ISSN 0001-4338, Izvestiya, Atmospheric and Oceanic Physics, 2017, Vol. 53, No. 9, pp. 1132–1141. © Pleiades Publishing, Ltd., 2017.
Original Russian Text © V.V. Kozoderov, T.V. Kondranin, E.V. Dmitriev, 2016, published in Issledovanie Zemli iz Kosmosa, 2016, No. 6, pp. 45–55.

METHODS AND MEANS OF PROCESSING AND INTERPRETATION OF SPACE INFORMATION

Comparison Analysis of Recognition Algorithms of Forest-Cover Objects on Hyperspectral Air-Borne and Space-Borne Images

V. V. Kozoderov[a], *, T. V. Kondranin[b], and E. V. Dmitriev[c]

[a] *Moscow State University, Moscow, Russia*
[b] *Moscow Institute for Physics and Technology (State University), Dolgoprudny, Russia*
[c] *Institute of Numerical Mathematics, Russian Academy of Sciences, Moscow, Russia*
**e-mail: vkozod@mail.ru*

Received January 12, 2016

Abstract—The basic model for the recognition of natural and anthropogenic objects using their spectral and textural features is described in the problem of hyperspectral air-borne and space-borne imagery processing. The model is based on improvements of the Bayesian classifier that is a computational procedure of statistical decision making in machine-learning methods of pattern recognition. The principal component method is implemented to decompose the hyperspectral measurements on the basis of empirical orthogonal functions. Application examples are shown of various modifications of the Bayesian classifier and Support Vector Machine method. Examples are provided of comparing these classifiers and a metrical classifier that operates on finding the minimal Euclidean distance between different points and sets in the multidimensional feature space. A comparison is also carried out with the "*K*-weighted neighbors" method that is close to the nonparametric Bayesian classifier.

Keywords: hyperspectral air-borne and space-borne imagery, natural- and anthropogenic-object pattern recognition, optimization of data processing

DOI: 10.1134/S0001433817090171

INTRODUCTION

The automation of recognition of natural and man-made objects based on their texture and spectral features in the processing of hyperspectral air-borne and space-borne imagery requires the optimization of computing processes with a view to improve the efficiency of the use of computer tools. Based on the example of recognition of objects in the forest cover of different species composition and age using air-borne hyperspectral images of the test territory, the studies (Kozoderov et al., 2014a, 2014c) show the properties of the optimization of calculations. Optimization by texture features is reduced to finding the neighborhood of resolution elements (pixels) of the first, second, and higher orders using ideas about Markov random fields (short-range action) for a given class of objects. Optimization by spectral characteristics makes it easier to join neighboring channels without significantly reducing the accuracy of recognition of such objects. Cognitive remote-sensing technologies that are used in this case (Kozoderov et al., 2014b, 2015b) embody the experience of the previous development of this branch of science, which is related to the development of the alphabet of classes of objects, dictionaries of their signs, and mathematical procedures attributing the current pixel to those known a priori in machine-learning algorithms for data processing.

Along with the recognition of objects, cognitive remote sensing technologies (RS) also make it possible to restore the parameters of the state of recognized objects (Kozoderov et al., 2015a). Such parameters for forest-cover objects include the phytomass volume of foliage/needles and the related biomass volume of different fractions of forest vegetation (Kozoderov and Dmitriev, 2012a). To solve the emerging applied problems, a hardware–software system for processing the data of hyperspectral air-borne and space-borne sensing is created (Kozoderov et al., 2013a, 2014b). Validation elements (ground-based confirmation) of the information products obtained by processing air-borne and space-borne imagery are also processed (Kozoderov et al., 2015d). It becomes necessary to understand the information capabilities of different classifiers (computational procedures) (Kozoderov et al., 2015c).

The listed aspects of the solution of applied problems based on appropriate methods, algorithms, and software for the joint processing of RS data and ground-based surveys originate from the principles of pattern recognition (Tu and Gonzalez, 1978). A later publication (Schowengertt, 2010) notes the main

methods and models for the thematic classification of air-borne and space-borne imagery. Let us dwell in more detail on the optimization procedures of existing applications.

STATEMENT OF THE PROBLEM

At the initial stage, the contours (open and closed, true and false) of the objects on the optical image under consideration are constructed to define classes. Forest classes are represented by different species and ages of stands, which are perceived as alternating pixels belonging to insolated tops, shaded areas of the canopy, and intermediate cases where there are partially insolated and shadowed areas. Such a representation characterizes the random distribution of pixels of different forest classes on optical RS images.

The proposed method of recognition is based on the well-known principle of Bayesian classification. Let D be a set of features (these can be weighted spectral brightnesses and related values), L be a set of class labels (names), and a be the function acting from D to L (i.e., a classifier). Then the general form of algorithms of this type is represented as

$$a(d) = \arg\max_{l \in L} P(l)p(d|l), \quad (1)$$

where $d \in D$ is the element of the feature space, $l \in L$ is the class label, $P(l)$ is the a priori probability of the class l, and $p(d|l)$ is the probability distribution density of the features d that belong class l. The a priori probabilities of the classes are estimated on the basis of analysis of the image structure in the most contrasted spectral channel. This procedure allows us to increase the recognition accuracy for the main classes, such as water surfaces, bare soils, meadow vegetation, and stands.

The texture analysis used can be formulated as follows. For some selected image, the vision theory involves considering a discrete set S of m positions, labels that are represented as a discrete set L^* of the size J, and a set of standard brightnesses D^* of the size m, which allows us to determine the set of functions

$$f = \{f_{ij}\}_{i=1,j=1}^{m,J},$$

that act from $S \times L^*$ in D^*, i.e., $f : S \times L^* \to D^*$. Here and below, the index i corresponds to the nodes of the grid (which is considered spatially regular) and is responsible for the spatial distribution. The index j identifies the classes under consideration. Thus, the set $\{f_{ij}\}_{i=1}^{m}$ is a known texture reference that corresponds to the object j.

