the use of artificial neural network in remote sensing field has been steadily increased during the last few years. Artificial neural networks can be regarded as highly parallel dynamic systems consisting of multiple simple units that can perform transformation by means of their state response to their input information. How the transformation is carried out depends on the Neural Network (NN) model and its way of learning the transformation. A neural network is presented iteratively with a set of samples, known as training set from which the network can learn the values of its internal parameters.

Feature extraction. Features should be easily computed, robust, insensitive to various distortions and variations in the image. They should also support the discrimination of the land cover/land use classes of remotely sensed data. The two basic feature groups used in this study are statistical features showing the intensities and intensity variations of pixels, and texture features based on gray level concurrence matrix.

Statistical features. The most basic of all image features is some measure of image amplitude in term of luminance, spectral value, or other units. One of the simple ways to extract statistical features in an image is to use the first-order probability distribution of the amplitude of the quantized image. They are generally easy to compute and largely heuristic. The first order histogram estimate of $p(b)$ is simply

$$p(b) = \frac{N(b)}{M} \tag{1}$$

where b is a gray level in an image, M represents the total number of pixels in a neighborhood window centered around the pixel and $N(b)$ is the number of pixels having gray value b in the same window. The following measures have been extracted using first order probability distribution.

Mean as the first feature is the average of pixel intensities within the image window.

$$S_M = \bar{b} = \sum_{b=0}^{L-1} bP(b) \qquad (2)$$

Standard deviation as the second feature is the standard deviation of pixel intensities within the same image window.

$$S_D = \sigma_b = \left[\sum_{b=0}^{L-1} \left(b - \bar{b}\right)^2 P(b) \right]^{1/2} \qquad (3)$$

Skew-ness as the third feature provides information on the shape of the distribution of intensity values within window. The skewness characterizes the degree of asymmetry of the intensity distribution around the mean intensity. If skewness is negative, the data spread out more to the left of the mean than to the right and vice versa.

$$S_S = \frac{1}{\sigma_b^3} \sum_{b=0}^{L-1} \left(b - \bar{b}\right)^3 P(b) \qquad (4)$$

Kurtosis as the fourth feature also provides information on the shape of the distribution of intensity values within window. Kurtosis measures the relative peakness or flatness of the intensity distribution relative to the normal distribution. The kurtosis of the normal distribution is 3. Distributions that are more outlier-prone than the normal distribution have kurtosis greater than 3.

$$S_K = \frac{1}{\sigma_b^4} \sum_{b=0}^{L-1} \left(b - \bar{b}\right)^4 P(b) - 3 \qquad (5)$$

Energy as the fifth feature is useful to examine the power content (repeated transitions) in a certain frequency band.

$$S_N = \sum_{b=0}^{L-1} \left[P(b)\right]^2 \qquad (6)$$

Entropy as the final feature is a common concept in many fields, mainly in signal processing .

$$S_E = -\sum_{b=0}^{L-1} P(b) \log_2 \{P(b)\} \qquad (7)$$

Texture features. Many land cover/land use classes can be distinguished from their characteristics on shape or structure. Therefore it is important to extract features which are able to describe relevant texture properties of the classes. Different kinds of texture features such as multi channel filtering feature, fractal based feature and features extracted from concurrence matrix have already been proposed [5]. In this study, the cooccurrence features are used as the basic texture feature detectors due to their good performance in many pattern recognition applications in remote sensing. When computing concurrence features from images with relatively great number of possible pixel intensity values, it is not suitable to use all possible discrete signal levels [2]. If all original intensity levels were employed, the derived texture information could be easily blurred by a noise in the image. Hence it is preferable to transform the original intensity values into smaller number of possible levels either by a scalar or vector quantization method. In this case, the original image pixels intensities are transformed from 256 to 32 different discrete values [1]. A gray level concurrence matrix is defined as a sample of the joint probability density of the gray levels of two pixels separated by a given displacement. In Cartesian coordinates the displacement of the concurrence can be regarded as a vector (Δx, Δy). The gray level concurrence matrix N is defined as:

$$N(i,j) = \{\#\ of\ pair\ (i,j) \mid image(x,y) = i \text{ AND } image(x+\Delta x, y+\Delta y) = j\}$$

where *i and j* are gray levels. In this study, Δx=0,2,2 and Δy=2,2,0 have respectively been used to generate the generalized co-occurrence matrix.

