OBJECT CLASSIFIERS FOR FOREST CLASSIFICATION

Gintautas Palubinskas

Data Analysis Department
Institute of Mathematics and Informatics
Akademijos 4, Vilnius 2600, Lithuania
Commision III WG III/3
E-mail: raudys%ma-mii.lt.su@fuug.fi

ABSTRACT :

The aim of this research is to compare the performance (probability of misclassification) of several object classifiers (some of them developed by author) and classical per-pixel maximum likelihood classifier for forest (deciduous, coniferous and mixed) classification.

The new analytical method (derived by author) is applied for the selection of object classifier based on calculating the probability of misclassification.

Object classifiers are supervised maximum likelihood classifiers incorporating spatial characteristics of an image during classification based on Markov random field model.

The research is carried out on Landsat TM data received from IFAG, Frankfurt a.M.

During investigation the Image Analysis and Classification System *IMAX* (developed by author and his group) is used.

First results show the complexity of the problem and the need for further investigation.

KEYWORDS : Thematic information extraction, Spatial characteristics of images, Pattern recognition, Landsat TM.

1. INTRODUCTION

The thematic information extraction from remote sensing images is important for solving many practical problems. Usually supervised maximum likelihood classifiers which assign each pixel of an image to one of m known classes, i.e. so called per-pixel classifiers, are used (Jensen, 1986; Richards, 1986). But often the quality of such classifiers is not satisfactory.

One of the ways to solve this problem is to incorporate spatial characteristics of an image in the process of classification. There are some reviews on these questions (Landgrebe, 1981; Swain, 1985; Alfiorov, 1989; Palubinskas, 1990a). We can see that there are three approaches how to use the spatial information: textural, object and contextual. This work is concerned with the object approach.

Object classifiers assign the whole object of an image or the central pixel of an object to one of m known classes. But there are difficulties in realization of object approach in general case because of the high dimensionality of the vector to be classified. Assumptions about the kind of dependence between intensity values of neighboring pixels must be made. Object classifiers first were introduced in (Ketting, 1976; Landgrebe, 1980) in the case of independent pixels of an object. Another assumption is based on Markov type dependence, particulary on the separable correlation model. All known object classifiers are based on this model.

In (Palubinskas, 1988a; Palubinskas, 1989) the systematization of image models based on the separable correlation model is made and on this basis some original classifiers are proposed. So in total 14 object classifiers (some of them are in the publications of Guyon and Yao, 1987; Mardia, 1984; Switzer, 1980) were investigated theoretically. The quality of classifiers is usually measured by the probability of misclassification (PMC). The new analytical method to calculate the PMC is proposed in (Palubinskas, 1988b; Palubinskas, 1992) which allows to compare the performance of several classifiers in the same conditions. This method is much more cheaper than the traditional method, when the PMC is evaluated on modeled data. This theoretical analysis helped us to select 5 object classifiers from 14 for further investigation on real data.

There are some papers (Landgrebe, 1980; Kalayeh and Landgrebe, 1987; Mardia, 1984; Switzer, 1980 and Palubinskas, 1990b) where some of these object classifiers are tested on real remote sensing imagery. These experiments allow to select 3 object classifiers from 5 and they can be recommended for practical use. However, this conclusion is valid not for all situations.

In this work the above mentioned analytical method of selection of a classifier is used in the analysis of remote sensing imagery. At first the statistical characteristics of training data and statistical charateristics which are used for a classifier design are calculated. Then the analytical PMC of this classifier is calculated. So the analytical method allows us to select the object classifier from a set of available with a very little computer time expenses. Then the selected classifier can be run on the full data set.

The work of this method is illustrated on the example of classifying the Landsat TM image of Frankfurt am Main surroundings recorded on 30 July 1984 (received from IFAG - Institut fur Angewandte Geodasie).

In Section 2 the object classifiers are described briefly. Section 3 presents the new analytical method for classifier selection. In Section 4 the experimental results are presented and finally Section 5 provides the concluding remarks.