The joint probability $P\left(\{f_{ij}\}_{i=1}^{m}\right)$ and conditional probability $P\left(f_{ij}\left|\{f_{i'j}\}_{i'=1;\ i'\neq i}^{m}\right.\right)$ are used to determine the context in the labeling problem. If the labels are independent (the context is absent), then the joint probability is expressed by the product of local probabilities

$$P\left(\{f_{ij}\}_{i=1}^{m}\right) = \prod_{i \in S} P\left(f_{ij}\left|\{f_{i'j}\}_{i'=1;\ i'\neq i}^{m}\right.\right). \quad (2)$$

The labels become dependent due to the spatial influence of individual pixels on the image, and the Markov random field theory (MRF) allows us to describe such a context. The pair of "a pixel and its N-neighbors" is usually written in the form of a graph $(S, N) \sim G$ that contains the corresponding positions and neighborhood relations. This graph defines related connections as a subset of the set S. The connections can be considered for a single site, for their pair, triplet, etc., for regular and irregular nodes, thus forming the basis for describing the neighborhood between the pixels of the first, second, and higher orders.

Let $\{d_i\}_{i=1}^{m}$ be the recorded brightnesses on the image for the corresponding classes, which are characterized by pixels in the form of random fields. It would be convenient to introduce the spatial argument x and consider $\{f_{ij}\}_{i=1}^{m}$ and $\{d_i\}_{i=1}^{m}$ as the discrete values of continuous functions $d(x)$ and $f_j(x)$. It is typical in pattern recognition to make a decision based on finding the maximum of a posteriori probability (MAP)

$$j^* = \arg\max_{j \in L^*} P(d(x)|f_j(x))P(f_j(x)). \quad (3)$$

The conditional probability $P(d(x)|f_j(x))$ is usually modeled on the basis of the Gaussian distribution using the maximum likelihood method and a priori probability by the MRF methods. The solution to the problem of marking the MAP-MRF is based on analytical properties of the theoretical distribution $f_j(x)$ in the case of a comparison with the recorded brightnesses $d(x)$, and the estimate of MAP is usually equivalent to the minimization of a posteriori energy for the corresponding class on the image.

Problems for a string, core, and other similar problems discussed below are classical in the applications of functional analysis for the statistical theory of automated systems. This branch of science is very popular in scientific studies related to programming. Its essence consists in understanding the behavior of systems when considering not only the distribution functions, but also their derivatives.

If we consider the Gaussian noise model,

$$d(x) = f_j(x) + \varepsilon(x), \quad (4)$$

for which the observation is represented as "truth" plus an independent Gaussian noise, which is characterized as $\varepsilon(x)$. It is implied that there are mean and standard deviations for the noise. Analytic properties of the functions $f_j(x)$ must also be taken into account.

Our approach (Kozoderov et al., 2015b) to considering the neighborhood of second-order pixels (the eight-point scheme within the window for each pixel with allowance for the diagonal elements, i.e., the "core") differs from the traditional first-order approach (the four-point scheme: pixels within the window on the grid on the left, on the right, above, and below, i.e., the "string"). Then, when minimizing a posteriori energy, we can use, along with the likelihood term, a regularizer in the form of an integral of the square of the derivative of the theoretical (trial) distribution function.

The emerging smoothness of the a priori constraints results in finding a regularized solution (in general, a solution may be not unique which requires such regularization) for a posteriori energy, which must be minimized. The category of such energy consists of two components that correspond to the a priori hypothesis and the difference between the theoretical distribution function and recorded brightnesses.

Thus, we must find

$$j^* = \arg\min_{j \in L^*} E\left(f_j(x)\right), \tag{5}$$

where a posteriori energy is expressed as follows:

$$E(f_j) = \sum_{i \in S^*} \left[f_{ij} - d_i\right]^2 + \lambda \int_a^b \left[f_j^{(n)}(x)\right]^2 dx. \tag{6}$$

Here set S^* contains indices of training data in points i from a certain class, $\lambda \geq 0$ is the weight coefficient between these two terms of the given quadratic representation of the distribution function on a certain interval of integration $[a, b]$, and $n \geq 1$ is the order of the derivative function. The first term in (6) characterizes the measure of proximity of the solution and initial data, and the second term is the regularizer itself that introduces an a priori restriction on the smoothness of the functions used.

If we consider a posteriori energy in the form

$$E(f_j) = \sum_{i \in S} \left[f_{ij} - d_i\right]^2 + \lambda \sum_{i \in S} \sum_{i' \in N_i} g\left(f_{ij} - f_{i'j}\right) \tag{7}$$

(the N_i set of neighbors of the node i), where an additional restriction is introduced for the derivative of the new g function

$$\lim_{\eta \to \infty} \left|g'(\eta)\right| = \text{const} < \infty, \tag{8}$$

we come to minimizing the energy in the Sobolev functional space W_2^n, each point of which is assigned by the function $f_j(x)$; the $(n-1)$th derivative $f_j^{(n-1)}(x)$ is absolutely continuous; and the nth derivative $f_j^{(n)}(x)$ is quadratically included in the integral. The transition to this functional space opens up new ways to solve problems of optimizing texture modeling and recognition based on the proposed image-processing methods, which were implemented for computations on high-performance computers.

