

УДК 528.85

The Performance of Classifiers in the Task of Thematic Processing of Hyperspectral Images

Egor V. Dmitriev^{*a,b} and Vladimir V. Kozoderov^c

^a*Institute of Numerical Mathematics RAS*

8 Gubkina Str., Moscow, 119333, Russia

^b*Moscow Institute for Physics and Technology (State University)*

9 Institutskiy per., Dolgoprudny, 141700, Russia

^c*M.V. Lomonosov Moscow State University*

1 Leninskiye Gory, Moscow, 119991, Russia

Received 19.05.2016, received in revised form 27.07.2016, accepted 26.08.2016

The performance of the spectral classification methods is analyzed for the problem of hyperspectral remote sensing of soil and vegetation. The characteristic features of metric classifiers, parametric Bayesian classifiers and multiclass support vector machines are discussed. The results of classification of hyperspectral airborne images by using the specified above methods and comparative analysis are demonstrated. The advantages of the use of nonlinear classifiers are shown. It is also shown, the similarity of the results of some modifications of support vector machines and Bayesian classification.

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

Citation: Dmitriev E.V., Kozoderov V.V. The performance of classifiers in the task of thematic processing of hyperspectral images, J. Sib. Fed. Univ. Eng. technol., 2016, 9(7), 1001-1011. DOI: 10.17516/1999-494X-2016-9-7-1001-1011.

© Siberian Federal University. All rights reserved

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

Эффективность классификаторов в задаче тематической обработки гиперспектральных изображений

Е.В. Дмитриев^{а,б}, В.В. Козодеров^в

^а*Институт вычислительной математики Российской академии наук
Россия, 119333, Москва, ул. Губкина, 8*

^б*Московский физико-технический институт (Государственный университет)
Россия, 141700, Долгопрудный, Институтский пер., 9*

^в*Московский государственный университет им. М.В. Ломоносова
Россия, 119991, Москва, Ленинские горы, 1*

Проводится анализ эффективности методов спектральной классификации в задаче гиперспектрального дистанционного зондирования почвенно-растительного покрова. Обсуждаются особенности реализации метрических классификаторов, параметрических байесовских классификаторов и многоклассового метода опорных векторов. Демонстрируются результаты классификации гиперспектральных аэроизображений указанными методами и приводятся данные сравнительного анализа. Показаны преимущества использования нелинейных классификаторов. Демонстрируется близость результатов некоторых модификаций метода опорных векторов и байесовской классификации.

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

Introduction

At the present time, remote sensing measurements are widely used in forest inventories. Traditional approaches are based on the concept of vegetation indices calculated with the use of multispectral aerospace images in visible and near infrared region (VNIR). The most part of actually existing sensors are oriented to obtaining multispectral images in 3-5 key VNIR spectral bands. Such methods allows obtaining assessments of the structure and productivity of forest stands for large enough areas [1]. However, it is known that the employment of such kind of measurements provides not enough accurate estimates because of the coarse spectral resolution of multispectral instruments [2, 3].

The development of optical remote sensing systems is steadily evolve towards increasing the spectral and spatial resolution. The hyperspectral remote sensing is the new promising technology which can be successfully used for the indirect forest inventory. Standard hyperspectral sensors have hundreds of contiguous narrow bands that makes it possible to find out finer differences among land cover objects and significantly extend the traditional set of recognized classes [4]. The experiments represented in [5] have indicated that it is possible to select different information layers formed on a particular test site using the collected ensembles of spectra depending on the age of different coniferous and deciduous species. The automation prospects are opened up of the recognition for such complex objects as the forest ecosystems with different species and age using the hyperspectral images.

Modern methods of processing of optical images with the high spatial and spectral resolution are implemented using machine-learning algorithms of the recognition of natural and artificial objects.

The basis of computer vision models consists of numerical optimization procedures, the necessity of which appears due to uncertainties of the obtained remote sensing data conditioned by intrinsic noise of measuring instruments and possible radiometric, geometric and other distortions of the obtained images. The related applications need a deeper understanding of the information content of hyperspectral data.