2. OBJECT CLASSIFIERS

Consider a two-dimensional multispectral image, where the pixels of an image are q-dimensional vectors

$$x = (x^1, x^2, ..., x^q)'$$

and x^l is the lth feature of the pixel x. The group of adjacent pixels defines an object of an image.

Let us denote the object of an image by X, where

$$X = (x_1, x_2, ..., x_n)'$$

is a vector of size $Nq\times 1$ and N is the number of pixels in the object.

Suppose that the pixels of an object have Gaussian distribution, i.e.

$$X \sim N(\mu_i, \Sigma_i),$$

where

$$\mu_i = (\mu_i^1, \mu_i^2, ..., \mu_i^q)$$

is the mean of the pixel x of the ith class and Σ_i is the covariance matrix of x.

The aim of image recognition is to classify each pixel of an object or the whole object of an image into one of m possible classes.

The classical procedure is a per-pixel classifier whichs assigns pixel x to the class i, when

$$p(x|\omega_i) = \max_{1\leq j\leq m} p(x|\omega_j), \tag{1}$$

where p is the class-conditional density function for the class ω_i. This decision rule classifiers pixels alone using only the spectral characteristics of pixels of an image.

The decision rule for the object classifier is the following

$$p(X|\omega_i) = \max_{1\leq j\leq m} p(X|\omega_j). \tag{2}$$

In Gaussian case $p(X|\omega_i)$ is characterized by K - covariance matrix of an object X of size $Nq\times Nq$ and $M = (\mu, ..., \mu)'$ - mean of X of size $Nq\times 1$.

The usage of object classifier (2) in the general case is very restricted because for rather large N and q it is difficult to estimate matrix K because of a limited size of the learning sample. To overcome these difficulties, one has to make certain assumptions about the structure of K.

The solution of this problem is based on the two following assumptions about the structure of matrix K.

First, it is assumed that the correlation between the pixels of an object does not depend on q, i.e. $K = R\otimes\Sigma$, where R is the spatial correlation matrix of size $N\times N$, $\otimes$ is the Kroneker product.

Secondly, assumptions about the structure of matrix R are made. Often, it is assumed that the pixels inside the object are independent (Ketting, 1976; Landgrebe, 1980). In this case R is the identity matrix and the spatial characteristics are employed indirectly. Another popular assumption is that an object of an image is a Markov random field (Switzer, 1980; Mardia, 1984; Guyon and Yao, 1987; Kalayeh and Landgrebe, 1987; Palubinskas, 1988a) which is represented by causal autoregressive model. In this case matrix R is characterized by few parameters, the number of which depends on the order of Markov model. All object classifiers are based on the popular separable correlation function

$$\mathrm{corr}(x_{ij}, x_{kl}) = \rho_1^{|i-k|}\rho_2^{|j-l|},$$

where ρ_1 and ρ_2 are spatial correlation coefficients between adjacent pixels of an object in the horizontal and vertical directions, respectively.

So we see that depending on the structure of matrix R there can be a wide variety of object classifiers. The problem of selection of object classifier is actual. From Section 1 we see that there are a lot theoretical investigations of object classifiers, but there are only few results on real data.

So object classifiers must be investigated more thoroughly on real data. For this purpose *IMAX* - image analysis and classification system was developed in Data Analysis Department, Institute of Mathematics and Informatics (Palubinskas, Cibas, Repsys, 1991) on PC computer.

IMAX offers the choice between 10 classifiers: one conventional per-pixel maximum likelihood classifier (PIX), one conventional per-pixel minimum distance classifier, 6 object maximum likelihood classifiers and 2 object minimum distance classifiers. Object classifiers classify the central pixel of an object. Then the window is moved by one pixel. Four object classifiers are for cross-shaped block and other four are for square-shaped block. In each group there are the following object classifiers:
- object classifier incorporating the spatial characteristics of an image directly on the base of causal Markov random field model of the first and third order, respectively (OMARK1, OMARK3),
- object classifier based on the assumption that the pixels inside the block are independent (OIND1, OIND3),
- object classifier which classify the mean of the block under the assumption that the pixels inside the block are independent (OMEANIND1, OMEANIND3),
- object classifier which classify the mean of the block under the assumption that the pixels inside the block are independent and with covariance matrix equal identity matrix (OMEAN1, OMEAN3).

