IMPROVING MAXIMUM LIKELIHOOD CLASSIFICATION ACCURACY USING A-PRIORI PROBABILITY

Alesheikh, Ali A.*
Fariba Sadeghi Naeeni Fard**

*Assistant Professor,
Dept. of Geodesy and Geomatics Eng.,
K.N. Toosi University of Technology,
Email:ali_alesheikh@hotmail.com
Tel: (+98 21) 8779473 Fax: (+98 21) 8779476

**Graduate Student
Dept. of Geodesy and Geomatics Eng.,
K.N. Toosi University of Technology,
Email: sadeghi@ncc.neda.net.ir
Tel: (+98 21) 8779473 Fax: (+98 21) 8779476

KEY WORDS: Maximum Likelihood Classification, A-Priori Probability, External Knowledge

ABSTRACT
Remotely sensed images are major sources of information, and as such, are used in many fields like morphology, geology, and agriculture. So far, many efforts have been performed to extract information from satellite images, and various methods have been developed in the field. Two main approaches are visual interpretation and quantitative analysis (digital interpretation). Among digital techniques, classification is a common and powerful information extraction, which is used in remote sensing. There are many classification methods that have their own advantages and drawbacks. Standard classification methods usually take pixels as fundamental elements and try to label the pixels based on their spectral properties. It is clear that using spectral properties alone, may not lead to adequate accuracy. Maximum Likelihood Classification (MLC) is perhaps the most widely used classification method. The underlying assumption on performing MLC is that the prior probability of land cover is equal, due to insufficient information. However, aprior occurrence probability gives a crucial effect to classification results. As long as the class showing the highest likelihood is allocated to a pixel, misclassification errors are unavoidable. The objective of this paper is to improve the accuracy of Maximum Likelihood Classification method using apriori information. Estimates of a-prior probability through, crop areas, crop calendar, soil type information and some a-prior probability about agricultural practices have been used in assigning the probability to pixels before classification. Methods of gathering information on a-prior probability , together with the results of the research have been presented and analyzed in this paper. An industrial agricultural field, Moghan Plain, in North Western Iran has been selected for testing the methods. Validation results demonstrate that this way is effective to improve classification errors.

INTRODUCTION

In the recent years, maximum likelihood classification has found wide application in the field of remote sensing, based on multivariate normal distribution theory. Future work may well produce an integrating method from which a user can select a mix appropriate to the spatial, spectral, and temporal resolution of the data in hand and information output desired (Richards, 1993).

The purpose of this paper is to show how the use of prior information about the expected distribution of classes in a final classification map can be used to improve classification accuracies. Prior information is incorporated through the use of prior probabilities -that is, probabilities of occurrence of classes which are based on separate, independent knowledge concerning the area to be classified. Used in their simplest form, the probabilities weigh the classes according to their expected distribution in the output dataset by shifting decision space boundaries to produce larger volumes in measurement space for classes that are expected to be large and smaller volumes for classes expected to be small.

The incorporation of prior probabilities into the maximum likelihood decision rule can also provide a mechanism for merging continuously measured observations (multi spectral signatures) with discretely measured collateral variables. In this way, the classification process can incorporate discrete collateral information into the decision rule through a model contingent on an external conditioning variable. Thus, prior probabilities provide a powerful mechanism for merging collateral datasets with multi spectral images for classification processes (Gorte, 1998).

Study Area and Data

The study area is located in the north-western of Iran which is called Moghan (Figure 1). Moghan Agro Industry and Livestock Co has under taken various activities in the field of crop production, horticulture, animal husbandry and related industries. About 300,000 tons of various crops such as wheat, barely, maize (seed, grain and forage), sugar beet and alfalfa is produced annually in 18000 ha of irrigated farms (Figure 2). In this research, three parts of irrigated farms are used. These three parts are flat with 2% slope.