There are different methods that can be used for the recognition of ground objects from hyperspectral remotely sensed images. In the last years, a number of papers have been published on the comparative analysis of different classification algorithms as applied for the assessment of the composition of forest stands. In particular, the performance of spectral angle mapper, artificial neural network and support vector machine classifiers was studied for the tropical forest stands with the use of hyperspectral images obtained from EO-1 [6]. It was found that the classification results coming from artificial neural network and support vector machine methods are quite similar in showing the distribution of eight considered vegetation classes. It was also shown that SVM classifier can be effectively used in the considered particular problem without any reduction of the dimensionality of the feature space.

The comparative analysis of the support vector machine and random forest classifiers was carried out in [7]. The test area located in the North of Karlsruhe in the federal state of Baden-Württemberg in Germany contains forest stands with characteristic for Central Europe composition of species. It was found that both these classification methods can be considered as equally reliable, however the random forest method outperforms the support vector machine classifier. In this study we consider that it is more important to compare the fundamental approached which form the basis for many other recognition methods and algorithms.

Classification methods

The problem of the construction of the supervised classifier can be formulated as follows. Let us denote the set of features as X and the set of object labels as Y . In the considered case, the features are the measured spectral radiances or associated values. Let we have some prior information represented in the form of the finite set of pairs of elements of sets X and Y : $X^N = \{x(i), y(i)\}_{i=1}^N$, where N is the number of such pairs. It is necessary to construct an algorithm $s = a(x)$ (where $x \in X$, $s \in Y$) which provides the result that is the best in the set X^N in some certain sense. The process of the optimization of parameters of the classifier is called as ‘training’, and the set X^N is called as ‘training set’. In this paper we compare the different modifications of metric classifiers (MC), parametric Bayesian classifiers (BC), the method of k weighted neighbors (KWN) and the support vector machine (SVM) [8].

Applying MC we suppose that some reference characteristics of the spectral reflectance can be assigned for each of recognized objects. The training of MC consists in following stages. At first we need to define a metric ρ in the feature space. Then we calculate the positions of centroids (mean values of corresponding features) for each class on the basis of the set X^N and distances from features to relevant centroid. Using the distances obtained, we construct quantiles 0.95-0.99, perform the filtering of features getting into the critical domain for each class and recalculate positions of the centroids. This operation is necessary because the data of measurements can contain some significant outliers. Further the values of maximum possible distances ρ_{\max} to the centroids of each

class are calculated. The algorithm of the classification of the measurement $x \in X$ can be expressed as follows

$$a(x) = \begin{cases} y(i), & k \left(\min_i \rho(x, c_i) \right) \leq \rho_{\max}(i), \\ \text{NaN}, & \text{else} \end{cases}$$

where i is the index of the class, $c_i \in X$ is the vector with coordinates of corresponding centroid, k is the scaling coefficient, NaN signifies objects outside of the set Y (called as ‘unrecognized objects’).

The use of BC implies that features can be considered as random variables or vectors. As a matter of fact, the measured spectral radiance reflected from recognized complex objects can be represented as an aggregate of radiances reflected from optically homogeneous elementary components. For instance, if some area of the forest canopy corresponds to the resolution cell of the hyperspectral instrument then the measured spectrum should be dependent on the volume distribution of foliage and branches, variations of the reflectance of leaves, branches and underlying surface, and the balance of mentioned elements, even in the case of homogeneous species and age composition of the forest stand. Since in the reality we do not have any exact information about the proportion of elementary components and their mutual spatial arrangement, the formation of spectra of resolution cells of the hyperspectral instrument has essentially random nature.

The general form of the BC algorithm is represented by the expression

$$a(x) = \arg \max_{y \in Y} P_y p_y(x),$$

where P_y is prior probability of the class y and $p_y(x)$ is the probability density function of features of this class. General form of BC is optimal because the solution obtained has the minimum total probability of the classification error. The training of BC consists in the estimation of prior probability values and distributions of features for all considered classes.

