The Performance of Texture Features in the Problem of Classification of the Soil-Vegetation Objects

Egor V. Dmitriev[1,2], Vladimir V. Kozoderov[3]; Anton A. Sokolov[4]

[1] Marchuk Institute of Numerical Mathematics of the Russian Academy of Sciences, Moscow, Russian Federation; e-mail: yegor@mail.ru
[2] Moscow Institute of Physics and Technology (National Research University); Dolgoprudny, Moscow Region, Russian Federation; e-mail: yegor@mail.ru
[3] Lomonosov Moscow State University, Moscow, Russian Federation; e-mail: vkozod@mail.ru
[4] Laboratoire de Physico-Chimie de l'Atmosphère Université du Littoral Côte d'Opale, Dunkerque, France; e-mail: anton.sokolov@univ-littoral.fr

Abstract. An analysis of the performance of texture features is carried out in the problem of supervised classification of soil and vegetation objects based on panchromatic images of WorldView-2. The 19 commonly used Haralick texture features calculated for different directions of adjacency are considered. The mutual dependencies of features and the sensitivity to the choice of adjacency direction are investigated by using correlation analysis. The most informative features which allowed us to achieve a sufficiently high accuracy of thematic processing (classification error is less than 1%) are selected.

Keywords: remote sensing, pattern recognition, texture analysis, very high resolution images, soil-vegetation cover.

1 Introduction

The development of aerospace optoelectronic systems for monitoring the Earth's surface in the visible and near infrared (VNIR) spectral range resulted in creating devices with very high spatial resolution (VHSR). A number of commercial satellite systems, such as WorldView-2, 3, 4, GeoEye-1 and Pleades have a spatial resolution of 1.24-2 m in multispectral channels and 0.31-0.5 m in a panchromatic channel. The use of this VHSR images opens up new possibilities for solving various problems dealing with remote sensing of soil and vegetation cover.

VHSR allows taking into account the distribution of illumination of elements of the forest canopy, consider a wider range of texture features and use the results of segmentation of crowns of individual trees when developing methods for thematic processing of images of forest territories. The use of VHSR satellite imagery (VHSRSI) ultimately contributes to the creation of the technology of accurate remote sensing forest inventory having high relevance to the Russian Federation and several other countries.

Questions of the efficiency of the use of VHSRSI are discussed in various scientific publications of recent years. Much attention is paid to the possibility of retrieval of forest structure parameters, such as the size and density of the crown, the height of the tree, the diameter of the trunk and the characteristic distance between the trees. For example, a technique proposed in [1] for thematic processing of VHSRSI from Quickbird and Pleades performs the search for linear dependences of forest structure parameters of pine forests using spectral and texture features of Haralick. The technique allowed achieving acceptable accuracy: the average error of retrieval of the diameters of the crowns was 1.1 m, of the distance between the trees – 0.9 m, of the height – 3 m and of the trunk diameters – 0.06 m. The Fourier texture features obtained by processing the VHSR photo were used in [2] to assess the aboveground biomass of forest stands of northeastern China. The comparison with lidar data showed that the accuracy of the proposed method was about 78%. A similar problem was also considered in [3] for the tropical forests of Cambodia. The authors used the Haralick, Fourier and Gabor texture features as applied to images provided by Google Earth.

The problem of optimizing the feature space arises in various works of this kind. The redundancy of the features used causes the problem of the curse of dimensionality in the training of classifiers and regression models. In this paper, we consider the problem of determining the effective dimension of a feature space and choosing the most informative set of features when processing panchromatic VHSRSI with the aim of classifying the soil-vegetation objects.

2 Texture Classification Technique

The texture analysis technique described here was first proposed in [4]. The technique is intended primarily for processing images in grayscale. The processing scheme is shown in Figure 1.

