Remote sensing images classifications based on decision fusion

Behzad Moshiri and Farhad Besharati
Control and Intelligent processing Center of Excellence
Electrical and Computer Engineering Department
University of Tehran, P.O. Box 14395/515, Tehran, IRAN

***Abstract*-With developing of remote sensing technology and creatation of new various imaging sensors it has been possible to access to various images from one scene. Because of some reasons, such as limited number of training samples, the results of classical classification methods used for these images are not satisfactory. A multiclassifier system is used for classification of these images. In this system, decision fusion strategy is utilized which is the highest level of data fusion approach. By a proper mechanism proposed in this paper, using of one multiclassifier system for classification of extracted data from one source has been available. For this purpose, initially, existed bands are categorized based on minimum correlation criterion and before primary classification in multiclassifier system and data of each group are used for primary classification careless to imaging sensors. In final classification stage, the results of source's primary classification are combined according to different statistical and neural algorithms and final decision for one pixel's class is made. Morever, the effect of using multi classifier in primary classification stage is investigated. This procedure is implemented on real data and the results are analyzed qualifiedly and quantitively using various criterion such as accuracy degree and classification provided map. The results show that this classification approach and specially using various classifiers in primary classification stage, can compensate limited amount of training samples in remote sensing data.**

***Keywords:* Decision fusion, multiclassifier systems, limited training samples remote sensing**

1. Introduction

Each sensor used in remote sensing for imaging has some bands. The reflected waves from earth's surface are registered in these bands and in the other words, surface spectrum of the bands is reconstructed by sensor. The more the number of the imaging bands, the closer reconstructed spectrum obtained related to real one and more information will be accessible from the surface [2]. One way to get more reliable data and decision accuracy in earth's surface covers is the using of provided images simultaneously by different imaging systems. Obtaining of these images from one scene on earth's surface has been facilitated by progressing of remote sensing technology. Each sensor has its exclusive properties. Some of sensors have good location accuracy and the others have proper spectrum accuracy. Using the advantages of different sensors can help decisioner to make accurate and reliable decision. Multisensor systems have a considerable capabilities compared to unisensor systems [3]:

- In these systems, the probability of unobserved events has reduced, e.g., observation region expands.
- Increased number of sensors and observations causes separation of observation or testing regions.
- The noise can be decreased effectively assuming its independency in different observations.
- In multisensor systems, the complexity and cost can be significantly low compared to the similar unisensor system to access the desired quality, accuracy and precision.

The resultant of such imaging systems is a set of images. To interpret or classify of these images, following problems must be considered:

- Maybe different sources have different natures. So, they can not be expressed by a proper multivariate statistical model.
- Different data sources have different levels of validity degree. So, it is better to weight different sources according to their validations for classification.
- A minimum number of training samples is required for classification of these images based on classical approaches (for example maximum likelihood method). Providing this minimum number may be impossible or complicated and expensive.
- If the size of investigated image were large, then using of classical methods can be very time consuming process. One of the most important methods for classification of mentioned data is the using of multiclassifier systems [4,5,6]. The followed approach in this paper is using of a multiclassifier system careless to imaging sensors in which data fusion technology is utilized. By applying the proposed system, one can use the information of different sensors together of several parallel processing systems. Therefore the required time of classification is reduced and all the capabilities of multiclassifier systems can be served to classify the sensor's information. The first considered case in data fusion is matching information. For this purpose, information is brought to the same scale and same ordinate regard to recognition unit, location coordinator and other cases. This is the first and one of the most important steps in data fusion, because if a problem exists in this matching, the results of fusion's mathematical models will not be true. In this paper, it is assumed that required preprocessing is applied on images. The other subject including in data fusion is injecting the resulted information from different data sources. Finally, a data fusion system is perfected by preparing a fusion model. This model depends on specified application and is defined according to information nature. Having auxiliary data in a data fusion system can be very important which improve fusion process. General diagram of a data fusion system is shown in Fig. 1.