The estimation of P_y is a kind of the adjustment process of the ‘rigidity’ of the selection of features attributed by classifier for corresponding classes. For limited discriminant surfaces, the uniform scaling occurs. For instance, if the considered class is discriminated by elliptical surface, increasing of the prior probability value for the given class leads to the proportional increasing of all its axes. Frequently P_y values are specified empirically on the basis of some suppositions concerning the possibility of the presence of the object in the considered scene. The more general approach is the estimation of P_y using the results of preliminary texture classification. This allows us to decrease the probability of gross errors in the course of the recognition of main types of objects. Pixels classified with expectedly low accuracy using texture analysis are considered as having identical P_y values for all classes.

Parametric approach implies that the probability density function is known up to parameters – $p_y(x) = \varphi(x, \theta)$, i.e. the parametric family of distributions Φ^θ is defined. Thus, the optimal value of the parameter vector θ can be obtained from the principle of maximum likelihood if the regularity conditions are satisfied.

In the case when the estimated probability density functions belong to the family of normal distributions (i.e. $p_y(x) \in N(\mu_y, \Sigma_y)$, μ_y is the expectation vector, Σ_y is the covariance matrix of features of the class y), the obtained algorithm is called as the quadratic normal BC or simply the normal BC

$$a(x) = \arg \max_{y \in Y} \left(\ln(P_y) - \frac{1}{2} (x - \hat{\mu}_y)^T \hat{\Sigma}_y^{-1} (x - \hat{\mu}_y) - \frac{1}{2} \ln(\det(\hat{\Sigma}_y)) \right).$$

If the covariance matrices are identical for all classes ($\Sigma_y = \Sigma, \forall y \in Y$), then we have the linear normal BC

$$a(x) = \arg \max_{y \in Y} (x^T \alpha_y + \beta_y),$$

where $\alpha_y = \hat{\Sigma}^{-1} \hat{\mu}_y, \beta_y = \ln(P_y) - \frac{1}{2} \hat{\mu}_y^T \hat{\Sigma}^{-1} \hat{\mu}_y$.

As a rule, in practice the discriminant surface cannot be considered as a hyperellipsoid, however it can be approximated by aggregate of hyperellipsoids. In this case, the Gaussian mixture model is used instead of the multivariate normal distribution

$$p_y(x | \{w_i, \mu_i, \Sigma_i\}_1^K) = \frac{1}{\sqrt{(2\pi)^{N_x}}} \sum_{i=1}^K \frac{w_i}{\sqrt{\det(\Sigma_i)}} \exp \left[-\frac{1}{2} (x - \mu_i)^T \Sigma_i^{-1} (x - \mu_i) \right],$$

where K is the number of components of the mixture, w_i, μ_i and Σ_i are respectively the weight parameter, the expectation vector and the covariance matrix of the component i of the mixture for the class y . The expectation-maximization (EM) algorithm is used for the estimation of the parameters w, μ, Σ .

Features that are far from samples in the training dataset correspond most likely to some other unknown group of objects. Thus, it is necessary to use only closed surfaces restricting some finite area in the feature space. For this purpose, the BC algorithm must be supplemented with additional constraint

$$\max(p_y) < P_{\min},$$

which allows us to introduce the special class of ‘unrecognized objects’. In this case, all discriminant surfaces of the quadratic normal BC will be hyperellipsoids.

The KWN method (K is a positive integer, typically small in comparison with the number of samples in the training set) is the most often used in this case dealing with K nearest neighbors (KNN). An object is classified by a majority vote of its neighboring pixels, with the object being assigned to the class most common among such neighbors. If we have a learning sample of pairs of remotely sensed pixels in the feature space and the names of classes, then we can use a certain measure to distinguish elements of the feature space, and a weighted function serves to separate the corresponding sub-class of the K weighted neighbors method. As a result, we can find the discriminant surface by these techniques, which are similar to the non-parametric Bayesian classifier using Parzen’s window.

