http://dx.doi.org/10.15576/GLL/2017.2.7

GLL

Geomatics, Landmanagement and Landscape No. 2 • 2017, 7–19

CLASSIFICATION OF FORESTS IN THE PRECARPATHIAN REGION USING QUICKBIRD-2 HIGH RESOLUTION SATELLITE IMAGE

Andriy Babushka, Khrystyna Burshtynska, Yuliya Denys

Summary

Based on the study of literature on the classification of forests using high resolution space images it was established that the separation of classes and classes close to the spectral brightness can not be identified with high accuracy. Classification using maximum likelihood algorithm, which generally gives better results compared with algorithms of spectral distance or Mahalanobis distance, does not lead to the definition of areas with a high probability. Therefore, the article examines a method of forest classification by post-processing. Experimental studies were carried out using an satellite image of the forested area of the Precarpathian region, obtained from QuickBird-2 (June 2010). Data collected during field research were used as verification data to determine areas of different objects. The controlled classification has been performed using the method of the maximum likelihood, size of signatures for 8 classes were selected from 100 to 400 points. For these classes a matrix of classes separation was calculated, and a significant correlation between next classes was found: young conifer plantings and pine and mixed forest, and deciduous young plantings and deciduous forest. Post-processing significantly improves the reliability of determination of area, and the procedure consists in assigning to all pixel of the selected neighbourhood brightness of most points, although reliability of determination of area depends on the size of the area. Accuracy of determination of areas are from 92 to 99%.

Keywords

supervised classification • divergence • separation of classes • reliability • training sample

1. Introduction

Protection and using of forest resources in Ukraine is governed by the forest legislation. The Forest Code of Ukraine, adopted in 1994, is the fundamental law on forests and forest management.

Forests are changing under the influence of both natural processes and anthropogenic factors. Forests play an important ecological role in protecting forests against erosion, climatic control, creating conditions for rest and recreation of people. Therefore, damage to forests due to inefficient management can lead to environmental disaster.

The Google Inc. presented the project of "Global Forest Watch" [http://www.globalforestwatch.org], developed jointly with the World Resources Institute and 40 other organizations. Using this project in real time, you can monitor deforestation across the globe. Map on the project website is made up of the NASA's satellite images. Besides, the image processing algorithms allow to thoroughly calculate the volume of lost and new forests in each territory and each country annually.

The total area of the forest fund in Ukraine is 10.4 million ha. Environmentalists say that in recent years, forest cover in Ukraine decreased by 11% [https://ecology.unian.ua].

One of the main problems in the forest sector is the usage of outdated accounting and forest mapping technologies, lack of thematic mapping information on forests health and lack of systematic and effective monitoring.

The world practice shows that at the very core of forest monitoring methods applied to objectively obtain information about their condition, dynamic changes and effective growth assessment lies a systematic approach, the main component of which is aerospace observations. Space satellite systems have opened enormous opportunities to ensure the accuracy, speed and frequency of measurements associated with the study of forest cover and the dynamics of its changes [Lyalko et al. 2006].

One of the components of forest mapping and creating thematic maps is classification. Classification is a procedure of computer interpretation of images, which consists in the automatic separation of image pixels into classes that correspond to different objects. In Lu and Weng [2007] the authors consider the main approaches that are based on nonparametric techniques, including neural networks and decision tree classifier and parametric knowledge-based classification. The authors point to the necessity of additional research to identify and reduce uncertainties in the process of image processing to improve the accuracy of classification. Global land cover studies using satellite images from Landsat with medium spectral resolution are presented in Gong et al. [2013]. Four classifiers that were freely available were employed, including the conventional maximum likelihood classifier (MLC), J4.8 decision tree classifier, Random Forest (RF) classifier and support vector machine (SVM) classifier. The SVM produced the highest overall classification accuracy (OCA) of 64.9%. It is important to note about the different possibilities of classification using satellite images with different spectral resolution. Viewegh et al. [2003] stated that ecological factors of environment must be taken into account for classification of forest. It is proposed to use ecological net. Giles M. Foody [Foody 2008] examined the questions of accuracy of classification since the literature presented different results. Insufficient accuracy of classification even disfeatures the areas of thematic maps based on materials of remote sensing. As we mentioned, especially it refers to the calculating of areas of different objects, what is important for creating thematic maps of forests. That is why, the presented research applies to methods of post-processing of results.

