Recognition of Forest Species
and Ages Using Algorithms Based
on Error-Correcting Output Codes

Egor V. Dmitriev[a], Vladimir V. Kozoderov[b],
Alexander O. Dementyev[a] and Anton A. Sokolov[c]
[a]Institute of Numerical Mathematics RAS
8 Gubkina Str., Moscow, 119333, Russia
[b]M.V. Lomonosov Moscow State University
1 Leninskiye Gory, Moscow, 119991, Russia
[c]Laboratoire de Physico-Chimie de l'Atmosphère Université
du Littoral Cote d'Opale,
189A Avenue Maurice Schumann, Dunkerque, 59140, France

The basic model of the recognition of forest inventory characteristics using spectral features is represented in the framework of the problem of hyperspectral airborne imagery processing. The algorithm of multiclass supervised classification based on the error-correcting output codes underlies this model. The support vector machine method is used as the necessary binary classifier. The method of the construction of training set by using mixed forest plots is represented. Results of the retrieval of species and age composition of forest stands from hyperspectral images are represented for the selected test area. The estimate of accuracy of the retrieval of the mixed forest composition is comparable with the accuracy of ground-based forest inventory data.

Keywords: Remote sensing, pattern recognition, spectral classification, hyperspectral measurements.

*   Corresponding author E-mail address: yegor@mail.ru, vkozod@mail.ru

Распознавание породного и возрастного состава древостоев с использованием алгоритмов на основе самокорректирующихся кодов

Е.В. Дмитриев[а], А.О. Дементьев[б],
В.В. Козодеров[а], А.А. Соколов[в]

[а]Институт вычислительной математики РАН
Россия, 119333, Москва, ул. Губкина, 8
[б]Московский государственный университет имени М.В. Ломоносова
Россия, 119991, Москва, Ленинские горы, 1
[в]Университет Литтораль
Франция, 59140, Дюнкерк, Авеню Морис Шуман, 189A

Представлена базовая модель распознавания таксационных характеристик древостоев по спектральным признакам в проблеме обработки гиперспектральных авиационных изображений. Основу модели составляет алгоритм многоклассовой обучаемой классификации с использованием самокорректирующихся кодов. В качестве необходимого метода бинарной классификации применен метод опорных векторов. Описана методика использования выделов со смешанным породным составом для построения обучающего ансамбля. Приведен пример восстановления породного и возрастного состава для выбранного тестового участка по данным гиперспектральных измерений. Оценка точности восстановления породного состава соответствует точности наземных данных лесотаксации.

Ключевые слова: дистанционное зондирование, распознавание образов, спектральная классификация, гиперспектральные измерения.

Introduction

The one of the important tasks of modern forest management is to obtain and update forest inventory data and related cartographic products. The main and most accurate source of information is the data of ground-based survey. Obtaining this kind of information is associated with significant time and financial expenses. Thus, remote forest inventory methods are actively developed in recent years. The current program of forestry development suggests that remote methods will provide about 80% of forest inventory data. Multispectral and hyperspectral aerospace images in the visible and near infrared spectral region are usually used to carry out the monitoring of vegetation cover.

The greatest number of applications relates to processing of multispectral images of medium and low spatial resolution (30-500 m). A significant part of studies follow the line of creation and modification of vegetation indices, which are used for the study of chemical processes associated with photosynthetic activity and the efficiency of using light by the plant to assess the content of pigments in the leaves. Despite all efforts, the possibilities of using vegetation indices are severely limited. Several studies have shown that the existing indices are not unambiguous measure of the content of pigments and are not always sensitive to the parameters describing the state of vegetation.

One of the most important problems, which frequently causes discussions at different scientific events on the remote sensing, deals with obtaining the particular practical results proofing the advantage

of the use of hyperspectral remote sensing data in comparison with multispectral data. There are different opinions on this matter. Let us give some results on the comparison of multispectral and hyperspectral approaches published in different scientific journals.