The binary SVM enables to find the most distant pair of hyperplanes discriminating two considered classes and passing through the area of boundary features (support vectors) of these classes. Let us define the set of labels of these classes as $y \in Y = \{-1, +1\}$ and write the equations of discriminant hyperplanes as $(w, x_i) - w_0 = y_i$. Then the optimization problem for the ‘soft margin’ SVM can be represented as follows

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

where the parameters $\xi_i \geq 0$ are penalties for the erroneous classification of the boundary points.

The training consists in the estimation of parameters of hyperplanes $w = \sum_{i=1}^N \lambda_i y_i x_i$ and

$w_0 = (w, x_s) - y_s$, where x_s and y_s are features and labels of classes of the support vectors, and the parameters λ_i can be obtained from the solution of the optimization problem

$$\begin{cases} -L(\lambda) = -\sum_{i=1}^N \lambda_i + \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \lambda_i \lambda_j y_i y_j (x_i, x_j) \rightarrow \min_{\lambda}; \\ \sum_{i=1}^N \lambda_i y_i = 0 \quad \text{and} \quad 0 \leq \lambda_i \leq C, \quad i = 1, \dots, N. \end{cases}$$

The algorithm of the linear classification using SVM can be written as

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

wherein it is noted that only the support vectors are used for the summation, since otherwise $\lambda_i = 0$.

SVM is easily generalized to the case of nonlinear discriminant surfaces using the transformation

of the scalar product (the kernel trick). In this case $w_0 = \sum_{i=1}^N \lambda_i y_i K(x_i, x_s) - y_s$ and the classification algorithm takes on the form

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

The following kernel functions $K(x, y)$: $K(x, y) = ((x, y) + 1)^2$ – the quadratic kernel and

$K(x, y) = \exp \left(-\frac{\|x - y\|^2}{2\sigma^2} \right)$ – the Gaussian kernel are considered for the comparison in this paper.

SVM can be extended to the case of multiple classes. For this the series of binary classification problems is solved for different groups of classes. The model matrix $M = (m_{ij})$ is constructed to formalize such approach. The columns of the matrix correspond to some binary SVM classifier. Elements of the matrix can have the following values: 1 – the target object, -1 – the background, 0 – other objects which are not involved in the classification. The strategy ‘one-versus-all’ is used in this paper. In this case, the diagonal elements of the matrix M equal 1, and the others equal -1. The classification algorithm has the form

$$a(x) = \arg \min_{y_i} \frac{\sum_{j=1}^L |m_{ij}| g(y_i, S_j(x))}{\sum_{j=1}^L |m_{ij}|},$$

where L is the number of models, $s_j(x)$ is the result of binary classification using the model corresponding to the column j of the matrix M , g is the loss function.

Reduction of the feature space

It is known that the learning process may become unstable when increasing the dimensionality of the feature space at the fixed number of samples of the learning set. The instability means that small perturbances of the learning set entail significant changes in the classification results. Classification approaches described above may suffer from this problem known as the ‘curse of dimensionality’. It becomes the most important for complicated multiparametric methods. The solution of the curse of dimensionality problem consists in the effective reduction of the feature space.

There are two different standard approaches: the principal component filtering and the stepwise forward selection. The first method projects the features into the basis of eigenvectors of the covariance matrix obtained from the learning set and uses for the classification only the components corresponding to the largest eigenvalues. The disadvantage of this method is found that the first the most informative components of this decomposition provide a good approximation of initial feature vectors, but do not ensure good separability of the classes.