2. Stated approach to classify data with limited training samples

It is assumed that one scene is imaged in N bands regardless of imaging sensor. Initially presented data is classified based on minimum correlation criteria and then each group's bands are considered as the bands of a new source and primarily are classified. To investigate the effect of using different classifiers in primary classification stage, implementations are accomplished in three cases as follows:

In first case, only maximum likelihood statistical classifiers are used for primary classification in each source.

In second case, only three-layer perceptron neural network with error back-propagation algorithms is used as sources primary classification.

In third case both three-layer perceptron neural network trained by error-back propagation algorithm and maximum likelihood classifiers is used, simultaneously. In each case, primary decision is combined in a decision fusion center for final decision pursuing. Algorithms established on different statistical techniques are used for primary decision fusion.

2.1. Maximum likelihood classifiers

In these classifiers, it is assumed that the probability distributions for the classes are of the form of normal models. However this is an assumption and one can not prove it as feature for all information and spectral classes.

This assumption makes mathematical calculation of the problem easier. Suppose that we have N spectral bands, therefore, the following relation would be true:

$$P(x|w_i) = (2\pi)^{-N/2} \left|C_{xi}\right|^{-1/2} \exp\left\{-\frac{1}{2}(x-m_i)^t C_{xi}^{-1}(x-m_i)\right\} \quad (1)$$

Where x is column vector of brightness value for the pixel. $P(x|w_i)$, m_i and C_{xi} are the posterior probability, mean vector and covariance matrix of class w_i, respectively obtained according to training data as:

$$m_i = E\{xi\} = \frac{1}{K}\sum_{k=1}^{K} xi_k$$
$$C_{xi} = E\{(xi-m_i)(xi-m_i)^t\} = \frac{1}{K-1}\sum_{k=1}^{K}(xi_k - m_i)(xi_k - m_i)^t \quad (2)$$

$(2\pi)^{-N/2}$ factor in equation (1) is a common term in all $g_i(x)$'s which has not any effect in their comparison. It can be omitted and hence final form of descriminant function is stated as:

$$g_i(x) = \ln P(w_i) - \frac{1}{2}\ln\left|C_{xi}\right| - \frac{1}{2}(x-m_i)^t C_{xi}^{-1}(x-m_i) \quad (3)$$

Implementation of this classifier is accomplished by using relation (3). The class with maximum descriminant function is selected as a sample class.

2.2. Neural network classifiers

These classifiers are nonparametric which do not need to any presumption about classes. Three-layer neural back-propagation perceptron neural network is used for image classification. Converting of pixel brightness depends on one of the preprocessing's gray code done on data. These preprocessing is performed because network training with gray code input is easier than network training with brightness level input. So, a network with gray code input were designed and used.

2.3. Evaluation criteria

After classification of training samples, two parameters named accuracy degree and α, are used to evaluate the classifier performance and is defined as follow:

$$\alpha = \frac{a}{A} * 100 \quad (4)$$

In above equation, "a" and "A" are the correct classified number of training samples and the total number of test pixels related to considered class, respectively. Total accuracy degree is defined as (5). In this equation, M is the total number of image classes.

$$T\alpha = \frac{\sum_{i=1}^{M} a_i}{\sum_{i=1}^{M} A_i} * 100 \quad (5)$$

3. Decision fusion algorithms

First algorithm: arithmetic mean

In this method, the arithmetic mean of calculated posterior probabilities is computed by primary classifiers. Descriminant function of this diffusion rule is as the following form [6]:

$$C_j(X) = \sum_{i=1}^{N} p(w_j / x_i) \quad (8)$$

In this equation N is the number of primary classifiers and w_j shows the class. The vectors $x_1, \ldots, x_N$ are the input data vectors and X is the pixel. $P(w_j|x_i)$ is the posterior probability of class w_j for input vector of i'th source. The class for which C is largest selected as the class of pixel.