Returning to algorithm (1), we determine the probability distribution densities of features on the basis of the Gaussian mixture model

$$p\left(d(x) \middle| \{w_k(l), \mu_k(l), \Sigma_k(l)\}_{k=1}^M\right) = \frac{1}{\sqrt{(2\pi)^\Omega}} \sum_{k=1}^M \frac{w_k(l)}{\sqrt{\det(\Sigma_k(l))}} \times \exp\left[-\frac{1}{2}(d(x) - \mu_k(l))^T \Sigma_k^{-1}(l)(d(x) - \mu_k(l))\right], \tag{9}$$

where $d(x)$ is the vector of the features of the size Ω that belongs to the set D introduced above; M is the number of components of the mixture; and parameters w_k, μ_k, and Σ_k are the weight coefficients, expectation vectors, and covariance matrices for each kth component of the mixture within class l, respectively. The feature space consists of weighted spectral brightnesses, which are normalized to an integral value according to a wavelength.

The estimates of the parameters $w_k(l)$, $\mu_k(l)$, and $\Sigma_k(l)$ are based on a priori information in the form of a set of spectra and corresponding names of classes that make up the learning set. This set is extracted from hyperspectral images of areas with the predominant homogenous species and age composition of stands. The corresponding locations are determined using the cadastral inventory data for the territories. The data set makes it possible to consider the problem of recognizing species and age variability of stands.

Let us introduce the indicator function for the event A

$$I(A) = \begin{cases} 1, & A \text{ is true}, \\ 0, & A \text{ is false}. \end{cases} \tag{10}$$

Then the estimate of the probability of erroneous classification for the class j can be expressed by the formula

$$\nu_j = 1 - \frac{1}{m} \sum_{i=1}^m I\left(l_{ij} = a(d_i)\right), \tag{11}$$

where l_{ij} is the label of the node i.

This is the statement of the basic model of the improved Bayesian classifier, which, together with the quadratic discriminant analysis, corresponds to the problems of statistical pattern recognition on the basis of hyperspectral sensing data. However, the model has limitations that are related to the requirement to have a representative ensemble of selected data for the implementation of the approach under consideration. An alternative to the use of small samples for classifier training is sought in the support vector method (Vapnik and Chapelle, 2000) in solving the minimax prob-

lem of searching for the saddle point known in mathematics, which characterizes the unique solution of the emerging quadratic programming problem. First, the "margin" between two hyperplanes in the multidimensional feature space is considered for the linear separation of peripheral objects, objects between these planes, and objects–violators that do not fall "in their class." Each measured spectrum is represented by a point in this feature space. Then, nonlinearity is introduced by replacing scalar products with kernel functions. The properties of this approach are expounded in the training manual (Kozoderov et al., 2013b). Other approaches to the recognition of objects by their multispectral/hyperspectral images are also expounded there. For example, of interest is the simplest metric classifier (Yuan et al., 2012), which is based on finding the minimal Euclidean distance between individual points of the feature space or between the sets of such points.

Completing the description of the basic model, we note other possibilities for improving the efficiency of computational methods, algorithms, and software for processing hyperspectral sensing data. The principal component method (Jolliffe, 2002) of decomposition of autocovariance matrices by eigenvectors ((EOF) empirical orthogonal functions) serves as an example of the possibility of reducing the dimensionality of the feature space.

Another way to reduce the feature space is based on the component-wise selection of the most informative features (in the English-language literature it is called the "step up method") (Fukunaga, 1990). The essence of the method is as follows: the recognition method and a priori probability for recognized classes are selected. The space of all features is divided into two groups—the features accepted in the model and the remaining features. For each feature from the set of the "remaining features," a recognition error is estimated if this feature is added to the model. The minimum error is selected from the obtained set of recognition errors and is compared with the error of the previous model. If a significant decrease in the error occurred, then the corresponding feature is accepted in the model; in the opposite case, the process stops.

Finally, let us also consider a classifier known as the *K*-weighted neighbors method (Cost and Salzberg, 1993). This method is interesting in that it is very close to the Parzen window method (the nonparametric Bayesian classifier) (Parzen, 1962). In the multidimensional case, the widths of the Parzen windows are usually chosen differently for different basic directions. The *K*-weighted neighbors method also makes it possible to implement this possibility, but in a more complicated way, by using nonisotropic distances. Despite the noted proximity of these classifiers, these methods are not completely identical.

Let us consider the listed examples of different classifiers and their modifications for solving the problem of recognition of objects by their hyperspectral air-borne and space-borne imagery on the basis of flight tests of the domestic hyperspectral equipment, which was developed by the Lepton Scientific Production Organization in Zelenograd (Kozoderov et al., 2012).

DISCUSSION

One of the areas for flight tests of hyperspectral equipment is imaged in Fig. 1. The contours of main objects can be identified based on the RGB image. Open soils and various types of vegetation are contrasted in the 637-nm and 827-nm channels, respectively.

Corrected hyperspectral images are considered in which the filtration of radiometric interference (bands along the flight trajectory) was performed based on the statistical polynomial method, maintaining the mathematical expectation of the calibration coefficients for this equipment. Figure 2 shows the eigenvalues of the autocovariance matrix for the illustrated image, which correspond to the variances of the coefficients of decomposition on the basis of the EOF. The variance of the instrumental noise of the hyperspectral camera depends on the signal level and varies in the measurement range from approximately 10^{-8} to 10^{-10} if presented in brightness units (spectral-energy brightness squares). It can be seen that the fifth principal component (indicated by the arrow in Fig. 2) is close to the noise level of the equipment.

The data in Fig. 3 reproduce the four most informative components of the main objects on the earth's surface. These components contain 97% of the total variance of objects. The color scales on the right in Fig. 3 characterize the effect of these variances. The predominant red color of component 1 between scales 0 and 10^{-3} indicates the prevalence of vegetation on this scene. The predominance of the blue color for component 2 between scales -1 and 1 demonstrates the response of this component to the spatial variability of this scene. Components 3 and 4 also carry additional information on the spatial variability of the objects on the earth's surface. Starting with component 5, distortions of the decomposition coefficients become significant due to the intrinsic noise of the equipment. It was found that the contribution of components 8 and 12 did not exceed 0.1 and 0.01%, respectively.