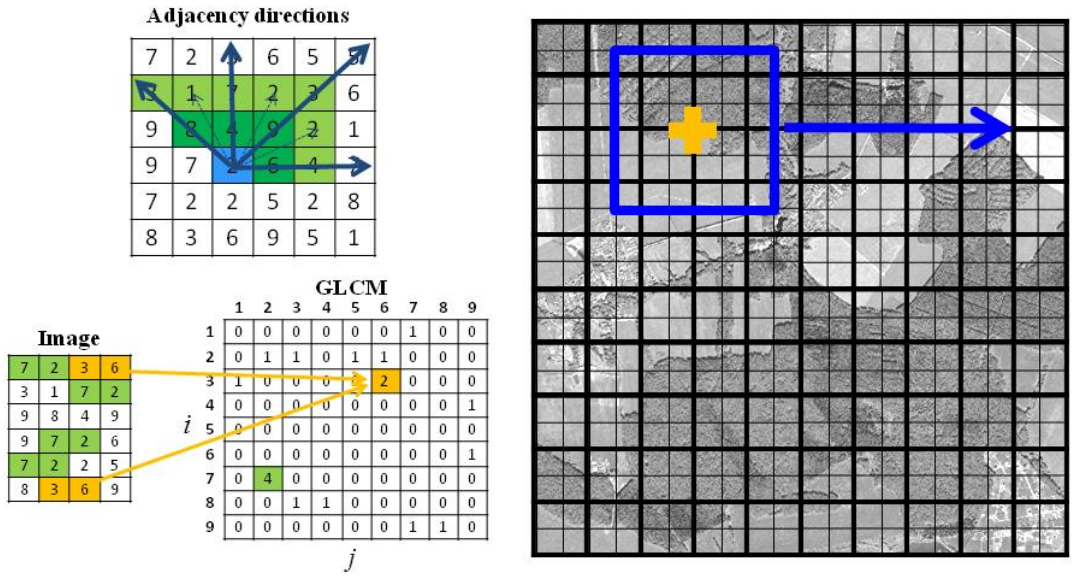

Figure 1. Scheme for calculating texture features using panchromatic VHSRSI.

At the first stage, it is necessary to evaluate the correct size of the moving window – a rectangular contour that selects the analyzed part of the image. The size of the window is determined by the characteristic scale of the analyzed textures. If the window size is chosen too small, the result of the texture classification will represent the high frequency noise and in some case it may resemble a classification based on the brightness of individual pixels. If the window size is too large, the calculation time increases and excessive smoothing of recognized objects occurs. Thus, the moving window should have the smallest possible size at which the analyzed textures are clearly distinguishable.

The panchromatic image is expanded to half the size of the moving window. The center of the window runs through all the points of the panchromatic image. When processing panchromatic and multispectral (or hyperspectral) images together, to reduce the amount of computation, it is sufficient to run only pixels whose coordinates correspond to the pixel centers of the multispectral image.

For each position of moving window, the gray-level co-occurrence matrix (GLCM) is calculated. GLCM elements are the frequencies of occurrence of brightness gradients in a given direction. An example of constructing such a matrix in the horizontal direction from left to right is shown in Figure 1. In this paper, we consider a symmetric method of constructing GLCM, when along with a given direction, the opposite is also considered. The normalized GLCM which is essentially a probability distribution function of the co-occurrence of a given number N of gray levels is calculated as

$$p(i,j) = \frac{GLCM(i,j)}{\sum_{i,j=1}^{N} GLCM(i,j)},$$

where i, j are indices GLCM elements.

Based on the values $p(i,j)$, statistics known as Haralick texture features are calculated. In this paper, the most frequently used 19 statistics are investigated. The corresponding calculation formulas are presented in Table 1. When calculating the statistics, the following parameters were used:

1) marginal expectation $\mu_i = \sum_{i=1}^{N}\sum_{j=1}^{N} i \cdot p(i,j)$, $\mu_j = \sum_{i=1}^{N}\sum_{j=1}^{N} j \cdot p(i,j)$;
2) marginal STD $\sigma_i = \sqrt{\sum_{i=1}^{N}\sum_{j=1}^{N} (i-\mu_i)^2 \cdot p(i,j)}$;
3) probability of difference $p_{i-j}(k) = \sum_{|i-j|=k} p(i,j)$;
4) probability of sum $p_{i+j}(k) = \sum_{i+j=k} p(i,j)$;