Comparison of estimates of the crop biomass obtained using hyperspectral EO-1/Hyperion images and multispectral images from IKONOS, GeoEye-1, WorldView-2, Landsat/ETM+ and MODIS having different spatial resolution is represented in [1]. For the analysis authors used the standard two-channel vegetation indices and multichannel indices constructed with the use of the singular decomposition in the initial data and the stepwise regression. The results obtained showed that the vegetation indices obtained from hyperspectral data allowed the achievement of higher accuracy for the considered crops then the indices obtained from multispectral data.

The problem of mapping the chlorophyll content for Mediterranean pine plantations in Spain is considered in [2]. Creation of maps of physiological conditions is useful for the recognition of stressed forest plots within healthy stands and it could play a critical role in the early detection of forest decline processes enabling their application in precision forestry. Hyperspectral and multispectral images (AHS, CHRIS/Proba, Hyperion, Landsat and QuickBird) with different spatial resolutions were used for the analysis with the red-edge and the simple ratio vegetation indices and PROSPECT-5 + DART modeling. It was shown that the data of high-resolution hyperspectral CHRIS/Proba and Hyperion sensors are feasible for the retrieval of vegetation biochemical parameters and multispectral images revealed less reliable but also significant results for obtaining biophysical variables.

The problem of segmentation crowns of individual trees and accurate classification of rocks arose with increasing spatial and spectral resolution of remote sensing instruments. The spatial resolution of the satellite hyperspectrometers does not exceed 30 m, due to technical constraints. Therefore, hyperspectrometers mounted on an aircraft can be used for obtaining more detailed regional information. The relatively low altitude flight provides the opportunity to combine high spatial and spectral resolution, and can significantly reduce the distorting effects of the atmosphere.

The forest monitoring implies mapping of species composition and types of stands. These new possibilities can be realized by employing latest remote sensing systems. The investigation described in [3] is aimed to contribute to understanding the innovation in forest and ecological research and proposes the guild mapping approach as a tool for the efficient monitoring the varied tropical forest resources, namely, the forests of in southwest part of Ghana: wet evergreen forests in Ankasa Conservation Area and moist semi-deciduous forest in Bia Conservation Area. Airborne hyperspectral images obtained from AISA Eagle sensor and simulated multispectral Sentinel-2 data were used for the recognition of dominant species and distinguishing between two available forest types. The classification results demonstrated significant improvement of the total accuracy for the Support Vector Machine over the Maximum Likelihood method. In this paper we also touch upon the comparison these two approaches.

Materials and methods

Let us consider the problem of recognition species and age composition of the forest stands on the basis of hyperspectral images of the territory of Savvatevskoe forestry (Tver, Russia) which were obtained in August 2011 during the aircraft measurement campaign using domestic hyperspectral

system produced by NPO Lepton (Zelenograd, Moscow, Russia). The AV-VD hyperspectrometer used has a swath width of 500 pixels and produces images in 287 spectral bands in the range of 401-1017 nm. The spectral resolution of channels is reduced to 5 nm, which corresponds to 87 combined spectral bands. The flight was at an altitude of about 2 km, thus the spatial resolution across the flight path is 1.1 m. The species composition of the forest on the test territory is represented by pine, birch, aspen and spruce. The detailed description of this measurement campaign is given in [4].

We used the supervised classification algorithm based on the technique of error correcting output codes (ECOC) [5] for solving the problem of recognition of the species and age composition of forest stands. This method uses several approaches from the information coding theory for the formalization of extending binary classifiers to the multiclass case.

Let $x \in R^M$ are features of classes and $x \in R^1$ are names of classes. Let us also introduce the matrix $C = (c_{ij})$ (coding design matrix), whose columns correspond to different ways of learning a binary classifier (or, better to say, some problems of binary classification, based on the available classes) and lines – to recognized classes. Elements of the matrix take values: "1" – the object, "–1" – the alternative, "0" – other objects which are not involved in the classification. Thus, each class is assigned to a code (a row of the matrix C). The following restrictions take place during the construction of the matrix C:

– permissible values of the elements are only -1, 0 and 1;
– each column includes at least one element 1 and one element -1;
– any two different column vectors are not collinear;
– any two different row vectors are not the same;
– for any two classes there must be a column in which these classes can be separated.