Another scene of the total test territory is shown in Fig. 4. The site includes significant areas of water surface, open sandy soil, and vegetation cover (mainly pine and birch mixed stands). One particular difficulty in the classification is caused by a complex orography of the sandy area in the lower right corner of the image. The area consists of two large sandy mounds separated by a dirt road and sparse vegetation.

The RGB-synthesized image (Fig. 4a) makes it possible to notice an artifact of attributing the shaded

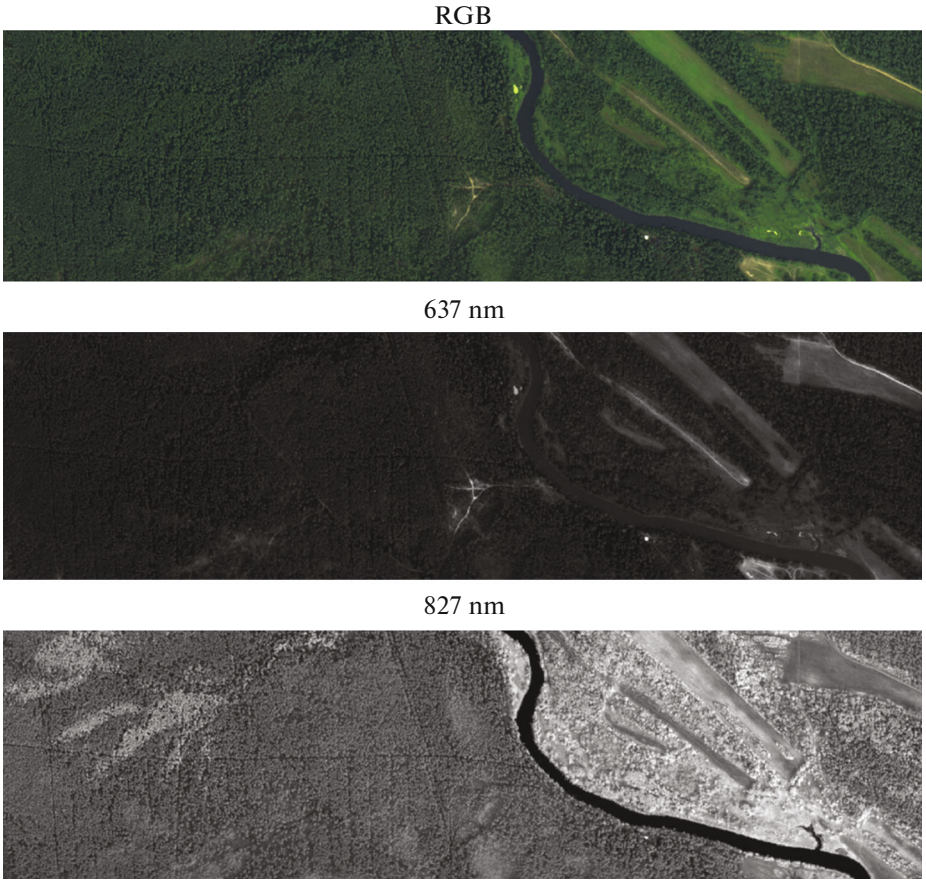

Fig. 1. RGB and single-channel images of the test area.

slope of a sandy mound as a road surface. In general, we can note the similarity of the classification results for the two presented methods (Figs. 4b, 4c). However, the classification results based on the *K*-weighted neighbors method (Fig. 4c) show that the use of more complex classifiers increases the accuracy of recognition. We can see that an artifact with a false recognition of a pavement at the edge of a mound is absent in Fig. 4c in comparison with Fig. 4b. Unclassified areas marked with the violet color are also absent in the area of the sandy mounds. Many more pixels of the road across the selected site are recognized. The pixels that are falsely classified as herbaceous plants in the border zones are significantly fewer in number. Nevertheless, it should be noted that data processing by the metric methods is performed at a much higher rate than in case of using the *K*-weighted neighbors method, even with consideration for acceleration due to incomplete search.

Let us give a few more examples of comparing the results of the above-described methods with more complex classifiers. Table 1 presents the mutual comparison of the metric classifier on the basis of the Euclidean distance and various types of Bayesian classifiers (BC) and the support vector method (SVM). Other designations are as follows: g. cm is the Gaussian mixture of pixels; lin. is the linear classifier; norm. is the normal distribution; sq. is the quadratic kernel; g. is the Gaussian kernel. The share of coinciding object names was used as a proximity measure.

Figure 5 presents the results from recognition of seven main classes of this scene: water surface (blue color), pavement (black color), soil (pink color),

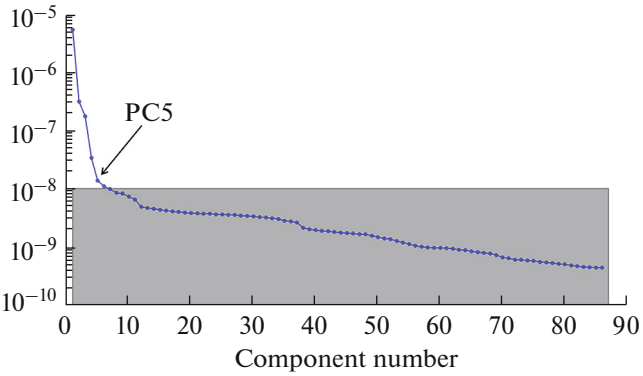

Fig. 2. Eigenvalues of the autocovariance matrix of a corrected hyperspectral image.

