Research of efficiency of the statistical non-parametric pattern recognition models for forest land classification

Aleksander P. Guk [1,*] , *Larisa G.* Evstratova [2]

[1] Siberian State University of Geosystems and Technologies, 630018, 10 Plakhotnogo Str., Novosibirsk, Russian Federation
[2] State University of Land Use Planning, 105064, 15 Kazakova str, Moscow, Russian Federation

Abstract. The principles of the creation of pattern recognition models, based on using multispectral imagery of forest land, have been analyzed. The statistical non-parametric model has been suggested as a basic pattern recognition model and the probability density function – as a recognition feature. Efficiency of the different quality criteria has been discussed. The main directions for improving the pattern recognition models are regarded.

Measured spectral brightness of objects recorded as brightness of image elements is key information for pattern recognition using multispectral images. As a result of processing these measurements, it is required to classify the objects and determine their qualitative characteristics. To directly determine these characteristics is not possible, therefore, based on the measurements, it is necessary to form features that uniquely determine the object and some of its properties. Thus, the object and its state are associated with measurements through the selected characteristics. It would be ideal to take measurements directly. However, the state-of-the-art Earth remote-sensing instruments do not allow doing it. Moreover, it is difficult to measure even the spectral characteristics of an object, since measurements are influenced by a lot of factors (atmosphere, location of the carrier and the survey system). Another important factor is that spectral reflectance of various objects is strongly correlated and fluctuated.

Thus, a complex problem arises that includes several stages: study of spectral properties of object reflection, description of object properties based on spectral brightness coefficients (formation of object features from spectral brightness coefficients), and development of a communication pattern between the measured brightness and features. As is obvious, to implement this scheme it is needed to study the object properties, to assess the influence of the atmosphere and other factors and to form models describing all these relations. Obviously these are probabilistic-deterministic models (deterministic chaotic models), the solution of which presents severe difficulties.

However, in practice, a different approach is used. A model is created that connects measurements and properties and it is assumed that the properties are selected in such a

* Corresponding author: guk_ssga@mail.ru

way that they uniquely determine an object. Often these are very simple models for interactive decryption that use the intuition of the decrypting operator. They are cluster analysis models. Statistical models such as the Bayesian classifier, the maximum likelihood method, etc. are also among the same type of models. To simplify the problem solution the majority of such models use the normal law. However this often leads to an inadequate result.

In [1, 2] a statistical nonparametric approach to deciphering forest tracts based on multispectral satellite imagery is suggested. The essence of the suggested approach is as follows. For all classes of objects that need to be recognized by their images on multispectral imagery reference features are created – the probability density function for the brightness of the corresponding elements of a multispectral image. Reference probability density functions are defined by sufficiently large samples of size $N \le N_{min}$. Measurements are performed from the images of the reference objects; the number of measurements N_{min} is taken with the certainty matching the criterion $\alpha \le 0.1$.

As is known, with non-parametric methods, the probability density functions are defined as a result of statistical studies by various methods (histogram, Parzen method, polygonal estimates, etc.).

Taking into account that, when processing images, a random function is given by discrete values with a given step on the bounded interval [0 - 255], [0 - 1023] or [0 - 4095], then the histogram method should be considered as the most acceptable for estimating the non-parametric probability density function. Indeed, all other methods are based on the approximation (interpolation) of sample values at some interval, which is useful for a large number of measurements of the function being evaluated.

The histogram approach allows us to obtain a discrete estimate of the probability density functions, the certainty of which depends on the sample size. In [2], as a result of the research, the values of the minimum sample size that can be used to classify the data are obtained (with the certainty 1- α, where $\alpha = 0.05$).

When assessing the reference probability density functions, it is necessary to check the distribution "for normality". If the distribution does not conform to the normal law, then it is advisable to use non-parametric statistics. Normality test is based on the criterion ω^2 [3]:

$$\omega_m^2 = m\int ....\int \psi^2\left[F(X_m,\mu)\right]\left[F_m(X_m) - F_m(X_m,\mu)\right]^2 dF(X_m,\mu) > \omega_\alpha^2 \qquad (1)$$

where $F_{\text{m}}(X_{\text{m}})$ is the empirical probability density function of the sample $\{x_i\}^{\text{m}}$; $F_{\text{m}}(X_m,\mu)$ the normal probability density function with the parametric variable μ and $\Psi^2[F]$ is the weight function.