Suppose that we have a training set of N pairs of elements x_i and y_i, $(i = 1,...,N)$, and we need to classify a new measurement x. Let us introduce the function $s(x)$, defining calculation of classification score for the measurement x, in accordance with the considered binary learner. For the support vector machine, $s(x)$ indicates the distance from the measurement x to the discriminant surface in the area of relevant class. We also introduce the loss function $g(y_i, s(x))$, for a class y_k from the source list. Then it is possible to construct a multiclass classification algorithm, which can be expressed as following

$$a(x) = \arg\min_{y_k} \frac{\sum_{j=1}^{L} \left| c_{kj} \right| g(y_k, s_j(x))}{\sum_{j=1}^{L} \left| c_{kj} \right|},$$

where L is the number of binary learners (columns of the coding design matrix). Thus, the algorithm selects the class corresponding to the minimum average loss for a given measurement x.

The result of the algorithm depends on the choice of the coding matrix. For example, the matrix

$$C = \begin{pmatrix} 1 & -1 & -1 \\ -1 & 1 & -1 \\ -1 & -1 & 1 \end{pmatrix}$$

corresponds to the known "one-vs-all" strategy as applied to the problem of recognition of some three classes, and the matrix

$$C = \begin{pmatrix} 1 & 0 & -1 \\ -1 & 1 & 0 \\ 0 & -1 & 1 \end{pmatrix}$$

corresponds to the "one-vs-one" strategy in the same problem. There are also several others standard coding design matrices. The "ordinal" coding matrix consists of upper and lower triangular matrices corresponding to the objects 1 and -1. This matrix provides the high training and classification speed however on the other hand it has a small distance between codes of neighboring classes. The opposite situation occurs for the "binary complete" and "ternary complete" coding designs, which provide well distinguishable but long enough codes of recognized classes. Thus, we need to find a balance between the classification speed and accuracy.

One of the most promising approaches to the construction of optimal coding design matrices is the method of random matrices, which can be formulated as follows. We select number L (it is usually proportional to the logarithm of the number of classes) and randomly generate a large number of coding matrices, in accordance with the rules formulated above. From these matrices we chose one having the largest possible minimum pairwise distance between lines in terms of the Hamming metric, i.e.,

$$D(k_1, k_2) = 0.5 \sum_{j=1}^{L} \left| c_{k_1 j} \right| \left| c_{k_2 j} \right| \left| c_{k_1 j} - c_{k_2 j} \right|.$$

In this paper, the non-linear support vector machine (SVM) [6] is considered as the basic binary classifier. Linear SVM allows finding the most distant parallel hyperplanes separating the given pair of classes, passing through the area boundary points (support vectors) of feature distributions of these classes. If the equation of separating hyperplanes is represented as $(w, x_i) - w_0 = y_i$, the optimization problem of the SVM with the "soft" margin can be represented as

$$\begin{cases} \dfrac{1}{2}(w, w) + C \sum_{i=1}^{N} \xi_i \to \min_{w, w_0, \xi} \\ y_i((w, x_i) - w_0) \geq 1 - \xi_i, \quad i = 1, \ldots, N \end{cases},$$

where parameters $\xi_i \geq 0$ are penalties for misclassification of boundary points. The algorithm for linear SVM has the form

$$a(x) = sign\left(\sum_{i=1}^{N} \lambda_i y_i (x_i, x) - w_0 \right).$$

Training of this algorithm consists in the estimation of the parameters λ (and associated w) and w_0. SVM can be easily generalized to the case of non-linear separating surfaces by replacement of the scalar product on $K(x, y)$. In this case, the classification algorithm has the form

$$a(x) = sign\left(\sum_{i=1}^{N} \lambda_i y_i K(x_i, x) - w_0 \right).$$

In this paper we use the Gaussian kernel $K(a, b) = \exp\left(-\dfrac{\|a - b\|^2}{2\sigma^2} \right)$. The function

$$s(x) = \sum_{i=1}^{N} \lambda_i y_i K(x_i, x) - w_0,$$

is considered as a score function for the binary SVM and the function

$$g(y, x) = \max\{0, 1 - y \cdot s(x)\}/2$$

is used as the loss function. Feature space was reduced in accordance with the method and results of the selection of the most informative spectral channels represented in [7].