5) entropies $HX = -\sum_{i=1}^{N} p_x(i) \cdot \ln p_x(i)$, $HY = -\sum_{j=1}^{N} p_y(j) \cdot \ln p_y(j)$, $HXY = -\sum_{i=1}^{N}\sum_{j=1}^{N} p(i,j) \cdot \ln p(i,j)$, $HXY1 = -\sum_{i=1}^{N}\sum_{j=1}^{N} p(i,j) \cdot \ln\left(p_x(i) \cdot p_y(j)\right)$, $HXY2 = -\sum_{i=1}^{N}\sum_{j=1}^{N} p_x(i) \cdot p_y(j) \cdot \ln\left(p_x(i) \cdot p_y(j)\right)$, where $p_x(i) = \sum_{j=1}^{N} p(i,j)$, $p_y(j) = \sum_{i=1}^{N} p(i,j)$.

To carry out the classification based on the above-described texture features, three standard methods were considered: the normal Bayesian classifier, the k-nearest neighbor method (KNN) and the multiclass support vector machine with a Gaussian kernel [5, 6]. The indicated methods have different problem statement, accuracy and computational complexity.

Table 1. Haralick texture features.

Name of feature	Formula
Autocorrelation	$\sum_{i=1}^{N}\sum_{j=1}^{N} i \cdot j \cdot p(i,j)$
Cluster Prominence	$\sum_{i=1}^{N}\sum_{j=1}^{N} (i+j-\mu_i-\mu_j)^4 \cdot p(i,j)$
Cluster Shade	$\sum_{i=1}^{N}\sum_{j=1}^{N} (i+j-\mu_i-\mu_j)^3 \cdot p(i,j)$
Contrast	$\sum_{i=1}^{N}\sum_{j=1}^{N} (i-j)^2 \cdot p(i,j)$
Correlation	$\sum_{i=1}^{N}\sum_{j=1}^{N} (i-\mu_i) \cdot (j-\mu_j) \cdot p(i,j) / (\sigma_i \cdot \sigma_j)$
Diffrence Entropy	$-\sum_{k=0}^{N-1} p_{i-j}(k) \cdot \ln p_{i-j}(k)$
Diffrence Variance	$\sum_{k=0}^{N-1} (k-\mu_{i-j})^2 \cdot p_{i-j}(k)$
Dissimilarity	$\sum_{i=1}^{N}\sum_{j=1}^{N} \lvert i-j \rvert \cdot p(i,j)$
Energy	$\sum_{i=1}^{N}\sum_{j=1}^{N} p(i,j)^2$
Entropy	$-\sum_{i=1}^{N}\sum_{j=1}^{N} p(i,j) \cdot \ln p(i,j)$
Homogeneity	$\sum_{i=1}^{N}\sum_{j=1}^{N} p(i,j) / (1+\lvert i-j \rvert)$
Homogeneity2	$\sum_{i=1}^{N}\sum_{j=1}^{N} p(i,j) / (1+(i-j)^2)$
Information Measure of Correlation 1	$(HXY - HXY1) / \max(HX, HY)$
Information Measure of Correlation 2	$\sqrt{1-\exp(-2(HXY2-HXY))}$
Maximum Probability	$\max_{i,j} p(i,j)$
Sum Average	$\sum_{k=2}^{2N} k \cdot p_{i+j}(k)$
Sum Entropy	$-\sum_{k=2}^{2N} p_{i+j}(k) \cdot \ln p_{i+j}(k)$
Sum Squares	$\sum_{i=1}^{N}\sum_{j=1}^{N} (i-\mu_i)^2 \cdot p(i,j)$
Sum Variance	$\sum_{k=2}^{2N} (k-\mu_{i+j})^2 \cdot p_{i+j}(k)$

We have performed a series of experiments (the description is beyond the scope of this article) in which the effectiveness of these classifiers for solving the considered problem was compared. As a result, an effective modification of the KNN method was chosen. The modification consists in the optimized search by using kd-trees which increase the calculation speed. The selected number of neighbors 49 provides a balance between classification accuracy and learning sustainability.

