A COMPARATIVE STUDY OF STATISTICAL AND NEURAL METHODS FOR REMOTE-SENSING IMAGE CLASSIFICATION AND DECISION FUSION

Safaa Mahmoud
National Authority of Remote Sensing and Space Science, Egypt
safaam_ahmed@yahoo.co.uk

Moumen T. El-Melegy
Electrical Engineering Department, Assiut University, Assiut 71516, Egypt
moumen@aun.edu.eg

Aly A. Farag
CVIP Lab, ECE Dept. University of Louisville, Louisville, KY 40292,USA
farag@cvip.uofl.edu

ABSTRACT

This paper focuses on evaluating a number of statistical and neural methods for supervised, pixel-wise remote-sensing image classification and decision fusion. Despite the enormous progress in the analysis of remote sensing imagery over the past three decades, still much is desired in the area of image classification as no specific algorithm is known to provide accurate results under all circumstances. Decision fusion may be pursued to combine the outputs of different classifiers applied on the same data, in the hope of combining the best of what each approach provides. We report the results of the comparison between several classification and fusion methods on two real datasets, one of which is the standard benchmark Satimage dataset. It is shown that the fusion approaches can indeed outperform the performance of the best classifier.

1. INTRODUCTION

Classification approaches based on statistical and neural networks have been applied successfully in remote sensing. The advantage of neural network classifiers over supervised statistical approaches is that neural networks need no priori statistical information about the input image. This is especially important for multispectral image, since it is very difficult to model the whole image by statistical methods. However, when a sufficiently accurate multivariate statistical model can be determined, statistical method should outperform neural networks in term of classification accuracies. Based on this, it is of interest to use statistical and neural network approaches together in the classification of remote sensing images and to investigate if higher accuracies can be achieved by using their combination. A decision fusion methodology is employed to combine the results of multiple classifiers. Decision fusion has been employed in recent years to increase the accuracy in classification of multispectral images beyond the level achieved by individual classifiers [1], [2], [3], [4].

In this paper, several experiments are conducted on supervised, pixel-wise image classification and decision fusion. A number of statistical and neural approaches have been used and compared on two real datasets, one of which is a standard benchmark set. There have been some other comparisons reported in the literature, e.g., [2]-[4], [7], [9]. However, these earlier efforts focused on statistical methods along with a few neural approaches, or vice versa. Our goal is to compare several major approaches from both families for both classification and fusion. Due to limited space, we give here a comparison between three statistical classifiers: Supervised Parametric Bayes, Non-parametric Bayesian Classifier using the Parzen density estimate, and k-nearest neighbors (k-NN), and three neural network classifiers: Multilayer Perceptron (MLP) with two one and two hidden layers, Radial Basis Function (RBF) networks; and Probabilistic Neural Networks (PNN). We also evaluate several fusion approaches: voting, Bayesian average, maximum and median rules, and neural networks.

2. CLASSIFICATION METHODS

2.1 Statistical approaches

In our investigation three statistical approaches were used. The following is a brief description of each approach.

Let $w_1 \dots w_C$ be the finite set of C classes for an image scene. The probability $P(w_i/\boldsymbol{x})$ gives the likelihood that the correct class is w_i for the d-dimensional feature vector $\boldsymbol{x}$. There are two issues to be considered. The first is the a *priori* probability of each class. Fortunately, this issue is not a critical matter since it can be estimated from the design data set or it can be assumed to be equal for all classes. The second and major issue is to estimate the class conditional probability $P(w_i/\boldsymbol{x})$, for each class. Towards that goal, two main directions are usually considered: parametric and non-parametric estimation [5].

0-7803-8554-3/04/$20.00 ©2004 IEEE.

Parametric Bayes Classifier
In this approach, an a priori form of the class conditional density $p(\boldsymbol{x}/w_i)$ is assumed; the parameters in this density are to be estimated. The design data are used to estimate these parameters. In the Gaussian case, which we assumed in our implementation, the only parameters needed to describe $P(w_i/\boldsymbol{x})$ are the *covariance* matrix Σ and the *mean* vector $\boldsymbol{\mu}$. There are many approaches in the literature to estimate these two parameters. One of the most common approaches is the maximum likelihood estimator [5], which is used in our implementation.