Discussion

The method described above was tried-out on the test area located in the territory of Savvatyevskoe forestry near Krasnogorsky sand quarry. Available ground-based data on the distribution of the species and age composition were presented by standard tabular and geospatial forest inventory data. Uniform forest areas with a closed canopy were primarily used for the construction of the training ensemble. However, such areas cannot be found for all the available species and ages. Thus, we additionally used forest plots with mixed species and age composition, to expand the training data set.

The mixed forest areas used must have the predominance of the species of interest and the presence of subdominant species having spectral characteristics which are strongly different from the dominant species spectra. These restrictions were imposed to reduce the likelihood of falling irrelevant elements into the training set. Exclusion the boundary pixels and pixels with low values of NDVI index has also allowed us to reduce errors due to linking aerial photos to forest inventory maps and to exclude other species and other objects (water, roads, houses, etc.) from training sample pixels. Forest inventory maps of Savvatyevskoe forestry and high-resolution satellite images on which one can see typical shadows and silhouettes of forest canopy were used for selecting these areas. An additional a priori information about specific sizes and spectral characteristics (in the range of 400 to 1000 nm) of crowns of the trees growing on the territory of the forestry was used for the improvement the selection by employing the training samples obtained from the pure forest plots in the Savatevskoe forestry and from the standard sets of spectra of natural objects (http://gis-lab.info/).

Images of the forest canopy have a characteristic texture which consists of the alternating sunlit and shaded parts of crowns and intercrown spacing. The spectral characteristics of these elements are usually very changeable. Sunlit and shaded portions of the crown can be separated on the basis of integral radiance values. For the obtained training samples, we calculated integral radiances and divided them into 3 gradations using corresponding percentiles. Thus, areas of the forest canopy were divided into sunlit, half-shaded and completely shaded parts.

The general form of the Bayesian classification algorithm was used as the reference method. The Bayesian classification implies that features can be considered as random variables or vectors. In the ideal case, prior probabilities and the probability density functions of features of considered classes are known and the general form of the Bayesian classifier is optimal because the solution obtained has the minimum total probability of the classification error. Otherwise, we should estimate the prior probability values and unknown parameters of distributions (the training process).

We performed some test experiments using model random data to evaluate the efficiency of ECOC SVM classifier for different kernels and coding design matrices. The data were generated using the

Table 1. Errors of ECOC SVM classifiers obtained from random number tests

		Coding matrices				
		ordinal	one-vs-all	binary complete	one-vs-one	ternary complete
Kernels	linear	0.75	0.57	0.54	0.53	0.56
	poly 2	0.48	0.46	0.41	0.38	0.38
	poly 3	0.45	0.43	0.38	0.32	0.36
	gaussian	0.34	0.3	0.3	0.3	0.3

Gaussian mixture model with the parameters obtained from real vegetation spectra. The estimates of total misclassification error are represented in the table 1. For these data, the error of optimum Bayesian classifier with exact probability density function is 0.3. As we can see from the Table 1, the Gaussian kernel leads to minimum errors for all coding designs considered. In this case, the use of any coding matrix presented in the Table 1, excepting the ordinal, allow us to reach the minimum possible value of the classification error. Thus, we used the one-vs-all coding design as the least computationally expensive method.

Results of classification of hyperspectral image of the considered test area using ECOC SVM method are shown in the Fig. 1. Spatial distribution of main types of objects can be seen in the Fig. 1*a*. Boundary of forest plots are depicted by white lines, white digits signify the numbers of forest plots in agreement with the forest inventory map. The characteristic feature of this classification consists in the presence of "unrecognized" pixels highlighted by magenta color. These pixels appear because of the constraint on posteriori probabilities and signify that the corresponding spectral portraits are unlikely or impossible on the training set. Unrecognized objects mainly correspond to the coastal zone of the sand pit, asphalt road shaded by tree crowns and slopes of the sand embankment. In general, the main object classes are recognized with high accuracy.