The reference functions are different for multispectral images obtained by different survey systems. Thus, for using the non-parametric approach it is required to create a data bank that includes probability density functions for all classes of objects that will be deciphered in the images. These functions must be obtained for different types of survey systems as well. As for the scale of the images, as a result of the research [2], we found out that when the image scale is changed within ±30%, the probability density function changes within the accuracy of the method.

After creating a bank of reference functions, you should decipher the images. Image segmentation is done for this. It is advisable to use area attributes, for example, textural ones. The number of elements in the selected segments must be at least n_0 and is determined from the condition of the minimum possible size of the section for which the difference from the reference distributions does not exceed $\alpha \le 0.05$. Then, for each selected segment, the histogram method is used to obtain estimates of the probability density function, which are compared with all the reference functions.

Thus, the essence of the method described is to obtain reference probability density functions, to compute the match value of the probability density function obtained by

measuring the object image in the deciphered image and assigning the object to a particular class.

The key problem of this method is to choose the decision rule and the quality criterion of the object function.

As is known [4], all types of decision rules are based on the formation of the likelihood ratio and its comparison with the certain threshold c, the value of which is selected in accordance with the quality criterion:

$$L = \frac{F(x_1, x_2 ... x_n / s_1)}{F(x_1, x_2 ... x_n / s_0)} \geq c \tag{2}$$

Practically when calculating statistics, it is necessary to take into account the form and structure of the probability density function.

Accordingly, statistics are divided into [4]:

a) Kolmogorov statistics, which estimate the maximum discrepancy between the values of the measured and reference functions:

$$D = \sup_x [F_0(x) - F_j(x)] \tag{3}$$

b) weighted Kolmogorov statistics of the form:

$$D = \int_{-\infty}^{+\infty} [F_0(x) - F_j(x)]^2 dF_0(x) \tag{4}$$

when more weight is given to the distribution tail area;

c) statistics of the form χ^2 (chi square):

$$D = \sum_{x_k}^{x_k + \Delta_l} \frac{F(x) - F_0(x)}{F(x)} \tag{5}$$

where k and l are quantities that allocate the necessary intervals from the probability density functions for quality analysis of the functions; $F(x_0)$ and $F(x)$ are the reference and measured probability density functions, respectively.

These statistics are easy to use for a univariate probability density function and the computational problem becomes quite complicated with increasing dimensionality of the measurement vector and the number of recognized objects. Moreover it is difficult to interpret the result obtained.

The probability density function has the properties of both element-by-element classification and area classification. It all depends on the choice of the parameters of the quality criterion. Indeed, if in (4) we properly choose the weights and integration limits, and the parameters k and l in the estimate (5), then we can select one or another part of the brightness as the main feature in recognition.

In the first stage, when using the suggested method [1], instead of statistics of the type "a", statistics based on definition of the correlation function $R_j(0) = \frac{F_0(x)F_j(x)}{F_0(x)}$ and the definition of $\max Rj_{\forall j}$ were used. This criterion, at least, allows us, in some sense, to estimate the general discrepancy of the probability density functions.

Approximately the same result is determined by the Fisher correlation method Z [3].

The recognition method suggested in [2] used the measurements of only one channel, which are completely characterized by a univariate probability density function. The channel informativity was determined visually – by the maximum difference of the probability density function curves for different classes in different channels. Although this approach is not mathematically justified, in practice its application is quite acceptable [4]. Nevertheless, it is clear that in general it is necessary to use the information from all the

channels of the survey system and, accordingly, it is required to estimate the multivariate density of information and to use the multivariate matching criteria.