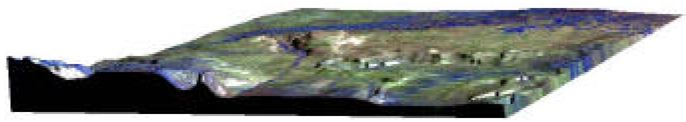

Figure 1. DEM of Moghan from 10 meters contour.

Field boundaries and their crop types are available from 1997 till 2001 (Figure 2).

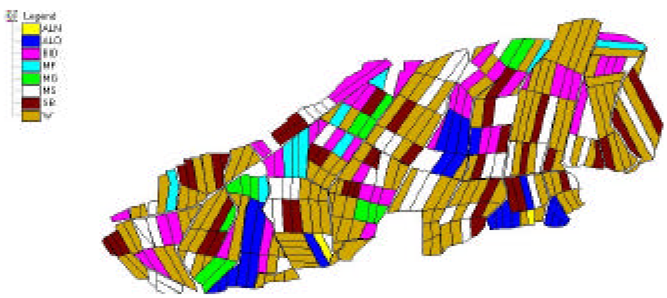

Figure 2. Field boundaries of Moghan agricultural fields

The ETM+ image of the study area which was acquired on 2001-05-23 and the 1/50000 map of it have been used in this study.

Review of Maximum Likelihood Classification

To understand the application of prior probabilities to a classification problem, the mathematics of the maximum likelihood decision rule must be understood. For the multivariate case, we assume each observation **X**(pixel) consists of a set of measurements on **p** variables (channels). Through some external procedures, we identify a set of observations which correspond to a class-that is, a set of similar objects characterized by a vector of means on measurement variables and a variance covariance matrix describing the interrelationships among the measurement variables which are characteristics of the class (Abkar, 1999). Multivariate normal statistical theory describes the probability that an observation **X** will occur, given that it belongs to a class **k**, as the following function:

$$\Phi\kappa = (2\pi)^{-p/} \Sigma_{\kappa}|^{-1/2} e^{-1/2(\chi-\mu)'\Sigma(\chi-\mu)} \qquad (1)$$

The quadratic product

$$(X_i - m_k)'D_k^{-1}(X_i - m_k) \qquad (2)$$

can be thought as a squared distance function which measures the distance between the observation and the class mean as scaled and corrected for variance and covariance of the class. As applied in a maximum likelihood decision rule, Equation (1) allows the calculation of the probability that an observation is a member of each of **k** classes. The individual is then assigned to the class for which the probability value is greatest. In an operational context, observed means, variances, and covariances substituted by the log form of the Equation (1).

$$Ln[\Phi\kappa(X_i)] = -p/2 \ln|2\pi| + -1/2 \ln|D_k| - 1/2((X_i - m_k)' D_k^{-1}(X_i - m_k) \quad (3)$$

Since the log of the probability is a monotonic increasing function of the probability, the decision can be made by comparing values for each class as calculated from the right hand side of this equation. A simpler decision rule, R1, can be derived from Equation (3) by eliminating the constants R1: Choose **k** which minimizes

$$F_{1,k} = \ln|D_k| + (X_i - m_k)' D_k^{-1}(X_i - m_k) \quad (4)$$

The use of prior probabilities in the decision rule. The maximum likelihood decision rule can be modified easily to take into account in the population of observations as a whole. The prior probability itself is simply an estimate of the objects which will fall into a particular class. These prior probabilities are sometimes termed "weights" since the modified classification rule will tend to weigh more heavily those classes with higher prior probabilities. Prior probabilities are incorporated into the classification through manipulation of the law of Conditional Probability (Alesheikh, 1998). To begin, two probabilities are defined: $P(w_k)$, the probability that an observation will be drawn from class w_k; and $P(X_i)$, the probability of occurrence of the measurement vector X_i. The law of Conditional Probability states that