3 Numerical experiments

For the calculations, panchromatic images of WorldView-2 of the territory of the Bronnitsky forestry (Moscow region) were used. Two test plots containing various groups of objects are considered. The Otra plot is located near the Tatarintsevsky pond and contains 4 main types of objects that differ in texture: water surface, field, natural mixed stand with a predominance of birch and spruce forest culture. The Lubninka plot is located near the settlement with corresponding name. It contains natural forests with a predominance of oak and birch, as well as part of the territory of the experimental area on which larch is grown. A distinctive feature of deciduous stands is strict ordering, trees are located along straight lines at equal distances from each other and have almost the same size of crowns. When conducting texture analysis, of particular interest is the ability to classify natural and cultural plantings.

The texture features presented in Table 1 (19 parameters) were calculated on the basis of panchromatic images of the test areas for 4 adjacency directions of 0, 45, 90, and 135 degrees. Thus, the initial attribute space has a dimension of 76. Most of the attributes turned out to be significantly dependent. Figure 2 shows the correlation matrices for 19 features in the set of directions. Correlation estimates between the features differ for the considered test plots, however, it can be seen that they have a similar structure.

The analysis of correlations by threshold values showed the following. 35% of the considered features have mutual correlations of more than 0.8 for both plots. The relationship between these variables is primarily explained by the way they are built. A relatively small part of the features has a weakly expressed mutual dependence. A correlation of less than 0.5 has 30% of the characteristics for the Otra plot and 25% for the Lubninka plot, and a correlation of less than 0.3 has 16% and 8% of the characteristics, respectively. Thus, the relationship between these signs significantly depends on the choice of scene.

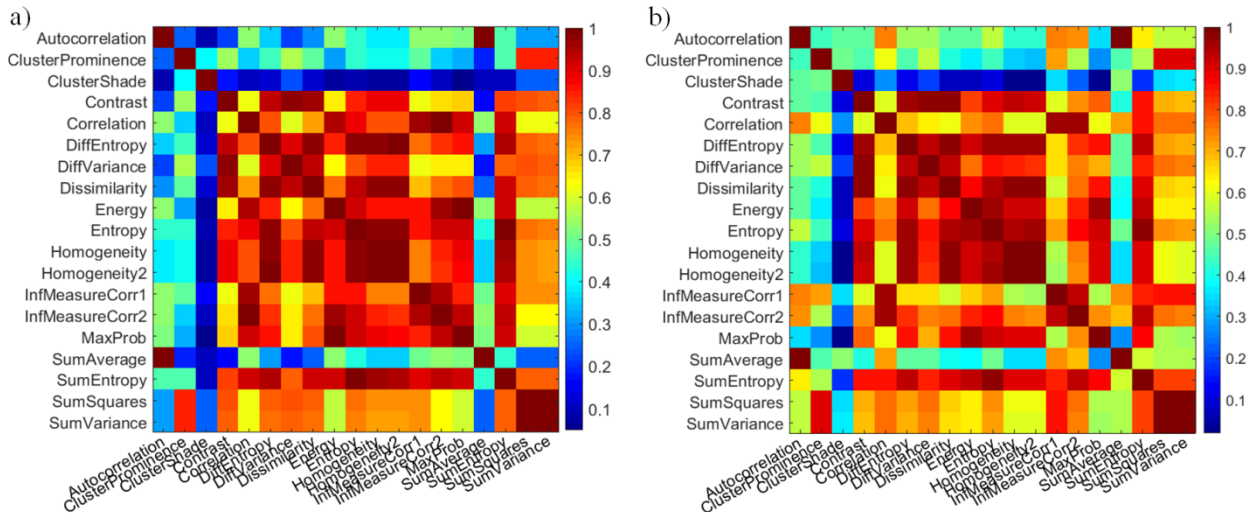

Figure 2. Matrices of correlation modules of texture features for test plots: a) – Otra, b) – Lubninka.