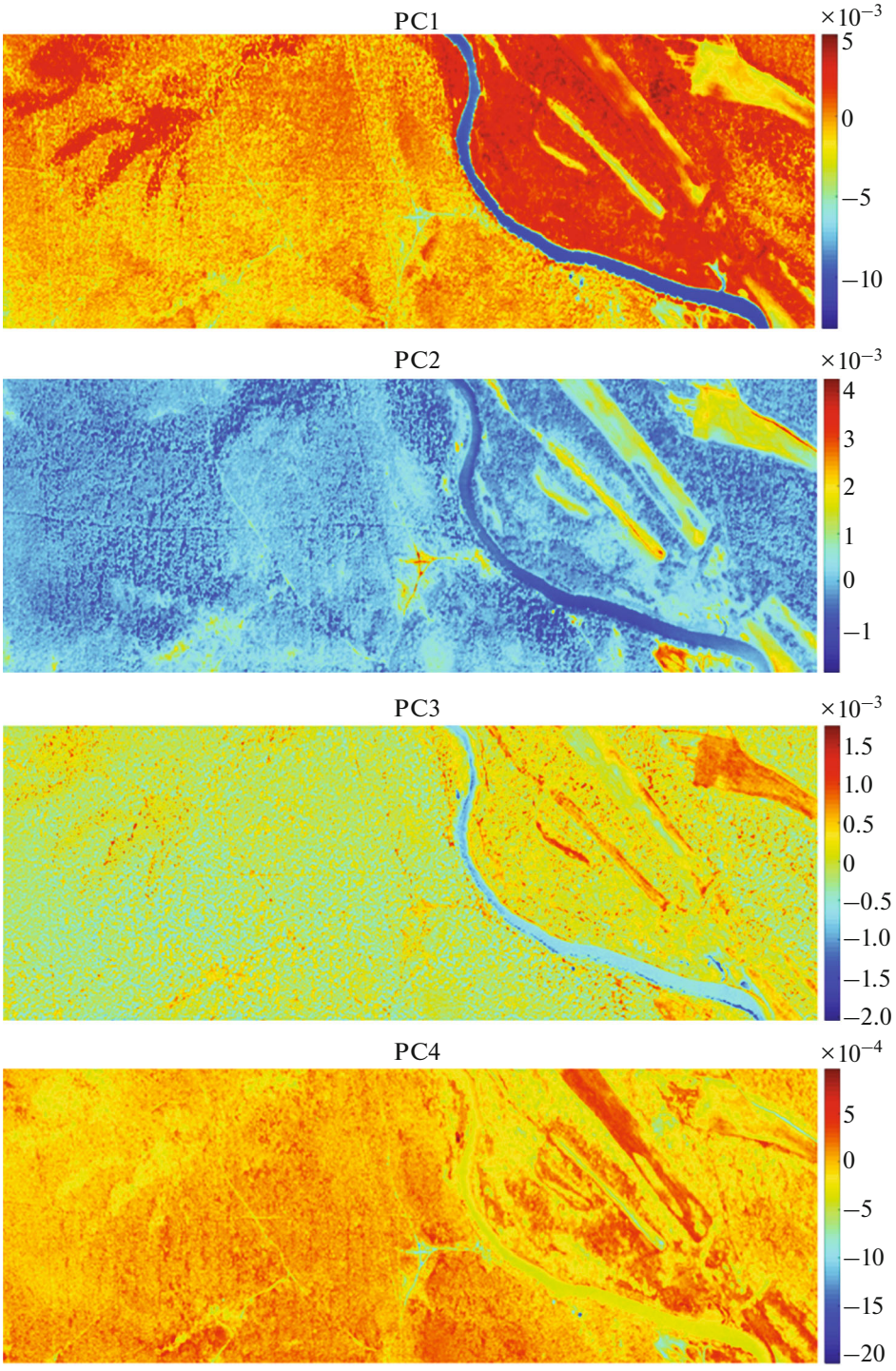

Fig. 3. Informative components of decomposition of a corrected normalized hyperspectral image on the basis of the EOF.

meadow vegetation (red color), pine (dark green color), birch (light green color), and aspen (orange color). An image in the IR channel (Fig. 5a), where these classes have a good contrast, is given for visual validation.

The greatest similarity is immanent to the BCs based on the normal distribution and Gaussian mixtures and the SVM with the Gaussian and quadratic kernel. These methods also show the greatest accuracy of classification. The results of the SVM show an improvement in the accuracy of classification of meadow vegetation in comparison with the BCs, but the accuracy of classification of the stand species composition is reduced. In addition, there are insig-