$$P\{w_k | X_i\} = \frac{P\{w_k, X_i\}}{P\{X_i\}} \quad (5)$$

the probability on the left-hand side of this expression will form the basis of a modified decision rule, since the ith observation is assigned to that class w_k which has the highest probability of occurrence given the p-dimensional vector X_i which has been observed. Using the law of Conditional Probability, we find that

$$P\{X_i | w_k\} = \frac{P\{w_k, X_i\}}{P\{w_k\}} \quad (6)$$

In this Equation, the left-hand term describes the probability that the measurement vector will take on the values X_i given that the object measured is a member of class w_k. This probability could be determined by sampling a population of measurement vectors for observations known to be from class w_k. However, the distribution of such vectors is usually assumed to be Gaussian. Thus, we can assume that $P\{X_i | w_k\}$ is acceptably estimated by (X_i) and rewrite Equation (6) as

$$\Phi\kappa(X_i) = \frac{P\{w_k, X_i\}}{P\{w_k\}} \quad (7)$$

Rearranging the Equation

$$P\{w_k, X_i\} = \Phi\kappa(X_i)\, P\{w_k\} \quad \Phi_k^*(X_i) \quad (8)$$

Thus, the numerator of Equation (5) can be evaluated as the product of the multivariate density function (X_i) and the prior probability of occurrence of class w_k. To evaluate the denominator of expression (5), and knowing that for all k classes the conditional probabilities must sum to 1,

$$P\{w_k | X_i\} = \frac{\Phi_k(X_i)\, P\{w_k\}}{\Sigma^K_{k=1} \Phi_k(X_i)\, P\{w_k\}} = \frac{\Phi_k^*(X_i)}{\Sigma^K_{k=1} \Phi_k^*(X_i)} \quad (9)$$

This Equation provides the basis for the decision rule which includes prior probabilities. Since the denominations remain constant for all classes, the observation is simply assigned to the class for which $\Phi_k^*(X_i)$ the product of $\Phi_k(X_i)$ and $P\{w_k\}$, is a maximum. In its simplest form, this decision rule can be stated as: R2: Choose k which minimizes

$$F_{2,k} = \ln|D_k| + (X_i - m_k)'D_k^{-1}(X_i - m_k) - 2\ln P\{w_k\} \quad (10)$$

It is important to understand how this decision rule behaves with different prior probabilities. If the prior probability $P\{w_k\}$ is very small, then its natural logarithm will be a large negative number; when multiplied by -2, it will become a large positive number and thus $F_{2,k}$ for such a class will never be minimal. Therefore, setting a very small prior probability will effectively remove a class from the output classification. Note that this effect will occur even if the observation vector X_i is coincident with class mean vector m_k. In such a case, the quadratic product distance function $(X_i-m_k)'D_k^{-1}(X_i-m_k)$ goes to zero, but the prior probability term $-2\ln P\{w_k\}$ can still be large. Thus, it is entirely possible that the observation will be classified into a different class, one for which the distance function is quite large.

As the prior probability $P\{w_k\}$ becomes large and approaches 1, its logarithm will go to zero and $F_{2,k}$ will approach $F_{1,k}$ for that class. Since this probability and all others must sum to one, however, the prior probabilities of the remaining classes will be small numbers and their values of $F_{2,k}$ will be greatly augmented. The effect will be to force classification into the class with high probability. Therefore, the more extreme are the values of the prior probabilities, the less important are the actual observation vector X_i.

EXPERIMENTAL WORK

Training data for each class Have been collected, and then the image is classified by maximum likelihood approach. It is assumed that a prior probability of the whole classes are equal. Figure 3 is the classified image.

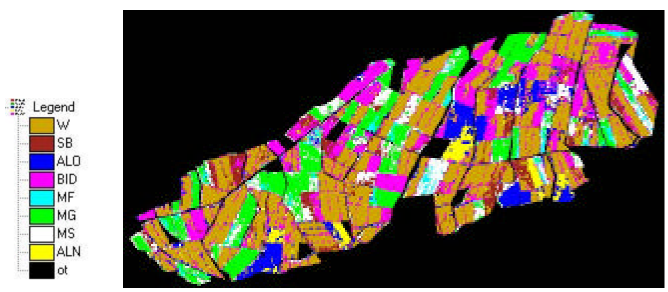

Figure 3. Classified image by maximum likelihood approach and equal a prior probability.