The results of the correlation analysis of characteristics for 4 selected directions are presented in Table 2. It can be seen that such features as Autocorrelation, Energy, Entropy, SumAverage, SumEntropy, SumSquares and SumVariance do not depend on the choice of direction. The most sensitive to the choice of direction are Contrast and DiffVariance. It should be noted that the above conclusions can be made for both test plots.

Table 2. The minimum and maximum correlation of the texture features of Haralik in the directions of the adjacency of pixels for the areas of Otra and Lubinka.

Feature name	Otra				Lubninka			
	ρ_{min}	$\alpha(\rho_{min})$	ρ_{max}	$\alpha(\rho_{max})$	ρ_{min}	$\alpha(\rho_{min})$	ρ_{max}	$\alpha(\rho_{max})$
Autocorrelation	1	90-0	1	135-45	1	90-0	1	135-45
ClusterProminence	0.99	90-0	1	135-90	1	90-0	1	135-45
ClusterShade	0.99	135-45	0.99	135-90	0.99	90-0	1	135-45
Contrast	0.81	135-45	0.95	135-90	0.83	135-45	0.94	90-45
Correlation	0.96	135-45	0.98	135-90	0.78	135-45	0.94	135-0
DiffEntropy	0.96	90-0	0.99	135-90	0.92	135-45	0.97	90-45
DiffVariance	0.72	135-45	0.94	135-90	0.79	135-45	0.92	90-45
Dissimilarity	0.93	90-0	0.97	135-90	0.9	135-45	0.96	90-45
Energy	1	90-0	1	135-45	0.99	135-45	1	135-0
Entropy	1	90-0	1	135-90	0.99	135-45	1	135-0
Homogeneity	0.96	90-0	0.99	135-90	0.92	135-45	0.96	90-45
Homogeneity2	0.96	90-0	0.99	135-90	0.92	135-45	0.96	90-45

InfMeasureCorr1	0.94	90-0	0.97	135-90	0.94	135-45	0.96	135-0
InfMeasureCorr2	0.98	90-0	0.99	135-45	0.95	135-45	0.99	135-0
MaxProb	0.99	90-0	1	90-45	0.95	90-0	0.96	90-45
SumAverage	1	90-0	1	135-45	1	90-0	1	135-45
SumEntropy	1	90-0	1	90-45	1	135-45	1	90-0
SumSquares	1	90-0	1	135-45	1	90-0	1	135-45
SumVariance	1	90-0	1	135-90	1	90-0	1	135-45

To effectively reduce the feature space, the regularized method of stepwise forward selection was used [7]. The problem with the standard method of stepwise forward selection is that the resulting sequence of the most informative features has high sensitivity to small changes in the training set. The regularized method allows getting a more stable result. Possible fluctuations in the selection results usually correspond to the least informative members of the sequence of characters. The stability of selection increases with an increase in the number of repeated calculations of locally optimal sequences of characters.

When processing data for the Otra plot, the following sequence of features was identified (the direction of adjacency is indicated in parentheses): Contrast (45), Autocorrelation (0), DiffEntropy (135), Correlation (90), Homogeneity2 (135), Dissimilarity (0) and Correlation (0). The results obtained are consistent with the data presented in Figure 3. The first 3 most informative features have the greatest probability of entering the ensemble of locally optimal sequences.

The results of thematic processing of test plots Otra and Lubninka are presented in Figure 4. You can see that the target objects were classified quite accurately. Black color on Figure 4b and 4d indicate other objects whose features are at a sufficiently large distance from the training set. The areas of other objects correspond mainly to the boundaries between the target objects, the road network, and the coastal shallow water (bottom visibility changes the texture of the water surface).