Second algorithm: geometric mean

In this fusion, the geometric mean of posterior probabilities is calculated. The related equation in this method is as the following form [6]:

$$F_j(X) = \prod_{i=1}^{N} p(w_j | x_i) \quad (9)$$

Parameters of this equation are the same as ones used in equation (8). The considered pixel is assigned to the class that its F is largest.

Third algorithms: fusion rule based on maximum probability

If the number of primary classifications is N and that's of classes is M, after primary classification for each sample, a N×M matrix is created from $P_{X/w}$ probabilities as (10):

$$P_{X/w} = \begin{bmatrix} P(x_1/w_1) & P(x_1/w_2) & .... & P(x_1/w_M) \\ P(x_2/w_1) & P(x_2/w_2) & .... & P(x_2/w_M) \\ . & . & & . \\ . & . & .... & . \\ . & . & & . \\ P(x_N/w_1) & P(x_N/w_2) & .... & P(x_N/w_M) \end{bmatrix} \quad (10)$$

In previous fusion algorithms, all of the posterior probabilities are used to determine the sample class; descriminant functions were calculated and a decision was made based on them. But in this method, final decision is made only by comparing calculated posterior probabilities. Corresponding equation to this rule is as:

$$\max_{i=1}^{N} p(x_i/w_j) = \max_{k=1}^{M} \max_{i=1}^{N} \left[p(x_i/w_k)\right] \quad (11)$$

The probabilities matrix $P_{X/w}$ of eq. (10) is calculated for each pixel. Then, these matrixes are investigated and specified in a element with maximum value. The column number of this maximum element specifies the class of desired pixel.

Fourth algorithm: Fusion rule based on minimum probability
The form of minimum based fusion rule is:

$$\min_{i=1}^{N} p(x_i/w_j) = \max_{k=1}^{M} \min_{i=1}^{N} \left[p(x_i/w_k)\right] \quad (12)$$

The performance of this rule can be explained as follows:
$P_{X/w}$ probability matrixes is searched and the minimum in each column is specified. Then between these minimum elements, an element with maximum value is determined. The column number of maximum element determines the pixel's class.

4. Experiment

To evaluate the performance of different fusion algorithms in this paper, a test image with 80×120 pixels and12 bands is provided. It has extracted from an agricultural segment of Indiana State and has 256 gray levels consisting of 8 classes. Two classes of image have 2400 pixels and the rest have 800. The wavelength bound of the image's bands is presented in table I. To implement all of categorized methods described in this paper, 18 training samples per each class are used. These training regions are specified by brown color in Fig 2a. Initially, all of 12 bands are classified by applying the maximum likelihood and three-layer perceptron neural network method with error back-propagation training algorithm. The corresponding results of this classification are illustrated in table III. In neural network classification, it is required to 96 neurons at input layer since the gray code is used. The number of neurons at output layer is 8 because we have 8 classes.

The maximum and minimum correlation criteria are used for band classifications and in both cases, three fourth-band groups are created. These groups are introduced in table II. In continue, these groups are classified primarily. Some experiments are done to investigate the effect of many numbers of primary classifiers. In first case, only maximum likelihood classifiers are used. In second case, only neural network classifiers are used and for the third case, both mentioned classifiers are utilized, simultaneously as the primary classifiers. In primary neural network classifications, because of using the gray code and having 8 classes, 32 and 8 neurons are needed at input and output layers, respectively. In intermediate layer, 16 neurons are used.