Overall accuracy of this approach is 52%. In this stage rule maps of the 8 crops can be calculated which is the basis for decision making for the software. For example Table 1 can show the rule matrix (the probability of each pixel for class W).

W =

	63	64	65	66	67	68	69	70	71
39	0	0.04	0.04	0.03	0.02	0.1	0	0	0
40	0	0	0	0.04	0.05	0.21	0	0	0
41	0	0	0	0.04	0.05	0.21	0	0	0
42	0	0	0	0	0.13	0.15	0	0	0
43	0	0	0	0	0.05	0.26	0.09	0	0
44	0	0	0	0	0.05	0.26	0.09	0	0
45	0.02	0.13	0.13	0	0	0.13	0.22	0.34	0.34
46	0.15	0.09	0.09	0.09	0	0	0.17	0.36	0.36
47	0.23	0.36	0.36	0.02	0.37	0.67	0.18	0.06	0.06
48	0.23	0.36	0.36	0.02	0.37	0.67	0.18	0.06	0.06
49	0.34	0.12	0.12	0.15	0.29	0.47	0.51	0	0
50	0.18	0.14	0.14	0.21	0.08	0.44	0.41	0.26	0.26
51	0.18	0.14	0.14	0.21	0.08	0.44	0.41	0.26	0.26

Table 1. Rule matrix for class W.

Since the sum of rule matrices for whole classes must be one, Table 1 will be modified to Table .2:

$W + SB + ALO + BID + MF + MG + MS + ALN =$

	21	22	23	24	25	26	27
0	1	1	0.99	1	1	1	1
1	1	1	0.99	1	1	1	1
2	1	1	1.01	1	1	1	1
3	0.99	0.99	1	0.99	1	1	1
4	0.99	0.99	1	0.99	1	1	1
5	1.01	1.01	1	1	1	1	1.01
6	1.01	1.01	1	0.99	1	1	1
7	1.01	1.01	1	0.99	1	1	1
8	1	1	1	0.99	1	0.99	1
9	1.01	1.01	1	1.01	1	0.99	1
10	1	1	1	0.99	0.99	1	1
11	1	1	1	0.99	0.99	1	1
12	1	1	1	0.99	1	0.99	1

Table 2. Sum of rule matrices for 8 classes.

Prior Probabilities Contingent on a Single External Conditioning Variable

Having shown how to modify the decision rule to take into account a set of prior probabilities, it is only a small step to consider several sets of probabilities, in which an external information source identifies which set is to be used in the decision rule. Hence, a third variable v_j, is introduced, which indicates the state of the external conditioning variable (e.g. crop calendar) associated with the observation. It is expected to find an expression describing the probability that an observation will be a member of the class w_k, given its vector of observed measurements and the fact that it belongs to class v_j of the external conditioning variable, namely.

$$P\{w_k | X_i, v_j\} \qquad (11)$$

In deriving an expression to find this probability, we can make the assumption that the mean vector and dispersion matrix of the class will be the same regardless of the state of the external conditioning variable. Considering this assumption and expanding Equation 11, it results in

$$P\{w_k | X_i, v_j\} = \frac{\Phi_\kappa(X_i)\, P\{w_k, v_j\}}{\Sigma^K_{k=1} \Phi_\kappa(X_i)\, P\{w_k, v_j\}} \qquad (12)$$

This result is analogous to Equation (5); note that the denominator remains constant for all k, and need not be calculated to select the class w_k for which $\Phi_\kappa^{**}(X_i)$ is a maximum.