Non-Parametric Estimation using Parzen Window
In the non-parametric approach, no a priori structural form is assumed for $P(w_i/\boldsymbol{x})$. The Parzen window approach estimates $P(w_i/\boldsymbol{x})$ for any $\boldsymbol{x}$ using the number of samples in a hypercube of dimension h around $\boldsymbol{x}$.

Non-Parametric Estimation using k-Nearest Neighbor
The approach directly estimates the a posteriori probabilities [5]. The k-Nearest neighbor rule classifies a point $\boldsymbol{x}$ by assigning it to the class that is most frequently represented among the k nearest samples.

2.2 Neural networks approaches

Multi-Layer Perceptons (MLP)
Feed-forward, multi-layer percepton networks are one of the most popular designs for neural networks. Network architecture or topology plays an important role for the neural network classifier; Setting the number of hidden layers and their neurons often requires some experimentation. The number of neurons in the input and output layers is given by the feature space dimension d and number of classes C. We set the output layer transfer function to exponential in order that the net output is representative for the posteriori.

Radial Basis Functions neural network (RBF)
A radial basis function network has linear transfer function at the output units. When an input feature vector $\boldsymbol{x}$ is presented to the net, the first layer computes distances from the input vector to the training input vectors, and produces a vector whose elements indicate how close the input is to a training input. The second layer sums these contributions for each class of inputs to produce as its net output a vector of posteriori probabilities.

Probabilistic neural network (PNN).
The PNN operates similar to RBF network except that the second layer is a competitive layer.

3. DECISION FUSION ALGORITHMS

In our study we used three decision fusion schemes, namely the majority vote, statistical rules and, neural networks. Typically, decision fusion involves a method that combines the outputs of L classifiers into a single one.

3.1 Fusion by voting schemes

Voting schemes can be applied to the formed set of multiple classifiers assuming that each classifier gives a single class label as an output [1], [3], [4]. In the *majority-voting* rule a sample is assigned the class, for which there is a consensus among all classifiers, otherwise the sample is rejected. However, in the c*omplete agreement* mode, when *all* the individual classifiers agree on the classification of a sample, the sample is classified to that class, otherwise, the sample is rejected.

3.2 Fusion by statistical rules

The largest group of classifier fusion methods operates on classifiers that produce so-called soft outputs. These outputs can be used to describe decision uncertainty. Effectively, the fusion methods in this group try to reduce the level of uncertainty maximizing suitable measures of evidence. In the following, we assume that for each input sample, the classifiers are able to provide an estimation of the posterior probability $P(w_i/\boldsymbol{x})$.

Bayesian Average
Assign $\boldsymbol{x}$ to w_i where $P_{av}(w_i/\boldsymbol{x})$ as defined below is maximum:

$$P_{av}(w_i / x) = \frac{1}{L}\sum_{k=1}^{L} P_k(w_i / x) \tag{1}$$

Maximum Rule
Assign $\boldsymbol{x}$ to w_i where $P_{max}(w_i/\boldsymbol{x})$ is maximum:

$$P_{\max}(w_i / x) = \max_{k=1}^{L} P_k(w_i / x) \tag{2}$$

Median Rule
Assign $\boldsymbol{x}$ to w_i where $P_{med}(w_i/\boldsymbol{x})$ is maximum:

$$P_{med}(w_i / x) = med_{k=1}^{L} P_k(w_i / x) \tag{3}$$

3.3 Fusion by neural network schemes

Fusion can be based on supervised schemes using the output of the formed set of multiple classifiers as the new features. It will apply the backpropagation (BP) algorithm on the training samples for a second learning process to estimate the required parameters (network weights) for the merging rule [8], [9]. Unlike [8] that used classifiers' hard outputs (i.e., class labels) as input to the MLP, we use, similar to [9], the classifiers soft outputs (i.e., posteriori probabilities) as inputs.

4. EXPERIMENTAL RESULTS

The goal here is to compare the performance of the previous classification and fusion approaches, and to show if the fusion approaches are able to improve on the classification accuracy.