These 8 object classifiers were investigated in this work. For detailed description of object classifiers see (Palubinskas, 1988a).

3. ANALYTICAL METHOD FOR EVALUATION OF CLASSIFIER PERFORMANCE

The conventional method for evaluation of classifier performance (the probability of error) is mathematical modeling. First, one has to programm the classifier, then to run it on test data in order to calculate the probability of error. But it is rather time consuming way.

The analytical way to calculate the probability of error requires only calculation of statistical characteristics of data (mean and covariance matrix). So it can be much more faster. In (Fukunaga, 1972) two analytical methods of calculating the probability of error are proposed. First calculates the exact probability of error of discriminant function using Imhof formula or characteristic function. Second approximates the probability of error under assumption of discriminant function normal distribution. But these formulae are in the case of classifier model and data model equality. In practice, usually, these models are different. In (Palubinskas, 1988b; 1992) these formulae are extended for the case of classifier model and data model inequality. So these formulae allow to compare several classifiers in the same conditions. Now we shall present these formulae in more details.

The purpose of pattern recognition is to determine which category or class a given sample belongs. Lets consider the two-class problem, i.e. each sample belongs to one of two classes, ω_1 or ω_2. The conditional density functions and the a *priori* probabilities are assumed to be known.

Let $X = (x_1, x_2, ..., x_n)'$ be an observation vector. The Bayes decision rule (1 or 2) can be written as follows:

$$h(X) = -\ln p(X|\omega_1) + \ln p(X|\omega_2) \lessgtr t$$

$$= \ln \frac{p(\omega_1)}{p(\omega_2)} \to X \in \begin{cases} \omega_1 \\ \omega_2, \end{cases} \tag{3}$$

where $p(\omega_i)$ - *a priori* probabilities and $p(X|\omega_i)$ - conditional density functions.

The probability of error evaluates the performance of a decision rule. The probability of error can be calculated as follows:

$$\varepsilon = p(\omega_1) \cdot \varepsilon_1 + p(\omega_2) \cdot \varepsilon_2, \tag{4}$$

where

$$\varepsilon_1 = \int_t^{+\infty} p(h|\omega_1)dh, \tag{5}$$

$$\varepsilon_2 = \int_{-\infty}^{t} p(h|\omega_2)dh. \tag{6}$$

Formulae (5,6) do not lead to straightforward calculation of probability of error, because we need to know the density function of $h(X)$. But there are some cases when this can be done.

When the $p(X|\omega_i)$ are normal with expected vectors M_i and covariance matrices Σ_i, the Bayes decision rule (3) becomes

$$h(X) = \frac{1}{2}(X - M_1)'\Sigma_1{}^{-1}(X - M_1)$$

$$-\frac{1}{2}(X - M_2)'\Sigma_2{}^{-1}(X - M_2) + \frac{1}{2}\ln \frac{|\Sigma_1|}{|\Sigma_2|} \lessgtr t \to X \in \begin{cases} \omega_1 \\ \omega_2. \end{cases} \tag{7}$$

From (7) we see that the discriminant function of $h(X)$ depends upon the following parameters: mean vectors M_i and covariance matrices $\Sigma_i, i = 1,2$ (classifier model).

Let X, which is to be classified, is normally distributed vector with the true parameters: M_i^T and $\Sigma_i^T, i = 1,2$ (data model).

3.1. Exact Probability of Error

The probability of error for the first class using Imhof formula can be expressed as

$$\varepsilon_1 = \frac{1}{2} + \frac{1}{\pi}\int_0^{+\infty} \frac{\sin\theta(u)}{u\rho(u)}du, \tag{8}$$

where

$$\theta(u) = \frac{1}{2}\sum_{i=1}^{n}[\tan^{-1}(d_{i,T}u) + v_{4,i}^2 d_{i,T}u(1 + d_{i,T}^2u^2)^{-1}] - \frac{1}{2}C_2u,$$

$$\rho(u) = \prod_{i=1}^{n}(1 + d_{i,T}^2u^2)^{\frac{1}{4}} e^{\{\frac{1}{2}\sum_{i=1}^{n}(v_{4,i}d_{i,T}u)^2/(1+d_{i,T}^2u^2)\}}.$$

The probability of error for the second class is obtained analogically.