Second-order histogram features are based on the gray level co-occurrence matrix. The histogram estimate of the second-order distribution is

$$P(a,b) \cong \frac{N(a,b)}{M} \tag{8}$$

where M is the total number of pixels in the measurement window and $N(a,b)$ is produced from cooccurrence matrix. If the pixel pairs within an image are highly correlated, the entries in $P(a,b)$ will be clustered diagonally in the matrix. Various measures may be used as measures that specify the energy spread around the diagonal of $P(a,b)$.

$$S_A = \sum_{a=0}^{L-1}\sum_{b=0}^{L-1} abP(a,b) \tag{9}$$

$$S_C = \sum_{a=0}^{L-1}\sum_{b=0}^{L-1} \left(a-\bar{a}\right)\left(b-\bar{b}\right)P(a,b) \tag{10}$$

where

$$\bar{a} = \sum_{a=0}^{L-1}\sum_{b=0}^{L-1} aP(a,b) \qquad \text{and} \qquad \bar{b} = \sum_{a=0}^{L-1}\sum_{b=0}^{L-1} bP(a,b)$$

$$S_I = \sum_{a=0}^{L-1}\sum_{b=0}^{L-1} (a-b)^2 P(a,b) \tag{11}$$

$$S_V = \sum_{a=0}^{L-1}\sum_{b=0}^{L-1} |a-b|\; P(a,b) \tag{12}$$

$$S_F = \sum_{a=0}^{L-1}\sum_{b=0}^{L-1} \frac{P(a,b)}{1+(a-b)^2} \tag{13}$$

$$S_G = \sum_{a=0}^{L-1}\sum_{b=0}^{L-1} \left[P(a,b)\right]^2 \tag{14}$$

$$S_T = -\sum_{a=0}^{L-1}\sum_{b=0}^{L-1} P(a,b)\log_2\left\{P(a,b)\right\} \tag{15}$$

In this study, measures such as autocorrelation (9), covariance (10), inertia (11), absolute value (12), inverse difference (13), energy (14) and entropy (15) have been used to specify texture features.

Self-organized learning. Self-organized learning or unsupervised learning paradigms are forms of the cluster analysis. Clustering generally describes a regression solution that attempt to optimally partition an input space of dataset of N elements into a compact representative set of K cluster centers, where $K<<N$. For example the input space data may represent the classified pixels in an image, and the cluster represent color image segments.

Self-organized maps. The self-organizing neural network, also known as Kohonen network, is a network that incorporates a topology scheme, i.e., it takes into account the topological structure among neurons. The input signals are n-tuples. There is a set of m cluster neurons. Each input neuron is fully connected to all output neurons which response differently to the input pattern. At the time of each input in the training phase, the cluster neuron with weights that best match the input pattern is selected the winner (usually in a minimum Euclidean distance sense). This winning neuron and the neighborhood around it are then updated in such a way that their internal weights be closer to the presented input. The adopted updating factor is not equal for all neurons, but stronger near the winning neuron, and decreasing for more distant neurons. Figure 1 shows the basic structure of self-organizing maps [7].

There are input components (white circles) connected to all cluster neurons (shaded circles). The neurons can assume any spatial distribution, which are usually linear or planar arrays. Weights are associated to each connection. With time, the gain factor must be reduced and also the neighborhood decrease in size. During the learning phase the node weights are changed in an ordered manner, in such a way that the main image features tend to be organized according to topological distribution in the network. Adjacent nodes response similarly, while distant nodes respond diversely. The convergence of the feature in the self-organizing map occurs considering some limitations on the gain factor while updating the weights.