5. Results

Investigation of presented results in table III indicates that statistical moments of classes in parametric classifiers -such as maximum likelihood- are estimated with an error, despite to limited training data. Consequently, the classification accuracy is reduced. In these conditions, nonparametric classification methods -such as neural network- represent better accuracy. The results of utilizing different band's classification criteria and decision fusion algorithms to classify the final image classifications are shown in table IV. Obviously, using the minimum correlation criterion for band's classification indicates better results. Since by this criterion, the bands selected for one group have less overlapping and consequently, presented information in these groups are more than the selected groups with maximum correlation criterion method. So, it is made a more accurate decision in primary decision and finally, the using of fusion algorithms represents more accurate results. In general, classification criteria are important and have considerable effect on the results. So, methods that can be independent of classification criteria will be very useful and advantageous. The results of using the multiple primary classifiers are presented in table IV. Investigation of the results shows an improved final classification with simultaneous using the maximum likelihood and neural network classifiers because of utilizing the advantages of both classifiers. The comparison between different classification schemes -shown in table III and IV- indicates that compensation of limited training samples considerably would be possible by using the proper criterion for band's classification.

References

[1] Richards A. John, and Xiuping Jia, *Remote sensing digital image analysis*, Third edition, Springer-Verlag Berlin Heidelberg, Printed in Germany, 1999.

[2] D. L. Hall and J. Llinas, *Handbook of multisensor data fusion,* CRC Press, 2001.

[3] Giorgio Giacinto and Fabio Roli, "An approach to the automatic design of multiple classifier Systems", ELSEVIER, Pattern Recognition Letters, vol. 22, pp.25-33, 2001.

[4] Xue-Hua Liu, A.K. Skidmore and H. Van Oosten, "Integration of classification methods for improvement of land-cover map accuracy", ISPRS Journal of Photogrammetry & Remote Sensing, vol.56, pp. 257–268, 2002.

[5] Giorgio Giacinto, Fabio Roli and Lorenzo Bruzzone, "Combination of neural and statistical algorithms for supervised classifcation of remote-sensing images", ELSEVIER, Pattern Recognition Letters, vol.21, pp. 385-397, 2000.

[6] Lui O. Jimenez, Anibal M. Morell and A. Creus, "Classification of hyperdimensional data based on feature and decision fusion approaches using projection pursuit, majority voting and neural network", IEEE Trans. Geosci. Remote Sens., vol. 37, no.3, pp.1360-1366, May 1999.

[7] Bassel Solaiman . "Multisensor Data Fusion Using Fuzzy Concepts", IEEE Trans. Geosci. Remote Sens., vol. 37, no.3, May 1999.

[8] Jon A. Benediktsson and Ioannis Kanellopoulos, "Classification of multisource and hyperspecrtal data based on decision fusion", IEEE Trans. Geosci. Remote Sens., vol. 37, no.3, pp.1367-1377, May 1999.

[9] L. Bruzzone, D. F. Pirto and S. B. Serpico, "A neural-statistical approach to multitemporal and multisource remote sensing image classification", IEEE Trans. Geosci. Remote Sens., vol. 37, no.3, pp.1350-1359, May 1999.

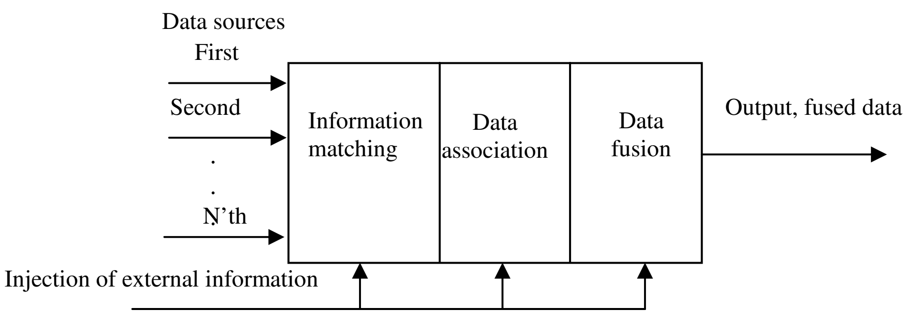

Fig.1. Block diagram of a data fusion system.

Table I. Wavelenght bands of test image.