The standard stepwise forward selection method allows the effective reduction of the feature space by selecting the most informative features (usually it spectral bands of the hyperspectral instrument). This method has the following disadvantages. The first one consists in that the classification error is estimated in the dependent manner – test samples are used for learning of the classifier. The second problem consists in the high sensitivity of the optimal sequence of features to small variations of the learning set. Especially, it concerns the last members of the sequence. In the paper [9], we have proposed the regularization for the standard stepwise forward selection with the use of optimization and resampling methods. The regularized solution is found to be much less sensitive to variations of the learning samples than the standard method. The robustness of the proposed method increases with the number of resamples. The results of the optimization of the feature space represented in [9] are also used in this paper.

Classification performance for different methods

Performances of the described above methods were compared employing the hyperspectral airborne images of the territory of Tver forestry. The test area contains forest stands mainly consisted of the pine and birch species with small presence of spruce and aspen. The learning sets were obtained from the areas with homogeneous ‘pure’ wood species and age composition. The most part of sets contains more than 2000 samples of the spectral radiances. The detailed description of the test and learning samples is given by [10].

Some results of the comparison are given in Fig. 1 for 7 major classes given by water, roads, sandy soils, forest species and grasses, which can be easily visualized in Fig. 1a. The MC gives details of the relevant vegetation classes with the unrecognized objects relating mostly to river coast (Fig. 1b). The KNN classifier demonstrates additional details of the distribution of the object classes with less number of the unrecognized pixels (Fig. 1c). The difference of the scene classification between the SVM Gaussian (Fig. 1d) and SVM polynomial (Fig. 1e) classifiers is not remarkable. The SVM linear classifier (Fig. 1f) reveals wrong classification of the most objects including water in the river in contrast