Results of the recognition of forest species obtained is represented in Fig. 1*b*. The classification error of the species composition taking into account mentioned above 3 gradations of the illumination of the forest canopy can be calculated by the formula

$$\varepsilon(i) = \sqrt{\frac{(p(i)_{pine} - \hat{p}(i)_{pine})^2 + (p(i)_{birch} - \hat{p}(i)_{birch})^2}{2}},$$

where i is the number of the corresponding region, p_{pine} and p_{birch} are the percentage areas of the pine and birch, respectively, according to the forest inventory, $\hat{p}_{pine}$ and $\hat{p}_{birch}$ are the similar values given by recognition results.

Estimates of errors of the species classification for different plots are represented in Fig. 2. The black curve represents a natural level of the errors due to the boundary pixels. We can see the contribution of the sunlit tops, shaded space and intermediate parts of tree's crowns for the selected regions. The errors are higher for the plots with prevailing birch stands. Also we can see that the presence of aspen in the plot 16 which is not confirmed by forest inventory data. The total error weighted on areas of corresponding plots can be calculated by the formula

$$\varepsilon_{total} = \sum_{i=1}^{20} \varepsilon(i)W(i),$$

where $W(i) = \dfrac{\text{Area of region}}{\text{Total area of regions}}$ are weighting functions of the plots.

The total weighted error amounts were calculated for the ECOC SVM classifier and the Bayesian classifier based on the Gaussian mixture model. It should be noted, that the Bayesian classifier is not optimal in this case in contrast with the test on the model random data and thus it cannot be considered as the reference method, but only just one of alternative methods used just for the comparison. Classification errors of both methods are represented in the Table 2. We can see, that the difference
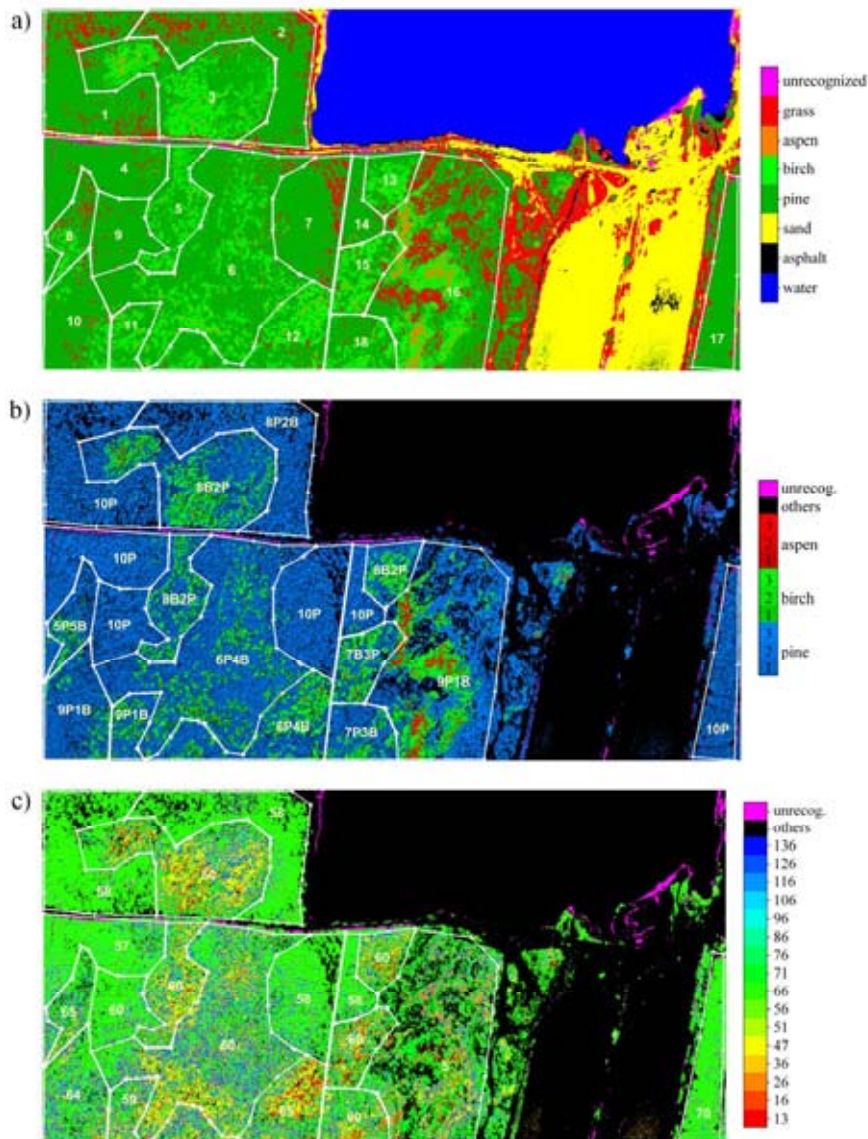