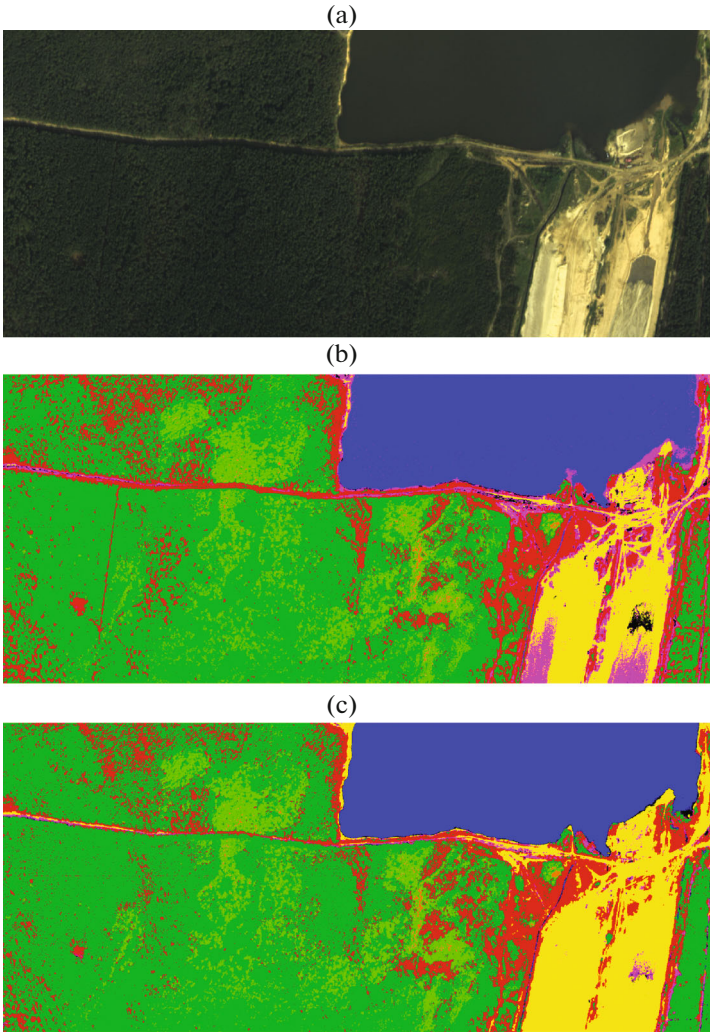

Fig. 4. RGB-synthesized hyperspectral image (a) and results from recognizing objects on this image using the metric classifier (b) and based on the *K*-weighted neighbors method (c). Blue is water, yellow is sand, black is road pavement, dark green is pine stands, light green is birch stands, orange is aspen stands, red is grassy vegetation, and violet is other objects.

nificant cases of false classification of the road surface for the SVM.

The linear SVM (Fig. 5d) proves to be almost unsuitable for solving the classification problem and is the worst of the methods considered. The MC (metric classifier) (Fig. 5b) also leads to significant errors and can only be used for classification at the "qualitative" level. In contrast to the linear SVM, the linear BC gives acceptable results. This takes place, in particular, due a restriction being imposed on a posteriori proba-

Table 1. Similarity of the results from classification by different methods

	MC (Euclid)	BK (g.cm.)	BK (lin.)	BK (norm)	SVM (lin.)	SVM (sq.)	SVM (g)
MC (Euclid)	1	0.7	0.59	0.7	0.39	0.62	0.64
BK (g.cm.)	0.7	1	0.6	0.92	0.34	0.74	0.76
BK (lin.)	0.59	0.6	1	0.61	0.38	0.56	0.55
BK (norm)	0.7	0.92	0.61	1	0.35	0.75	0.76
SVM (lin.)	0.39	0.34	0.38	0.35	1	0.31	0.28
SVM (sq.)	0.62	0.74	0.56	0.75	0.31	1	0.85
SVM (g)	0.64	0.76	0.55	0.76	0.28	0.85	1

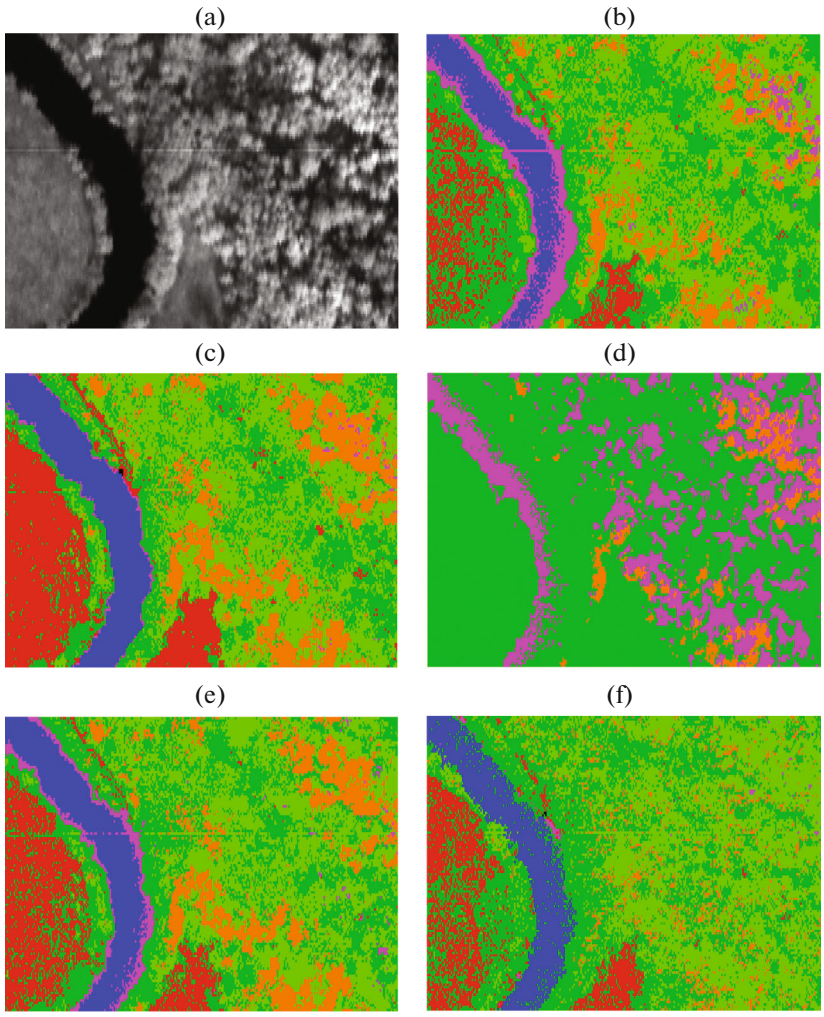

Fig. 5. Results from classifying a hyperspectral image using different methods: (a) image in the spectral channel of the near infrared region, (b) metric classifier (the Euclidean distance), (c) SVM with the Gaussian kernel, (d) SVM with the linear kernel, (e) BC based on the model of Gaussian mixtures, and (f) linear normal BC.

bility. Nevertheless, the linear BC gives significantly less accuracy than the nonlinear modifications of the SVM and BC.