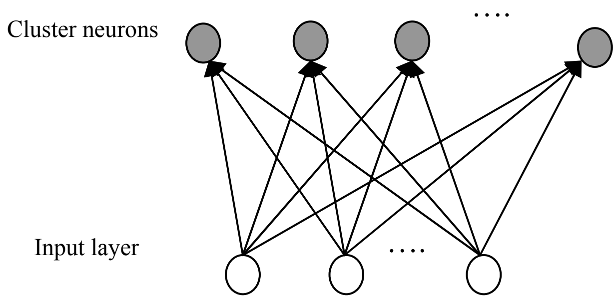

Fig. 1. Self-organizing map

Classification. Classification is the process of sorting pixels into a finite number of individual classes, or categories of data, based on their values. If a pixel satisfies a certain set of criteria, then the pixel is assigned to the class that corresponds to that criterion. The most widely used neural classifier is Multi Layer Perceptron (MLP) network. The MLP neural network model consists of a network of processing elements or nodes arranged in the layers. Typically it requires three or more layers of processing nodes: an input layer which accepts the input variables used in the classifier procedure, one or more hidden layers, and an output layer with one node per class.

In this study, a three-layer network has been selected with 120 neurons in input layer, 40 neurons in hidden layer and 5 neurons in output layer. The number of neurons in the input layer depends on the features vector. The number of neurons in the output layer depends on the number of classes.

There are several training algorithms for feed forward networks. These algorithms use gradient of the performance function to determine how to adjust the weights to minimized performance. The gradient is determined using back propagation technique which involves performing computational backwards through the network.

Resilient Back propagation training algorithm has been used in this study. It is a training algorithm that eliminates the harmful effect of having a small slope at the extreme ends of the sigmoid "squashing" transfer functions [3,6]. It is faster and more accurate than the standard back propagation algorithm for training. It can converge from ten to one hundred times faster than the standard algorithm using delta rules. This algorithm operates in the batch mode and is invoked using train.

Implementation and results. The flowchart of the classification algorithm has been shown in the figure 2. In order to implement the algorithm, LANDSAT - 7 ETM satellite images from a region in the north of Iran have been used. The proposed segmentation utilizes a self-organizing map to detect the main features present in the image. The features are represented by their chromaticity values which express colors hue and saturation, avoiding the luminance component.

In the Final step, the extracted statistical and texture features along with the intensity of the central pixel of each window are arranged in a vector which is then fed to the MLP network. The MLP identifies the class code of each segment.

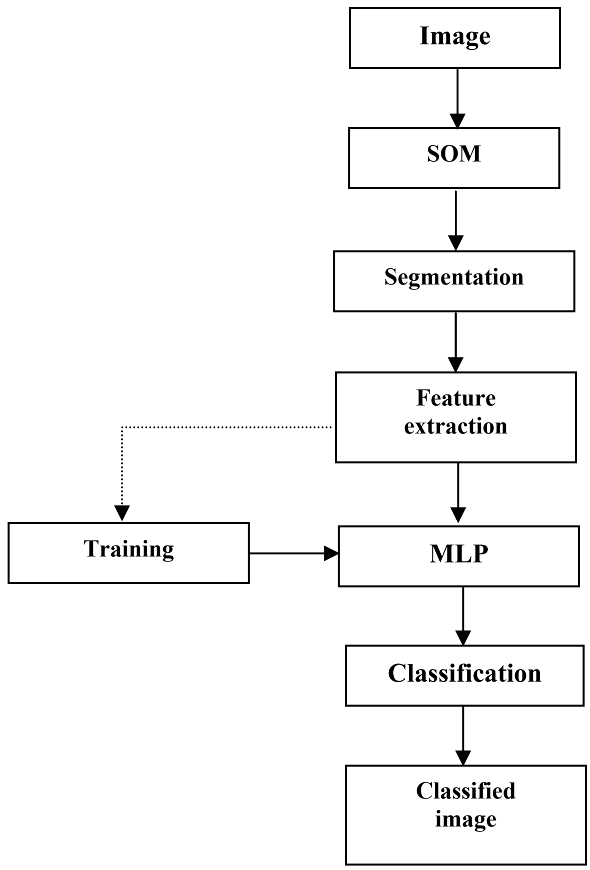

Fig. 2. The flowchart of land covers classification using ANN

Implementation results have been compared with Maximum Likelihood Classification (MLC) method which is one of the conventional classification methods based on statistical analysis of the image. The maximum likelihood decision rule is based on the probability of a pixel belonging to a particular class. The basic equation assumes that the probabilities are equal for all classes, and that the input bands have normal distributions.