Fig. 1. Results of the classification using ECOC SVM method: a) – main types of objects; b) – forest species (taking into account gradations of the canopy illumination); c) – forest ages
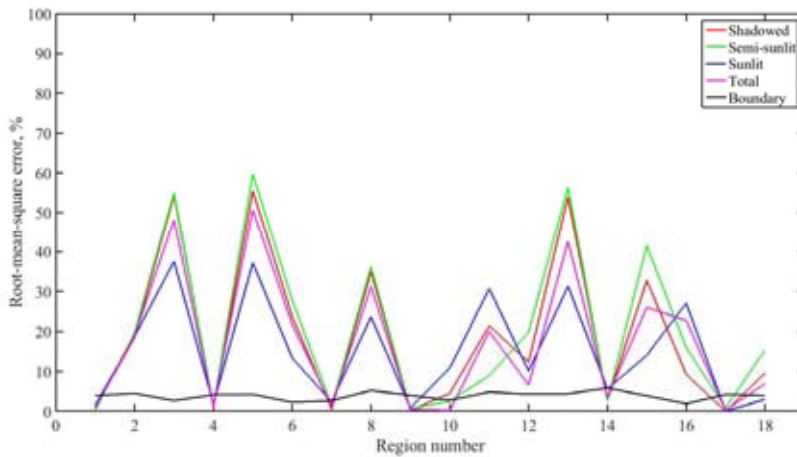
Fig. 2. Classification errors of the ECOC SVM method. The color curves represent the regional classification errors of the species composition (measured in percents) for the pixels corresponding to the shaded, half- shaded and sunlit areas and for all pixels. The black curve corresponds to the natural uncertainty because of the pixels on the boundaries between different regions

Table 2. Classification error estimates for Bayesian and ECOC SVM classifiers

Classifier		Bayesian	ECOC SVM
Weighted error	total	0.083	0.091
	shaded	0.097	0.093
	half-shaded	0.082	0.11
	sunlit	0.077	0.08
Aspen content		0.6%	0.9%

between the errors is small enough and depends on the illumination gradations. For the ECOC SVM method, the number of pixels erroneously classified as the aspen amounts to 0.9% from the total number of forest pixels. It is more than in the case of using the Bayesian classifier. In general, we can conclude that the accuracy of the classification of forest species by ECOC SVM is slightly lower than for Bayesian method. However, the both methods reveal weighted errors required for the forest inventory.

As for the classification of the age composition, the results obtained by Bayesian and ECOC SVM methods were significantly different. Ages of pine stands recognized by ECOC SVM in the plots 4, 7 and 9 (Fig. 1c) are enough homogeneous and near to the forest inventory. Ages of pine stands in the plots with the mixed species composition are significantly overestimated. Proportions of ages of the birch stands for ECOC SVM are following: 23% of pixels corresponding to birch stands are classified as 16 years old, 19% – as 51 years old and 58% – as 71 years old. Thus the portion of ripening birch stands significantly decreases at the expense of mature stands. The worst result is for the plot 16.