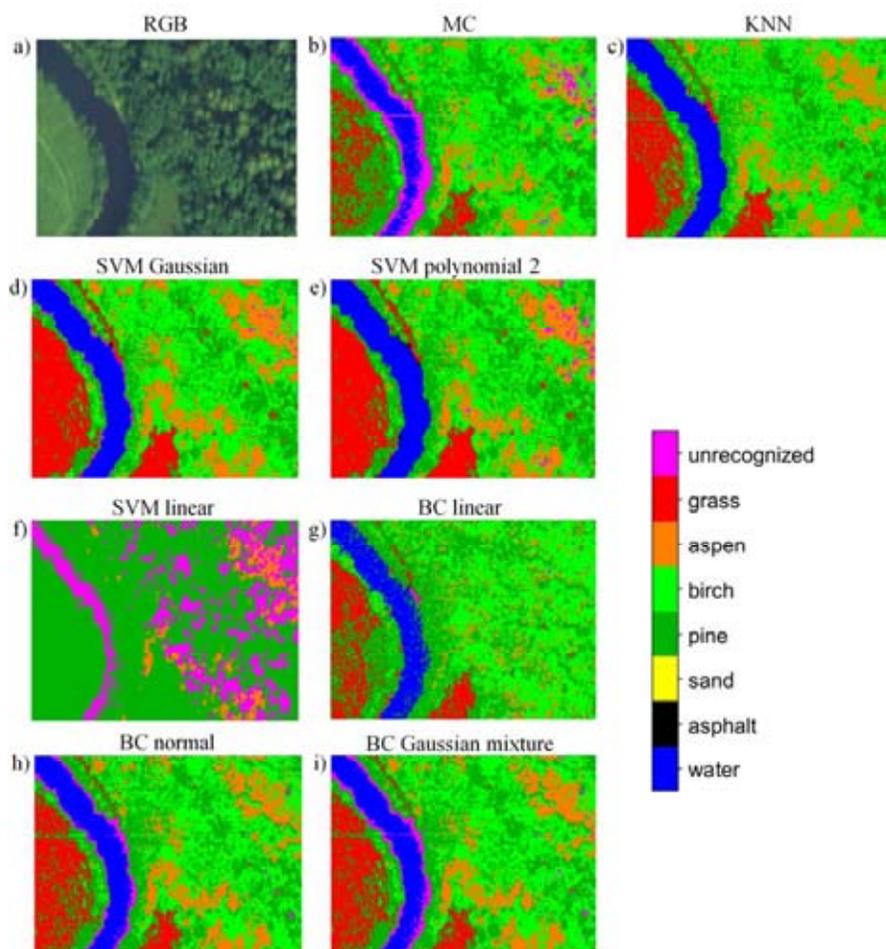


Fig. 1. Classification results of a hyperspectral image by different approaches: a – RGB-synthesized image; b – the metrical classifier (dealing with Euclidian distance); c – the KNN classifier; d – the SVM classifier with Gaussian kernel; e – the SVM classifier with polynomial kernel; f – the SVM classifier with linear kernel; g – the linear Bayesian classifier; h – the normal Bayesian classifier; i – the Bayesian classifier operating with Gaussian mixture model

with the linear Bayesian classifier (Fig. 1g) which shows much better results. The normal BC (Fig. 1h) and the BC with the Gaussian mixture of spectral radiances (Fig. 1i) highlight many additional details of the classes. The river coast is classified by them as unrecognized objects that can be explained by relatively higher rigidity due to constrains of posterior probability of classes.

The linear SVM classifier (Fig. 1f) occurs to be practically unapplicable to solve the classification problem being the worst from the methods considered. MC (Fig. 1b) also results in significant errors. Unlike the linear SVM classifier (Fig. 1f), the linear Bayesian classifier (Fig. 1g) gives satisfactory results. In general, the linear classifier gives lower accuracy as compared with non-linear classifiers (Fig. 1d,e,h,i).

The most similar are the results of the Bayesian classifier with Gaussian mixture model (Fig. 1i) and of the SVM classifier with Gaussian kernel (Fig. 1d). Both these methods demonstrate the highest classification accuracy. The SVM classifier is seen to be of the better accuracy for the meadow

Table 1. Similarity of the classification results by different methods: MC – the metrical classifier with Euclidian distance; BCG – the Bayesian classifier with Gaussian mixture of spectral radiances; BCL – the linear Bayesian classifier; BCN – the normal Bayesian classifier; SVML – the SVM classifier with linear kernel; SVMS – with square kernel; SVMG – with Gaussian kernel

	MC	BCG	BCL	BCN	SVML	SVMS	SVMG
MC	1	0.7	0.59	0.7	0.39	0.62	0.64
BCG	0.7	1	0.6	0.92	0.34	0.74	0.76
BCL	0.59	0.6	1	0.61	0.38	0.56	0.55
BCN	0.7	0.92	0.61	1	0.35	0.75	0.76
SVML	0.39	0.34	0.38	0.35	1	0.31	0.28
SVMS	0.62	0.74	0.56	0.75	0.31	1	0.85
SVMG	0.64	0.76	0.55	0.76	0.28	0.85	1

vegetation as compared with the Bayesian classifier, but of the lower accuracy while recognizing the tree's species.

The fact that the linear Bayesian classifier does not feel the river coast pixels as unrecognized objects is not good from the point of view of much errors may be present in this classification. The water body spectra along the coast differ from their learning ensembles due to the bottom and water plankton influence. This means that the pixels belonging to the coast should be classified as unrecognized. That is why the linear Bayesian classifier results in many wrong classified pixels within the river.

Table 1 gives information about the similarity of the classification results by the proposed methods. The similarity is a measure of coincidence of any two classifications compared. The value 1 means the exact coincidence, i.e. all pixels on any processed image were classified identically. This measure serves to highlight differences in classification results while employing different methods. If the results are not changed, this means that the classifier complication leads only to computer time consuming. If the relevant changes are essential, the next stage is the error comparison.

We can see from Table 1 that the maximal similarity (the level 0.7-0.9) is apparent between the metrical classifier and the normal Bayesian classifier, the normal Bayesian classifier and the Bayesian classifier with Gaussian mixture of spectral radiances, the Bayesian classifier with Gaussian mixture and the SVM classifier with square and Gaussian kernels, etc. The minimum similarity (the level 0.3-0.4) is distinctive between the SVM classifier with linear kernel and all other classifiers.