In their previous studies Burshtynska et al. [2014, 2015] established that automated determination of areas of different types of objects using controlled classification algorithm depends on the density of the forest cover and impact of soil surface, as well as on the selected testing area. Deciduous forest, for example, contains, to a greater or lesser degree, other objects (bushes, grass vegetation, pine trees, etc.). A major problem of classification is to determine areas with similar objects, where as a phenomenon of "salt-pepper", known from literature, significantly distorts the results. The classified image corresponds to the real picture, however, it does not allow to automatically measure the area with the dominant type of vegetation. This is a significant obstacle to solving the problems of forestry, such as estimation of the forest surface area [Sakhatsky et al. 2002].

The Earth's remote sensing data of high-resolution are required for the bodies engaged in the organization of forest exploitation, which are characterized by two trends: first – mapping and monitoring of forest damages (deforestation, forest fires and other damages), the second – updating national databases and GIS for monitoring forests and forestry. The primary products of these services are presented in the form of thematic maps: forest clearing, forest regeneration, forest cover, forest age [Sesin 2003, Slobodjanyk 2014].

Myklush et al. [2012] say that the forest area determination on the basis of the Earth's remote sensing data is connected with both the image resolution, and the methods of its processing. It has been proposed to use maps and surface information resources as a test material.

The key role in the automatic classification of forests is played by the methods of controlled and uncontrolled classification. Foody [2008] highlights the features of the classification using satellite images of high resolution, stressing the fact that the majority of controlled classification algorithms for broad classes that include the diversity of brightness and spectral characteristics of objects leads to a significant mix-up during their classification. The authors believe the two methods to be almost identical technologically by their complexity. Methods of controlled classification provide reliable results while requiring careful formation of training samples.

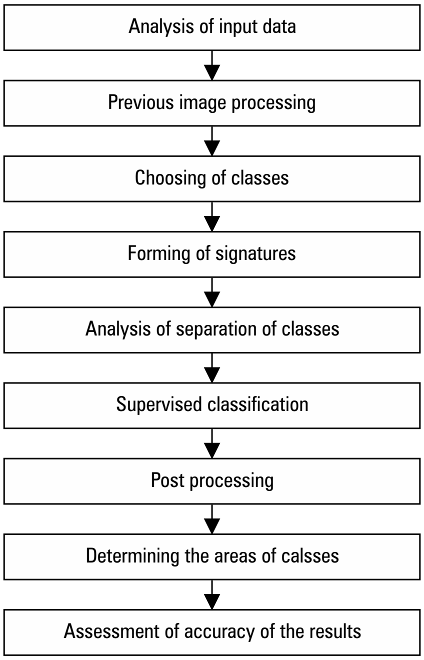
2. Presentation of basic material of the research and research results

2.1. Task of research

From analysis of the literature and our previous studies was found that phenomenon, called in the literature "salt-pepper", is significant obstacle to determine area for high resolution satellite images. In addition, classes, which are similar in spectral brightness, are slightly separating.

The task of the study was to analyze the separability of classes using method of transformed divergence and using post-processing with assessment of accuracy of classification.

For experimental investigation QuickBird-2 satellite image of the forested area of Precarpathian region was used. Technological scheme of classification and identification of areas is shown in Figure 1.



Source: authors' study

Fig. 1. Technological scheme of supervised classification and definition of area of classes

2.2. The choice of classes

Identification of objectives involves interpretation of objects to be decrypted on the image. The dominant role when selecting them and, thus, selecting the classes of objects, belongs to the image properties: number of channels of the imaging system, spatial fragmentation, radiometric fragmentation, etc.

Generally, standards are defined on the basis of field observations, which shall be conducted at the same time of the year as the imaging. Additional sources for the selection of standards may be thematic maps, images of higher resolution, deliverables and so on. In particular, in the simplest case, easily identified on the images objects are chosen as standards.