Finally it should be noted, that ECOC SVM method is applicable for the recognition of forest species and age composition from hyperspectral images as well as the Bayesian method based on the Gaussian mixture model. The difference between these methods is not obvious in the case of

classification of forest species when both of them have the accuracy acceptable for the forest inventory. In the case of the recognition of age classes which have poor separability, the difference between Bayesian and ECOC SVM methods becomes much more important. More reliable validation of these methods is not possible with the standard forest inventory data. Thus, more exact and detailed ground-based measurements are needed.

Conclusions

The problem of thematic processing of high resolution airborne hyperspectral images for the purpose of mapping species and age composition of forest areas is considered. The classification method proposed is based on some results of the coding theory and allows us to solve multiclass recognition problem using some selected binary classification algorithms. Kernel soft margin support vector machine classifier is used as a basic binary classifier. This method is frequently used for the processing of the remote sensing images because of its flexibility and efficiency. The technique of error correcting output codes allowed us to construct multiclass classifier ECOC SVM. Accuracy and calculation efficiency of this method depend on the choice of the coding design matrix. Numerical tests with the model data allowed us to select the coding matrix providing the high calculation speed and the accuracy which corresponds to the optimal Bayesian classifier with exactly known probability density functions of considered classes. The ECOC SVM classifier was applied for the processing of airborne hyperspectral images of the territory of Savvayesvskoye forestry (Tver, Russia). The extensive set of training spectra for different tree species and ages was constructed using the proposed method allowing the use of mixed forest plots. ECOC SVM classification errors were compared with the errors of Bayesian classifier based on the Gaussian mixture model for the same test area exemplified by forests of pine and birch species of the ages from young forest to mature forest with ten years resolution. The content of pixels classified as species missing in the forest inventory can be considered as an additional characteristic of the classification accuracy. The estimate of accuracy of the retrieval of the mixed forest composition is comparable with the accuracy of ground-based forest inventory data and the proposed technique can be recommended as an alternative to the laborious ground-based measurements.

Acknowledgments

References

[1] Marshall M., Thenkabail P. Advantage of hyperspectral EO-1 Hyperion over multispectral IKONOS, GeoEye-1, WorldView-2, Landsat ETM+, and MODIS vegetation indices in crop biomass estimation. *ISPRS Journal of Photogrammetry and Remote Sensing,* 2015, 108, 205-218.

[2] Navarro-Cerrillo R., Trujillo J., Sánchez de la Orden M., Hernández-Clemente R. Hyperspectral and multispectral satellite sensors for mapping chlorophyll content in a Mediterranean Pinus sylvestris L. plantation. *International Journal of Applied Earth Observation and Geoinformation*, 2014, 26, 88-96.

[3] Laurin G.V., Puletti N., Hawthorne W., Liesenberg V., Corona P., Papale D., Chen Q., Valentini R. Discrimination of tropical forest types, dominant species, and mapping of functional guilds by hyperspectral and simulated multispectral Sentinel-2 data. *Remote Sensing of Environment*, 2016, 176, 163-176.

[4] Dmitriev E.V. Classification of the Forest Cover of Tver' Region Using Hyperspectral Airborne Imagery. *Izvestiya, Atmospheric and Oceanic Physics,* 2014, 50(9), 929-942.

[5] Dietterich T.G., Bakiri G. Solving Multiclass Learning Problems via Error-Correcting Output Codes. *Journal of Articial Intelligence Research*, 1995, 2, 263-286.

[6] Shawe-Taylor J., Cristianini N. Kernel Methods for Pattern Analysis. *Cambridge University Press*, 2004, 462.

[7] Kozoderov V.V., Kondranin T.V., Dmitriev E.V., Sokolov A.A. Retrieval of forest stand attributes using optical airborne remote sensing data. *Optics Express*, 2014, 22(13), 15410-15423.