Let us consider one more result of comparison for the forest species recognition using BC with Gaussian mixture model. The recognition results shown in Fig. 2 are in accordance with 4 methods: the SVM with Gaussian kernel (Fig. 2a), the metrical classifier (Fig. 2b), the Bayesian classifier with Gaussian mixtures (Fig. 2c), the K weighted neighborhood classifier (Fig. 2d). The numbers at Fig. 2 after the listed classifiers denote the total pixels wrongly classified as the aspen while it is known that this species is not present at the scene. These numbers can be considered as a measure of accuracy of the compared classifiers which seems to be better than the direct comparison with the ground-based forest inventory data. For the KNN classifier we show two numbers depending on the nearest neighbors: 528 erroneous pixels for the case of 100 nearest neighbors and 878 erroneous pixels for the case of 1 neighbor.

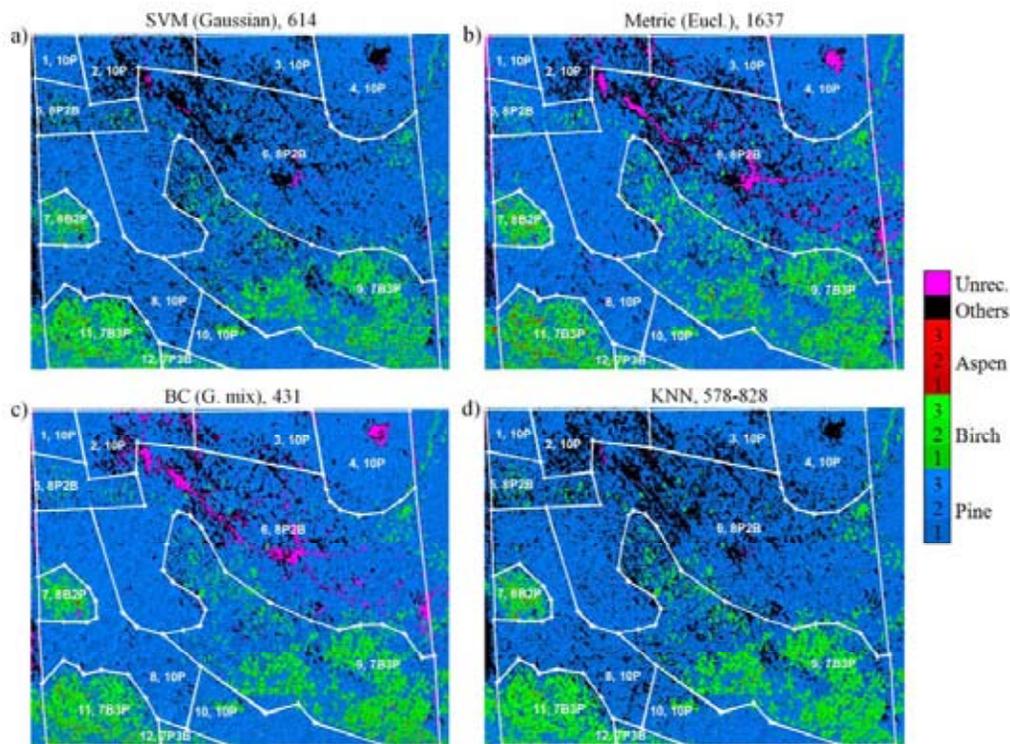


Fig. 2. The recognition results of species composition obtained by different classifiers: SVM with Gaussian kernel (a), MC (b), the BC with Gaussian mixture model (c), KNN classifier (d). Numbers in picture titles show the total amount of pixels classified as aspen. Numbers on the color scale denote gradations of the solar illumination from completely shaded pixels (1) to the sunlit tree's tops (3) with intermediate illumination condition (2). Pixels belonging to other and unrecognized objects are selected by different colors

The recognition was conducted taking different pixels into consideration: relating to the sunlit tops of trees, the completely shaded background and partially illuminated by the Sun and partially shaded tree's phytoelements. Contours of the forest inventory plots are denoted by white lines along with the white color notation of these plots (P – pine, B – birch with the resolution of 10 percent, thus 10P, for example, denotes the pure pine plot).