Wavelenght bands (μm)	Band	Wavelenght bands (μm)	Band	Wavelenght bands (μm)	Band
1.40–1.00	9	0.60–0.54	5	0.49–0.46	1
1.80–1.50	10	0.65–0.58	6	0.51–0.48	2
2.60–2.00	11	0.70–0.61	7	0.54–0.50	3
11.70–9.30	12	0.92–0.92	8	0.57–0.52	4

Table 2. The results of maximum and minimum criterion classification .

Maximum correlation criteria		Minimum correlation criteria	
7,3,2,1	First group bands	10 ,8,6,1	First group bands
12,6,5,4	Second group bands	12,11 ,4,2	Second group bands
11,10,9,8	Third group bands	9,7,5,3	Third group bands

Table 3. The classification results of 12 bands using maximum likelihood and neural network methods.

Total accuracy degree	Classification method
76.77	Maximum likelihood
81.20	Neural network

Table 4. The comparison of resultant accuracy using multiple classifiers

Simultaneous using		Neural network		Maximum likelihood		algorithm
Minimum correlation	Maximum correlation	Minimum correlation	Maximum correlation	Minimum correlation	Maximum correlation	
88.45	82.58	84.09	76.04	84.17	74.80	First
89.99	83.45	84.86	79.16	85.86	77.01	Second
81.19	67.47	80.05	66.17	82.38	68.81	Third
87.45	78.71	82.88	75.18	85.83	75.67	Fourth

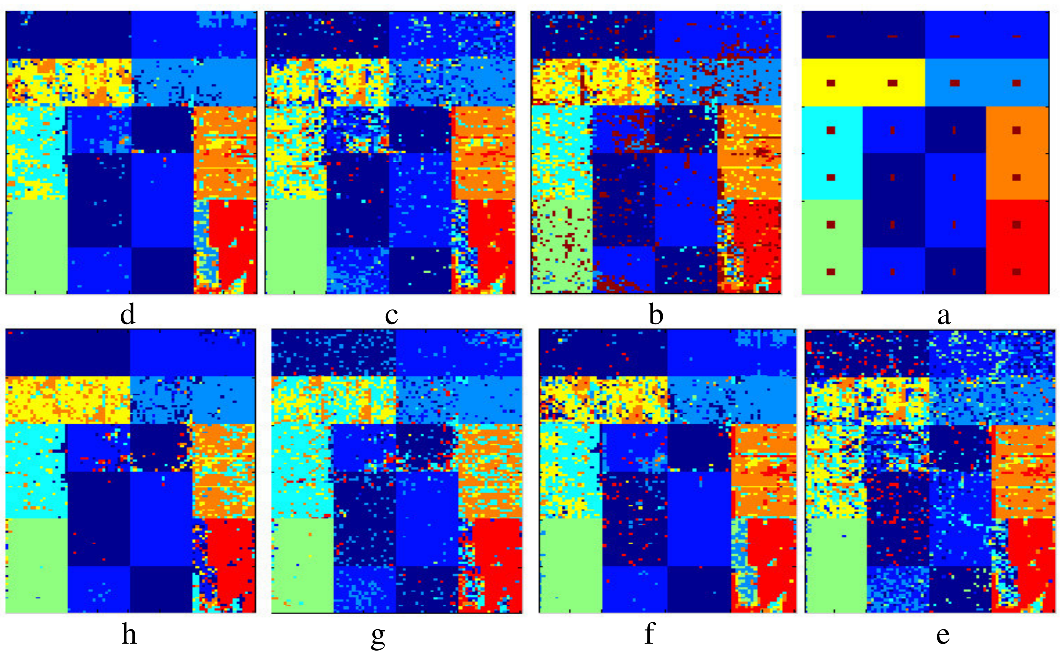

Fig. 2. (a) The reference map of test image with training regions. Class map produced by classification of 12 bands with (b) maximum likelihood (c) first algorithm of maximum correlation criterion. (d) first algorithm of minimum correlation criterion (e)third algorithm of maximum correlation criterion (f)third algorithm of minimum correlation criterion (g) second algorithm of maximum correlation criterion (h) second algorithm of minimum correlation criterion.