For each classifier, a careful designing phase was carried out in order to assess the best performances provided by single classifiers. For the *k*-nearest neighbor classifier, we carried out different trials with different values for *k*. For the MLP, over 80 different topologies with one or two hidden layers were tried. For the Parzen window approach we carried out different trials with *h*. The Bayes classifier, Radial Base Network, and the Probabilistic Neural Networks need no/minimal designing phases. All classifiers were trained on the same training data set. The Accuracy (AC) of any class is defined as the ratio between the positively true (pixels classified to be in a class and are truly in that class) and all pixels that are used as ground truth of this class. The first dataset used was based on the Satimage dataset. The second set used a multispectral image from Egypt.

4.1. Set 1

The standard benchmark *Satimage* dataset is available to the public at *ftp://ftp.dice.ucl.ac.be/pub/neural-nets/ELENA/databases*. It contains intensities of pixels derived from Landsat Multi-Spectral Scanner (MSS) satellite images that have been segmented into 6 classes. 4 spectral bands were used and the feature vector contains intensities of the central 8 surrounding pixels, altogether 36 features quantized from 0 to 255. The training set contains 4435 vectors and the test set has 2000 vectors. The classification accuracy on this set for individual classifiers is shown in Table 1.

TABLE 1: COMPARISON OF CLASSIFIERS ACCURACIES ON SET 1.

Method of Classification	Test AC
k-NN (k=5)	90.3500
Bayes	85.7000
Parzen window (h=10)	89.1500
MLP, one hidden layer (36-150-6)	90.4500
MLP, 2 hidden layers (36-60-200-6)	90.8000
RBN	89.0500
PNN	90.4000

TABLE 2: COMPARISON OF ACCURACIES OF FUSION APPROACHES ON BOTH SETS.

Fusion approaches	Set 1	Set 2
Mean	91.7500	95.8730
Max	90.9500	94.1270
Median	92.2000	96.3492
Complete agreement	73.6500	87.9365
Majority voting	92.3155	97.0779
MLP	90.8000	95.5556

The tabulated classifiers accuracies outperform those reported in [8]. This reflects our careful classifiers designing phase. Voting, statistical, and MLP network with topology (36-100-6) decision fusion approaches were applied to the 7 classifiers. The second column of Table 2 summaries the results obtained.

4.2 Set 2

The data set is acquired from Landsat Thematic Mapper (TM). A 252x251 image scene of El-Fayum (Egypt), all 7 bands are used including the infrared thermal band, so each pixel is characterized by seven-element feature vector containing the brightness values in the six optical bands and the one of the infrared thermal band. Seven distinct classes have been selected by human experts, who also selected 30 points per band, or 210 training points, per class and 90 test point per class. The results produced by the classifiers are summarized in Table 3. Voting, statistical rules, and MLP network with topology (42-15-7) decision fusion approaches are also applied to the 6 classifiers as was done on set 1. The obtained performance has been shown in column 3 of Table 2. A sample of the pixel-wise classified image with/without fusion is demonstrated in Figure 1.

5. CONCLUSION

In this paper, a comparison of the performance of several statistical and neural methods for both of supervised classification of remote-sensing images and decision fusion has been given. Even if all individual classifiers have been optimized, decision fusion approaches can improve the overall performance. However, the benefit from decision fusion may be limited when there is a small number of training data, or when the classification accuracy of an individual classifier is sufficiently high as the case in set 2. Reported results have showed that majority voting, despite being simple to implement, exhibited the best performance among other fusion scheme on both datasets. Although this observation was also made by some researchers (e.g., [9]), it calls for further investigation and more experiments.

TABLE 3: COMPARISON OF CLASSIFIERS ACCURACIES ON SET 2

Method of Classification	Test AC
k-NN (k=3)	94.9126
Bayes	93.9683
Parzen window (h=10)	95.2381
BPN (7-15-7)	94.2857
BPN (7-15-10-7)	94.7619
RBN	93.3333
PNN	95.7143

The next step in this research is to add more classification methods such as fuzzy based approaches, in addition to applying other decision fusion methods such as fusion using mutual information. A direction worthy of investigation is the use of the fusion MLP network to perform classifier selection [10] along with coming up with the best method to combine the classifiers selected from the pool of candidate classifiers. This can hopefully improve the complexity and/or the accuracy of the overall classification scheme!