All 4 illustrated classifiers are seen to recognize the species composition. The ground-based forest inventory maps are known to have the error near to 10 percent, but false classified pixels for each algorithm are shown in numbers near to each figure notations. MC occurs again to be the worst. The Bayesian classifier with Gaussian mixture is the best. The SVM and the K weighted neighborhood classifiers have commensurate errors, but the latter seems to be the nearest to the optimal Bayesian classifier.

Conclusions

Basic classification methods are considered in the framework of processing of hyperspectral images. Links and similarity of different classifiers are discussed. The given examples show the nonlinear classifiers are preferable for hyperspectral imagery processing in the case if we have enough training samples. Nonlinear SVM with Gaussian kernel and parametric Bayesian classifier based on the Gaussian mixture model revealed the highest accuracy. Quality of the classification by the KNN

method significantly changes for different scenes and, for some of them, the accuracy is high enough. For the scenes corresponding to higher errors, KNN reveals significantly lower calculation efficiency. Metric classifier demonstrates inferior classification quality and can be used only for qualitative analysis. However this method is preferable in the case of lack of training data. Linear SVM and naive Bayesian classifiers demonstrates the worst accuracy and they seem to be inapplicable for the problem considered.

Acknowledgments

These results are obtained under funding support from the Russian Science Foundation (No. 16-11-00007), Federal Target Program “Research and Developments of Priority Directions in Science and Technology Complex of Russia on 2014-2020” (Grant Agreement No. 14.575.21.0028, its unique identification number RFMEFI57514X0028), the Russian Fund for Basic Research (No. 14-05-00598, 14-07-00141, 16-01-00107).

References

- [1] Knorn J., Rabe A., Volker C., Radeloff C.V., Kuemmerl T., Kozak J., Hostert P. Land cover mapping of large areas using chain classification of neighboring Landsat satellite images. *Remote Sensing of Environment*, 2009, 113(5), 957–964.
- [2] Ellis E.C., Wang H., Xiao H., Peng K., Liu X.P., Li S.C., Ouyang H., Cheng X., Yang L.Z. Measuring long-term ecological changes in densely populated landscapes using current and historical high resolution imagery. *Remote Sensing of Environment*, 2006, 100(4), 457–473.
- [3] Kozoderov V.V., Dmitriev E.V. Remote sensing of soils and vegetation: pattern recognition and forest stand structure assessment. *International Journal of Remote Sensing*, 2011, 32(20), 5699–5717.
- [4] Turner W., Spector S., Gardiner N., Fladeland M., Sterling E., Steininger M. Remote sensing for biodiversity science and conservation. *Trends in Ecology and Evolution*, 2003, 18(6), 306–314.
- [5] Kozoderov V.V., Kondranin T.V., Dmitriev E.V., Kamentsev V.P. A system for processing hyperspectral imagery: application to detecting forest species. *International Journal of Remote Sensing*, 2014, 35(15), 5926–5945.
- [6] Vyas D., Krishnayya N.S.R., Manjunath K.R., Ray S.S., Panigrahy S. International Journal of Applied Earth Observation and Geoinformation. *International Journal of Applied Earth Observation and Geoinformation*, 2011, 13(2), 228–235.
- [7] Ghosh A., Fassnacht F.E., Joshi P.K., Koch B. International Journal of Applied Earth Observation and Geoinformation. *International Journal of Applied Earth Observation and Geoinformation*, 2014, 26, 49–63.
- [8] Hastie T., Tibshirani R., Friedman J. *The Elements of Statistical Learning, second edition*. New York: Springer, 2008. 739 p.
- [9] 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.
- [10] 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.