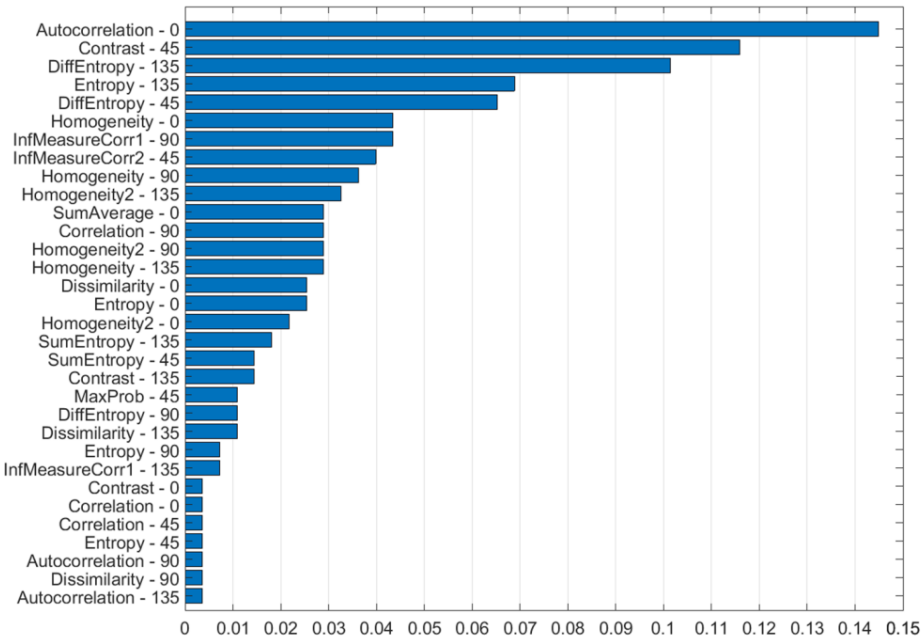

Figure 3. The probability of occurrence of characters in the ensemble of locally optimal sequences of features.

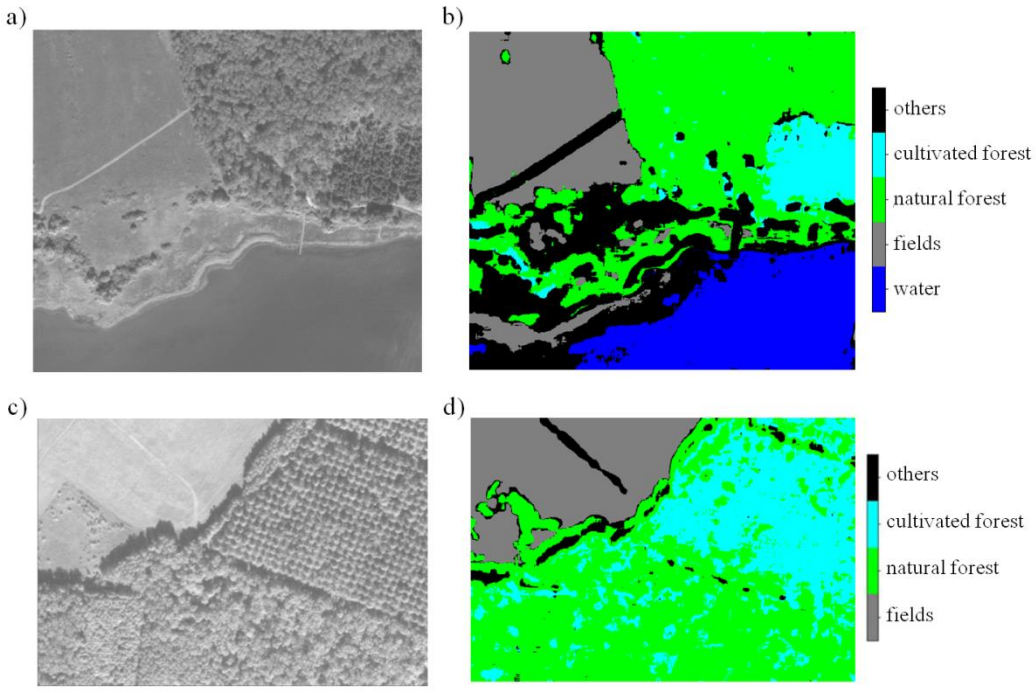

Figure 4. Recognition of target classes by texture features: a) and b) - panchromatic image and thematic map of the Otra plot; c) and d) - panchromatic image and thematic map of the Lubninka plot.