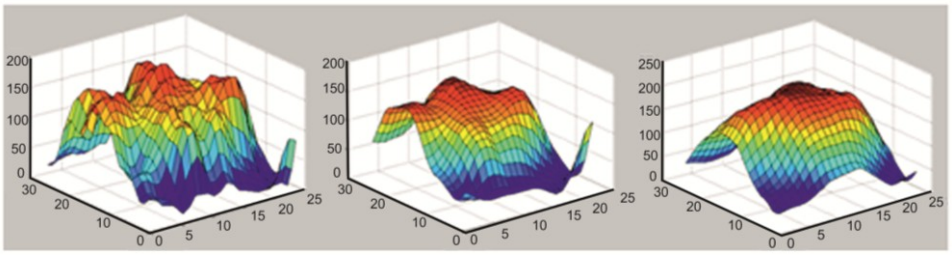
The method of using one channel [2] to recognize the survey forest characteristics made it possible to obtain a good result (with recognition accuracy of above 90%). And although this result is obtained for certain conditions and a limited range of characteristics of objects (trees), it is possible to get a better result.

Improving stability of the algorithm and accuracy of the estimates obtained is possible if the given data structure is improved, for example, by orthogonal transformation of the reference measurement matrix and using the above criteria for multivariate probability density functions. The improved statistical model can be applied to a wide class of natural objects. In particular, the technique can be used to obtain statistical reference and for other objects, for example, agricultural crops.

The research carried out in this area has shown that the PCA transformation alters the form of the probability density function. The distribution of the first image component for each vegetation type is close to normal. To describe the first component it is entirely justified to use normal distribution and apply the corresponding simplified recognition criteria. At the same time, the third component has the greatest difference between the probability density functions of different forest types. This leads to the conclusion that it is necessary to use multivariate probability density functions or construct a decision tree classifier that uses separate image components. Note that although multivariate functions are preferable, their practical production for reference functions is associated with technical difficulties.

The empirical research has also confirmed the conclusion that spectral measurements cannot uniquely characterize the type and other characteristics of forest. By measuring the spectral characteristics only (even in the case of ideal measurements free from the influence of external factors), the type of trees is not uniquely determined by these measurements. Accordingly, the linear transformation of these spectral measurements does not save from ambiguous solution, although at the same time it identifies the most important factors affecting the image of vegetation. Thus, in future it is necessary to determine additional characteristics that would help eliminate ambiguity of object description and improve recognition accuracy.

The suggested method can be used to solve other problems. For example, we studied various spectral descriptions of the image of individual trees (using the original images or transformed by the Gaussian operator (Fig.1)). Reference images of trees help determine most of survey forest characteristics from multispectral images.



a b c

Fig. 1. Spectral tree image using the Gaussian operator with various σ:
a) the original image, b) $\sigma = 1.5$, c) $\sigma = 2.0$.

It should be noted that when using the non-parametric approach a significant part of the recognition problem solution is transferred to the segmentation process, which is based on

using textural features of object images. (Note that the actual degree of homogeneity can be estimated using the probability density function.). If segmentation results in a high degree of homogeneity of the selected segment, the application of the non-parametric statistical method gives a high degree of object recognition (about 95%).

To improve the suggested approach and non-parametric statistical models it is necessary to:

1. determine the stable statistical characteristics of the brightness distribution of images of given classes of objects transformed in accordance with the a priori given probabilistic model of multispectral measurements – the method of principal components (PCA), independent components (ICA), Tasseled Cap (TC) or using other models;
2. choose the characteristic sections of the probability density function curve that determines the greatest class difference and develop appropriate criteria for model conformance evaluation;
3. choose the most informational characteristics and create effective space of features that provide the maximum for the criteria used that determine the differences in the probability density functions for the classes studied.

References

1. A.P. Guk, L.G. Evstratova, New statistical approach of forest image recognition, 14-16 (2016)
2. A.P. Guk, Izvestia vuzov Geodesy and Aerophotosurveying, Vol. **5**, 166–170 (2015)
3. K. Fukunaga, Introduction to statistical pattern recognition. (Academic Press, New York and London, 1972)
4. D.R. Cox, D.V. Hinkley, *Theoretical statistics* (Imperial college, London, 1974)