We see that the formula for the probability of error is rather complicated for computing. In the next Section the simple formula for the approximation of probability of error is presented.

3.2. Approximate Probability of Error

When $h(X)$ is a normal random variable (5,6) becomes

$$\varepsilon_i = \Phi[(-1)^i \frac{t - \eta_i}{\sigma_i}],\ i = 1,2, \tag{9}$$

where

$$\eta_i = \mathrm{E}\{h(X)|\omega_i\}$$

$$= \frac{1}{2}[\sum_{j=1}^{2}(-1)^{j-1}\{\mathrm{tr}\{\Sigma_j{}^{-1}\Sigma_i^T\}$$

$$+(M_i^T - M_j)'\Sigma_j{}^{-1}(M_i^T - M_j)\} + \ln\frac{|\Sigma_1|}{|\Sigma_2|}],$$

$$\sigma_i^2 = \mathrm{E}[\{h(X) - \eta_i\}^2|\omega_i] = \frac{1}{4}[2\mathrm{tr}\{(\sum_{j=1}^{2}(-1)^{j-1}\Sigma_j{}^{-1}\Sigma_i^T)^2\} +$$

$$4\{\sum_{j=1}^{2}(-1)^{j-1}(M_i^T - M_j)'\Sigma_j{}^{-1}\}\Sigma_i{}^T$$

$$\{\sum_{j=1}^{2}(-1)^{j-1}(M_i^T - M_j)'\Sigma_j{}^{-1}\}],$$

$$\Phi(x) = (2\pi)^{-\frac{1}{2}}\int_{-\infty}^{x} \mathrm{e}^{-\frac{\xi^2}{2}}d\xi.$$

The accuracy of approximation of the probability of error is investigated in (Palubinskas, 1992). There we want to note that the accuracy of approximation is strongly influenced by the concrete structures of the covariance matrices of classes. Also to calculate the approximate of the probability of error is much faster than to calculate the exact probability of error. In the following Section both analytical methods are used for object classifier performance evaluation.

4. EXPERIMENTS

In this Section first experimental results on forest classification, based on LANDSAT TM data recorded on 30 July 1984, are presented. The aim of this research is to distinguish forest types: deciduous forest, coniferous forest and mixed forest with the help of object classifiers. We have to note that the same problem on the same data set was solved in (Schulz, 1988; Pyka, 1990) with the help of per-pixel classifier. The potential of object classifiers for forest classification is also of interest.

Defining the training and control fields for supervised classification is rather difficult task, especially when there is no possibility to get true ground information. So from visual analysis of multispectral images and topographic map 1:50 000 (supplied by IFAG) two fields were defined for each class (forest type). Only one band: No. 4 was used for classification. The results of classification are shown in Table 1. PR is correct classification on training sample and PK - on control sample.

We see the great potential of object classifiers. The use of unsupervised classification can help to select training fields more accurately. It is planned to do in further experiments. The use of analytical methods for evaluation of the performance of classifier can help significantly to reduce computer time. This is also planned for further experiments.

TABLE 1. The results of forest classification of nine classifiers

Classifier	PR(%)	PK(%)
PIX	76.22	75.99
OMARK1	89.47	87.44
OIND1	88.29	85.40
OMEANIND1	80.81	80.14
OMEAN1	67.91	67.91
OMARK3	90.66	86.25
OIND3	90.49	88.29
OMEANIND3	82.68	82.34
OMEAN3	68.93	68.93

5. CONCLUSION

In this work object classifiers incorporating spatial characteristics of an image during classification and based on Markov random field model are introduced.

Two analytical methods for the evaluation of performance of object classifier based on calculating the probability of error are presented.