The application of this equation in classification requires that the joint probabilities $P\{w_k, v_j\}$ be known. However, a simpler form using conditional probabilities directly obtained from a stratified random sample can be obtained through the application of the Law of Conditional Probability.

$$P\{w_k, v_i\} = P\{w_k | v_j\}\, P\{v_j\} \qquad (13)$$

Thus, either the joint or conditional probabilities may be used in the decision rule: R3 : Choose k which minimizes

$$F_{3,k} = \ln|D_k| + (X_i - m_k)' D_k^{-1} (X_i - m_k) - 2\ln P\{w_k | v_j\} \qquad (14)$$

RESULTS

In the Moghan, crop production activities almost have disciplines, i.e. most farmers practice crop rotation to increase their yields, the rotational patterns, combined with information about crops-planted in the previous years, can be used to predict the current crop. The five years crop rotation matrix (transition matrix) is presented in Table 3.

TRANSITION MATRIX "1999-2000, 2000-20001"

	ALN	ALO	BID	MF	MG	MS	SB	W
ALN	0.01	0.99	0	0	0	0	0	0
ALO	0	0.97	0	0	0	0	0	0.03
BID	0	0	0.36	0.17	0.11	0.03	0.13	0.2
MF	0	0	0.48	0	0	0	0	0.52
MG	0	0	0.74	0	0	0	0	0.26
MS	0	0	0.05	0	0	0	0	0.95
SB	0	0	0.07	0.04	0	0	0	0.9
W	0.01	0	0.1	0.02	0.08	0.32	0.3	0.16

TRANSITION MATRIX "1997-1998, 1998-1999"

	ALN	ALO	BID	MF	MG	MS	SB	W
ALN	0.02	0.97	0	0	0	0	0	0
ALO	0	1	0	0	0	0	0	0
BID	0	0	0.22	0	0.36	0.07	0.15	0
MF	0	0	0.47	0	0	0	0	0.48
MG	0	0	0	0	0	0	0	1
MS	0	0	0	0	0	0	0	1
SB	0	0	0	0	0	0	0	1
W	0.05	0	0.17	0	0.07	0.23	0.39	0.06

TRANSITION MATRIX "1996-1997, 1997-1998"

	ALN	ALO	BID	MF	MG	MS	SB	W
ALN	0.02	0.98	0	0	0	0	0	0
ALO	0	0.71	0	0	0	0	0	0.29
BID	0	0	0.41	0.19	0.08	0.07	0.21	0
MF	0	0	0.19	0	0	0	0.11	0.57
MG	0	0	0.12	0	0	0	0	0.88
MS	0	0	0.02	0	0	0	0.02	0.96
SB	0	0	0	0.11	0	0	0	0.89
W	0.03	0	0.1	0.05	0.1	0.21	0.39	0.04

Table 3: Transition Matrix in 5 consequent years

Then by using produced transition matrices, which were described above and "1998-1999, 1999-2000" transition matrix (for estimating a prior probability) the classification calculated once again and validation results show that overall accuracy goes to 63%.

CONCLUSIONS

This paper has pointed out limitations of classification by MLM, and then proposed a classification method by considering a prior probability with discovering crop rotation schemes. Validation showed that the proposed method gave higher classification accuracy than traditional MLM. Further research must be conducted to explore other external knowledge to be included in the procedure and enhace the results.

REFERENCES

Abkar, A. 1999. Likelihhod-based segmentation and classification of remotely sensed Images. Ph.D. Thesis, University of Twente, ITC, Enschede, The Netherlands.

Alesheikh, Ali A. 1998. Handling Uncertainty in Object-Based GISs. Ph.D. Thesis, Departmene of Geomatics Engineering, The University of Calgary, AB, Canada

Gorte, B. 1998. Probability segmentation of remotely sensed images. Ph.D. thesis. Wayeningen Agricultural University (WAU), ITC, Enschede, The Netherlands.

Richards, J.A. 1993. Remote Sensing Digital Image Analysis: An Introduction. Second Edition, Springer, ISBN 0-387-5480-8