Conclusions. Based on the comparison made between the classification method implemented in this study and the MLC on LANDSAT - 7 ETM images, it has been shown that neural network method is more accurate than MLC method. Also, it has been concluded that the neural network method is more sensitive to the training site selections than MLC. The segmentation before classification has shown that the classified image presents more clear edges than MLC method. Obviously, classes that have not been trained for are supposed to be labeled unclassified, but using MLC method, there are no unclassified pixels left which may be regarded as a disadvantage of the MLC method for its misclassifications. It should be noted that in order to reduce processing time and data redundancy in features, it has been necessary to do segmentation before feature extraction. Resilient back propagation is generally much faster than the standard steepest descent algorithm. It also has a useful property that requires only a modest increase in memory requirements. Therefore, it has not been necessary to store the updated values for each weight and bias, which is equivalent to storage of the gradient.

1. *Pratt, W. K.,* 2001, Digital image processing, John Willy & Sons, Inc., 3nd ed.
2. *Ohanian, P.P. and R.C. Dubes,* 1992, Performance Evaluation for four classes of textural features, *Pattern recognition*, 25 (8), pp. 819-833
3. *Ibrahimova S.R.* Development of new mathematical methods for increase of accuracy rasters transformation. Reports National Academy of Azerbaijan.. Institute of Cybernetics, Institute of Information Technologies. Baku, 2003, pp.61-64.
4. *Ibrahimova S.R.* Methods of remote sounding of the ground with application new information technologies. Reports National Academy of Azerbaijan. 2002, v. XXII, №2-3, pp. 168-173.
5. *Ibrahimova S.R. The Mathematical Modeling of Sustainable Development of Environmentally Hazardous Area.* ISNET Journal of Space Sciences and Technology. Vol.1., No.1, 2004.
6. *Riedmiller, M., and H. Braun,* 1993, A direct adaptive method for faster back propagation learning, The RPROP algorithm, Proceedings of the IEEE International Conference on Neural Networks, San Francisco.
7. *Yin, H. and Allinson, N.M.,* 1995, On the Distribution and Convergence of Feature Space in Self-Organizing Maps, Neural Computation, 7, pp. 1178-1187.

LANDSAT - 7 ETM PEYK TƏSVİRİNİN NEYRON ŞƏBƏKƏNİN TƏTBİQİ İLƏ SİNİFLƏŞDİRİLMƏSİ

İBRAHİMOVA S.R., BADEA A.

Məqalədə neyron şəbəkənin köməyi ilə LANDSAT - 7 ETM peyk təsvirinin sinifləşdirilməsinə baxılır. Şəbəkənin girişinə 120 neyron yerləşdirilir. Şəbəkə 5 sinif üzrə öyrədilir və bu təsvirin tam sinifləşməsinə imkan verir.

КЛАССИФИКАЦИЯ LANDSAT - 7 ETM СПУТНИКОВЫХ ИЗОБРАЖЕНИЙ С ПРИМЕНЕНИЕМ НЕЙРОННОЙ СЕТИ

ИБРАГИМОВА С.Р., БАДЕА А.

В работе исследуется обратно распространенная нейронная сеть для классификации изображений. Статистические данные извлекаются от LANDSAT - 7 ETM. Извлеченные данные помещаются на входе сети, который состоит из 120 нейронов. Сеть обучена для пяти классов LANDSAT - 7 ETM изображения. Далее обученная сеть используется для классифицирования полного изображения.