6. REFERENCES

[1] J. Benediktsson and I. Kanellopoulos, "Classification of Multisource and Hyperspectral Data Based on Decision Fusion," *IEEE Tran. Geosci. Remote Sensing*, vol 37, pp.1367–1377, May 1999.

[2] A. A. Farag, R. M. Mohamed and H. Mahdi, "Experiments in Image Classification and Data Fusion," *Proceedings of 5th International Conference on Information Fusion*, Annapolis, MD, Vol. 1, pp. 299-308, July 2002.

[3] L.I.Kuncheva, J.C Bezedek, and R.P.W Dubin, "Decision Templates for Multiple Classifier Fusion: an Experimental Comparison," *Pattern Recognition*, vol 34, pp. 299-314, 2001

[4] L.I. Kuncheva, "Switching between selection and fusion in combining classifiers: An experiment," *IEEE transactions On Systems Man And Cybernetics,* Part B-cybernetics, vol. 32, pp. 146-156, August 2002.

[5] R. O. Duda, P. E. Hart and D. Stork, "*Pattern Classification*," 2nd edition, Wiley, 2001

[6] K. Woods, W. Ph. Kegelmeyer, and K.Bowyer, "Combination of Multiple Classifiers Using Local Accuracy Estimates," *IEEE Tran. Trans. Pattern Analysis and machine Intelligence*, vol. 19, pp. 405-41, April 1997

[7] J. Kittler, M. Hatef, R.P.W. Duin, and Jiri Matas, "On Combining Classifiers," *IEEE Tran. Trans. Pattern Analysis and machine Intelligence*, vol. 20, pp. 226-239, March 1997

[8] L.O. Jimenez, A. Morales-Morell, and A.Creus. "Classification of Hyperspectral Data Based On Feature and Decision Fusion Approaches Using Projection Pursuit, Majority Voting, and Neural Networks," *IEEE Tran. Geosci. Remote Sensing*, vol 37, pp.1360–1366, May 1999.

[9] G.Giacinto, and F.Roli, "Ensembles of Neural Networks for Soft Classification of Remote Sensing Images," *Proc. of the European Symposium on Intelligent Techniques*, Bari, Italy, pp.166-170, March 20-21, 1997.

[10] G.Giacinto, and F.Roli, "Methods for dynamic classifier selection," *10th International Conference on Image Analysis and Processing*, Venice, Italy, pp.659-664, 1999.

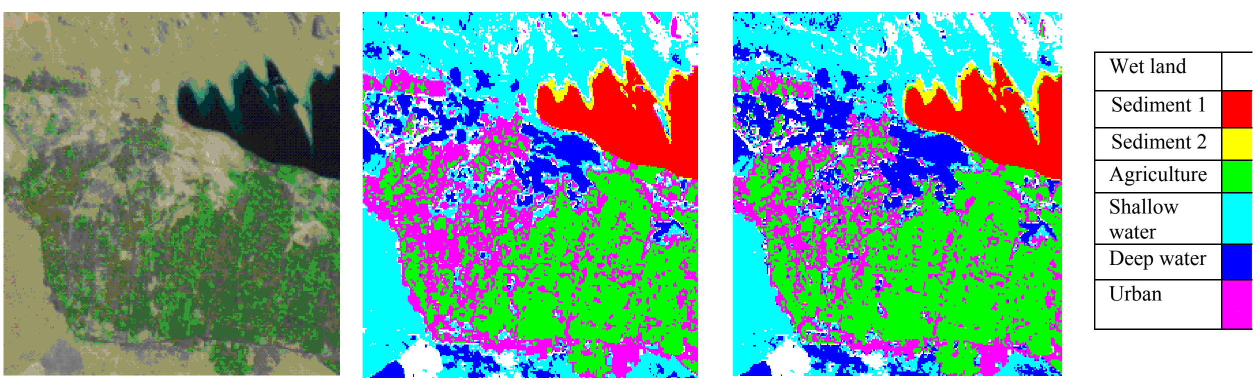

Fig. 1. Sample classification results on Set 2: (a) Original image, (b) Classification using the Parzen window classifier, (c) Result of fusing 7 classifiers using the median rule.