First experimental results on forest classification, based on LANDSAT TM data, show the great potential of object classifiers comparing with per-pixel classifier. On the other hand we have to note the complexity of the problem and the need for further investigation.

6. REFERENCES

+ References from **JOURNALS**:

Alfiorov, G.A., 1989. Contextual methods for the analysis of aerocosmic images. In: Advances in Science and Technics, Series: Technical Cybernetics, Moscow, vol. 26, pp. 105-136 (in Russian).

Kalayeh, H.M. and Landgrebe, D.A., 1987. Stochastic model utilizing spectral and spatial characteristics. IEEE Trans. on Pattern Analysis and Machine Intelligence, 9:457-461.

Ketting, R.L. and Landgrebe, D.A., 1976. Classification of multispectral image data by extraction and classification of homogeneous objects. IEEE Trans. Geosci. Electronics, 14:19-26.

Landgrebe, D.A., 1980. The development of spectral-spatial classifier for earth observational data. Pattern Recognition, 12(3):165-175.

Landgrebe, D.A., 1981. Analysis technology for land remote sensing. Proc. IEEE, 69(5):628-643.

Mardia, K.V., 1984. Spatial discrimination and classification maps. Communications in Statistics. Theory and Methods, 13:2187-2197.

Palubinskas, G. 1988a. A comparative study of decision making algorithms in images modeled by Gaussian Markov random fields. International Journal of Pattern Recognition and Artificial Intelligence, 2(4):621-639.

Palubinskas, G. 1988b. Probability of misclassification of the Bayes decision rule for normal random vectors in the case of model and data inadequacy. In: S.Raudys, Ed., Statistical Problems of Control, Issue 82, Institute of Mathematics and Cybernetics Press, Vilnius, pp. 32-42 (in Russian).

Palubinskas, G., 1990a. A review of spatial image recognition methods. In: Statistical Problems of Control, Issue 93, Vilnius, pp. 194-214 (in Russian).

Palubinskas, G., 1990b. Object classifiers in remote sensing. In: Statistical Problems of Control, Issue 93, Vilnius, pp. 215-231 (in Russian).

Palubinskas, G., 1992. On approximation of the error of the Bayes decision rule. To appear.

Swain, P.H., 1985. Advanced interpretation techniques for Earth data information systems. Proc. IEEE, 73(6):1031-1039.

Switzer, P., 1980. Extensions of linear discriminant analysis for statistical classification of remotely sensed satellite imagery. Mathematical Geology, 12:367-376.

+ References from **BOOKS**:

Jensen J.R., 1986. Introductory Digital Image Processing: A Remote Sensing Perspective. Prentice-Hall, Englewood Cliffs, pp. 177-233.

Fukunaga K., 1972. Introduction to Statistical Pattern Recognition. Academic Press, New York, pp. 50-67.

Richards J.A., 1986. Remote Sensing Digital Image Analysis: An Introductory. Springer-Verlag, Berlin, pp. 173-250.

+ References from **GREY LITERATURE**:

Guyon X., and Yao J.F., 1987. Analyse discriminante contextuelle. In: Proc 5th International Symposium on Data Analysis and Informatics, Paris-France, Vol.1, pp. 43-52.

Palubinskas, G., 1989. Object classifiers for textured images. In: Mathematical Research, Vol.55 (Proc. of the 3rd International Conference CAIP'89 on Automatic Image Processing " Computer Analysis of Images and Patterns ", Leipzig (GDR), September 8-10, 1989), eds. K.Voss, D.Chetverikov and G.Sommer, Akademie Verlag Berlin, pp.141-147.

Palubinskas G., Cibas T., Repsys A., 1991. *IMAX* - image analysis and classification system. Technical Report, Vilnius, 9 pages.

Pyka K., 1990. Data preprocessing influence on classification results (in relation to forest classification). In: Proc. Symposium of ISPRS Com. III in Wuhan, pp. 743-755.

Schulz B.S., 1988. Hypothesis-free land use classification from LANDSAT-5 TM image data. In: International Archives of Photogrammetry and Remote Sensing, Kyoto, Vol. 27, pp. 736-742.