To estimate recognition errors, k-fold cross-validation [6], resubstitution (training and test ensembles coincide), and independent validation (test and training ensembles are completely different) methods were used. For these estimates of error, the designations CV, Resub, and Indep are introduced, respectively. General characteristics of the quality of the trained classification are the total probability of error TE, the average omission error TOE, the average commission error TCE, and kappa [6]. These errors are presented in Table 3. The proximity of resubstitution and cross-validation errors indicates the stability of training. Independent error estimates are significantly greater than cross-validation errors. Thus, we can conclude that there are systematic changes in the values of texture features in the image. In general, we can talk about high classification accuracy, for both test plots the error estimates do not exceed 1%, and high Cohen kappa values indicate excellent agreement between the classification results and expert data.

Table 3. General characteristics of classification quality.

	Otra			Lubinka		
	CV	Resub	Indep	CV	Resub	Indep
TE	0.001	0.001	0.005	0.006	0.005	0.072
TOE	0.002	0.001	0.009	0.005	0.004	0.070
TCE	0.002	0.004	0.002	0.007	0.067	0.006
kappa	0.998	0.999	0.994	0.990	0.992	0.890

Independent estimates of OE and CE are presented in Table 4 for each considered class. For both test plots, the smallest accuracy is achieved with forest crop recognition. For the Lubninka test plot, the recognition errors for the territory of the experimental test plot are quite high; this is most likely due to the correspondence of the average size of crowns of natural stands and cultural plantings of larch.

Table 3. Class-wise characteristics of classification quality.

		water	fields	natural forest	cultivated forest
Otra	OE	0.000	0.000	0.002	0.035
	CE	0.000	0.000	0.011	0.006

Lubinka	OE	-	0.002	0.070	0.138
	CE	-	0.000	0.091	0.109

Acknowledgements. The studies were conducted with the financial support of the state represented by the Ministry of Education and Science of the Russian Federation (unique project identifier RFMEFI58317X0061).

References

[1] Beguet B., Guyon D., Boukir S., Chehata N. Automated retrieval of forest structure variables based on multi-scale texture analysis of VHR satellite imagery // ISPRS J. Photogramm. Remote Sens. 2014. Vol. 96. P. 164–178.

[2] Meng S., Pang Y., Zhang Z., Jia W., Li Z. Mapping Aboveground Biomass using Texture Indices from Aerial Photos in a Temperate Forest of Northeastern China // Remote Sensing. 2016 Vol. 8. P. 230.

[3] Singh M., Evans D., Friess, D., Tan B., Nin C. Mapping Above-Ground Biomass in a Tropical Forest in Cambodia Using Canopy Textures Derived from Google Earth // Remote Sensing. 2015. Vol. 7. P. 5057–5076.

[4] Haralick R.M., Shanmugam K., Dinstein I. Textural Features for Image Classification // IEEE Transactions on Systems, Man, and Cybernetics, SMC-3. 1973. N 6. P. 610-621.

[5] Duda R., Hart P., Stork D. Pattern Classification, Second Edition. New York, NY, Uand Sons. 743 p.

[6] Hastie T., Tibshirani R., Friedman J. The Elements of Statistical Learning. New York: Springer, 2001. 745 p.

[7] Dmitriev E.V. Classification of the Forest Cover of Tver' Region Using Hyperspectral Airborne Imagery // Izvestiya, Atmospheric and Oceanic Physics. 2014. Vol. 50, N 9. P. 929–942.