One more example concerning the accuracy of recognition of species composition by these methods is shown in Fig. 6. Four methods are considered: the SVM with the Gaussian kernel, the metric classifier based on the Euclidean distance, the Bayesian classifier based on Gaussian mixtures, and the *K*-weighted neighbors method. The pixels of forest canopy were classified making allowance for the gradation of illumination.

The white lines show the contours of stand areas according to the available forestation data. The species composition is indicated for each stand area with a white text. As we can see, the areas contain only pine and birch stands. In general, all the presented four methods have coped well with the task of recognizing the species composition. Since the forestation data that are used for validation are rather rough, it is difficult to compare the quantitative characteristics of errors on their basis (Kozoderov et al., 2015d). However, an indirect method can be proposed.

As we can see, some pixels were falsely classified as aspen stands. The forest taxation data have an intrinsic error of about 10%, but it is reliably known that there are no aspen stands in the first tier in this site. The number of false-classified pixels for each algorithm is represented by the numbers in the captions above the figures.

From this standpoint, the worst of the presented methods is the metric classifier. The best method is the BC based on Gaussian mixtures. The SVM and *K*-weighted neighbors method have somewhat larger but commensurate errors. It can be concluded that, of all the methods considered above, the *K*-weighted neighbors method has an accuracy commensurable with the accuracy of the nonlinear optimal classifiers

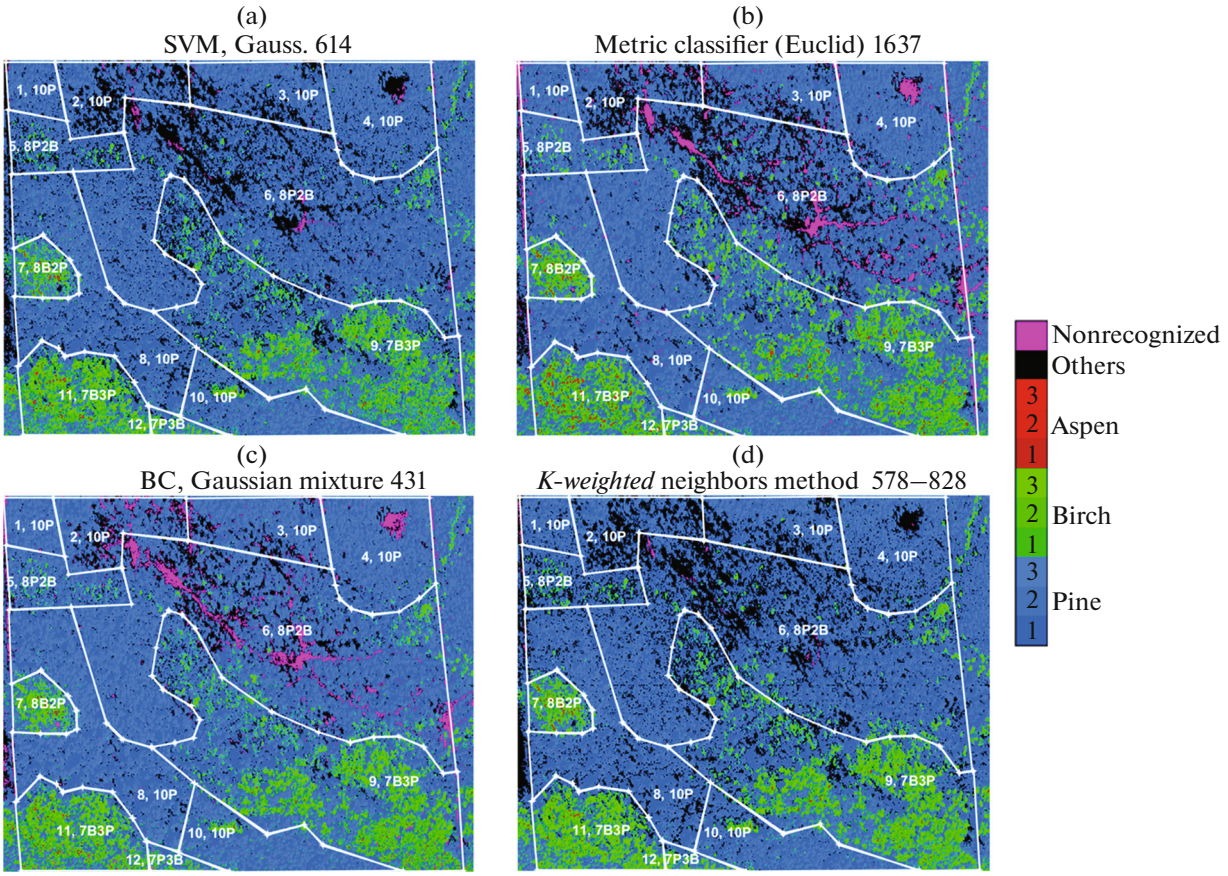

Fig. 6. Results from recognizing the species composition of stands with consideration for the gradations of illumination using different methods: (a) SVM with the Gaussian kernel, (b) metric classifier (the Euclidean distance), (c) BK based on the model of Gaussian mixtures, and (d) *K*-weighted neighbors method. The ciphers in the captions above the figures designate the number of pixels with a significantly mistaken classification. The ciphers in the color scale designate the gradations of illumination, varying from completely shaded crowns (1) to crowns insolated with direct sunlight (3).

and can be equally used for the applied problems of recognizing the forest cover of different species composition and age.

CONCLUSIONS

The basic model for recognizing forest-cover objects of different species composition and age was developed in the problem of processing the hyperspectral air-borne and space-borne sensing data. The effectiveness of the classification methods in the problem of hyperspectral RS of natural and anthropogenic objects has been analyzed. The properties of the implementation of the metric classifiers, parametric Bayesian classifiers, and the multiclass support vector method have been discussed. The results from classifying the hyperspectral air-borne imagery by the indicated methods in the selected test territory have been demonstrated. The results from the comparative analysis of different classifiers, as well as the advantages of using nonlinear classifiers, have been shown. The similarity of the results given by some modifications of the support vector method and Bayesian classification has been noted.

ACKNOWLEDGMENTS

This work was supported by the Russian Science Foundation (project 16-11-00007), Federal Target Program "Research and Development of Priority Areas of the Russian Science and Technology Complex for 2014–2020 (agreement 14.575.21.0028, unique identification number RFMEFI57514X0028), and grants from the Russian Foundation for Basic Research (project numbers 14-05-00598, 14-07-00141).

REFERENCES

Cost, S. and Salzberg, S., A weighted nearest neighbor algorithm for learning with symbolic features, *Mach. Learn.*, 1993, vol. 10, no. 1, pp. 57–78.

Fukunaga, K., *Introduction to Statistical Pattern Recognition*, New York: Academic, 1990.

Jolliffe, I.T., *Principal Component Analysis*, Springer, 2002.

Kozoderov, V.V. and Dmitriev, E.V., Remote sensing of forest cover: An innovative approach, *Vestn. Mosk. Gos. Univ. Lesa–Lesn. Vestn.*, 2012a, no. 1, pp. 19–33.

Kozoderov, V.V., Kondranin, T.V., Dmitriev, E.V., Kazantsev, O.Yu., Persev, I.V., and Shcherbakov, M.V., Processing of hyperspectral aerospace sounding data, *Issled. Zemli Kosmosa*, 2012, no. 5, pp. 3–11.

Kozoderov, V.V., Dmitriev, E.V., and Kamentsev, V.P., System for processing of airborne images of forest ecosystems using high spectral and spatial resolution data, *Issled. Zemli Kosmosa*, 2013a, no. 6, pp. 57–64.

Kozoderov, V.V., Kondranin, T.V., and Dmitriev, E.V., *Metody obrabotki mnogospektral'nykh i giperspektral'nykh aerokosmicheskikh izobrazhenii. Uchebnoe posobie* (Methods of Processing of Multispectral and Hyperspectral Aerospace Images: A Textbook), Moscow: MFTI, 2013b.

Kozoderov, V.V., Kondranin, T.V., and Dmitriev, E.V., Recognition of natural and man-made objects in airborne hyperspectral images, *Izv., Atmos. Ocean. Phys.*, 2014a, vol. 50, no. 9, pp. 878–886.

Kozoderov, V.V., Dmitriev, E.V., and Kamentsev, V.P., Cognitive technologies for processing optical images of high spatial and spectral resolution, *Atmos. Oceanic Opt.*, 2014b, vol. 27, no. 6, pp. 558–565.

Kozoderov, V.V., Kondranin, T.V., Dmitriev, E.V., and Sokolov, A.A., Retrieval of forest attributes using optical airborne remote sensing data, *Opt. Express*, 2014c, vol. 22, no. 13, pp. 15410–15423.

Kozoderov, V.V., Kondranin, T.V., Dmitriev, E.V., and Kamentsev, V.P., A system for processing hyperspectral imagery: Application to detecting forest species, *Int. J. Remote Sens.*, 2014d, vol. 35, no. 15, pp. 5926–5945.

Kozoderov, V.V., Dmitriev, E.V., and Sokolov, A.A., Improved technique for retrieval of forest parameters from hyperspectral remote sensing data, *Opt. Express*, 2015a, vol. 23, no. 24, pp. A1342–A1353.

Kozoderov, V.V., Kondranin, T.V., Dmitriev, E.V., and Kamentsev, V.P., Bayesian classifier applications of airborne hyperspectral imagery processing for forested areas, *Adv. Space Res.*, 2015b, vol. 55, no. 11, pp. 2657–2667.

Kozoderov, V.V., Dmitriev, E.V., and Sokolov, A.A., Cognitive technologies in optical remote sensing data processing, *Clim. Nature*, 2015c, no. 1, pp. 5–45.

Kozoderov, V.V., Kondranin, T.V., Dmitriev, E.V., and Kamentsev, V.P., Validation of information products of processing of aircraft hyperspectral images, *Issled. Zemli Kosmosa*, 2015d, no. 1, pp. 32–43.

Parzen, E., On the estimation of a probability density function and the mode, *Ann. Math. Stat.*, 1962, vol. 33, no. 3, pp. 1065–1076.

Shovengerdt, R.A., *Distantsionnoe zondirovanie. Modeli i metody obrabotki izobrazhenii* (Remote Sensing. Models and Methods of Image Processing), Moscow: Tekhnosfera, 2010.

Tou, J. and Gonzales R., *Pattern Recognition Principles*, Reading, MA: Addison-Wesley, 1974; Moscow: Mir, 1978.

Vapnik, V. and Chapelle, O., Bounds on error expectation for support vector machines, *Neural Comput.*, 2000, vol. 12, no. 9, pp. 2013–2036.

Yuan, G.-X., Ho, C.-H., and Lin, C.-J., Recent advances of large-scale linear classification, *Proc. IEEE*, 2012, vol. 100, no. 9, pp. 2584–2603.

Translated by L. Solovyova