The standards determine the quality of the learning sample and, consequently, the accuracy of the controlled classification. Therefore, when determining the thematic standard zones on the image, the following requirements have been taken into account [Swain and Davis 1978]:

- the number of pixels included in the standard zones should be 1–5% of all pixels of the image, so it is recommended to use for each class 10–100 times more pixels than the number of the image spectral zones;

- the land plot area is determined in such a way that it contains accurate and reliable information about the thematic class, while its area should not be too large, as in this case the probability of unwanted variations of values increases;
- the number of standard zones depends on the number of the recognizable objects, their diversity and additional information sources used to determine the standards. Typically, 5–6 standard zones are formed for each class to take account of spatial and spectral variability of objects within each class.

It is also useful to form several zones for each class, since some of them may be excluded from the analysis in the future. It is also noteworthy that it is better to use a large number of standard zones of small size than a small number of large-area standards.

2.3. Resolution of classes

One of the first measures of statistical resolution of classes used in the image recognition is divergence [Swain and Davis 1978]. Divergence is associated with likelihood ratio L_{ij} for the two classes: i and j:

$$L_{ij}(X) = \frac{p(X \mid \omega_i)}{p(X \mid \omega_j)} \tag{1}$$

Suppose that the classes have a normal probability density function:

$$p(X|\omega_i) = N(U_i, \Sigma_i), \quad p(X|\omega_j) = N(U_j, \Sigma_j) \tag{2}$$

where:

$N(U_i, \Sigma_i)$ – multinormal density function with a mean vector U_i (i signature mean vector) and covariance matrix Σ_i (i signature covariance matrix).

Let us write the expression for the divergence, which includes the mathematical expectation and covariance matrix:

$$D_{i,j} = \frac{1}{2} tr\left[(\Sigma_i - \Sigma_j)\ (\Sigma_j^{-1} - \Sigma_i^{-1})\right] + \frac{1}{2} tr\left[(\Sigma_i^{-1} + \Sigma_j^{-1})(U_i - U_j)(U_i - U_j)^T\right] \tag{3}$$

where:

$tr[A]$ – matrix trace function,

A – diagonal sum.

The transformed divergence, which is most commonly used as a measure of resolution classes, is determined from the formula 4:

$$D_{i,j}^T = 2[1 - \exp(-D_{i,j} / 8)] \tag{4}$$

2.4. Classification results confidence estimation

Classification of the Earth's remote sensing data cannot be considered complete until we estimate its accuracy, which determines the applicability of the results in the future.

The main objective of this assessment is to determine how accurately thematic classes on the picture correspond to the classes of real objects on the Earth's surface.

The standard form of representing information when classifying estimation of accuracy is the **confusion matrix**. On the basis of the confusion matrix, a number of quantitative indicators, the most common of which is the classification accuracy and kappa index, can be calculated.

The satellite image of the forest plot in Yavoriv district of Lviv region, taken by the optical-electronic imaging system from the QuickBird-2 (June 2010) satellite, served as an incoming research material. This space-based system forms an image in the five spectral band (panchromatic, blue, red, green and near-infrared). The QuickBird-2 system discrimination capacity is 2–0.61 m in panchromatic mode and 2.44 m – in multispectral mode.

The area is dominated by coniferous and deciduous forests of different ages and species, in the north-eastern and eastern parts of the image there are agriculturally used areas and human settlements. Besides, the image contains a number of areas cut over at different times, as well as grassy and shrubby lands. All this significantly changes the luminance characteristics of objects. The area is flat land, and, thus, there is no need to take account of the terrain effect on the image.

Data obtained during the field observations with the division of the image into testing areas and the relevant description of each of them with temporal interpolation of changes serves as the verification data. The verification data was obtained in 2011 [Burshtynska et al. 2015].

The data has been collected by the following criteria: general characteristics of the testing area (for example, mature coniferous forest, uncontrolled sprouting, etc.); the average tree height; the average stem thickness; the average distance between the trees; the percentage of the dominant species.

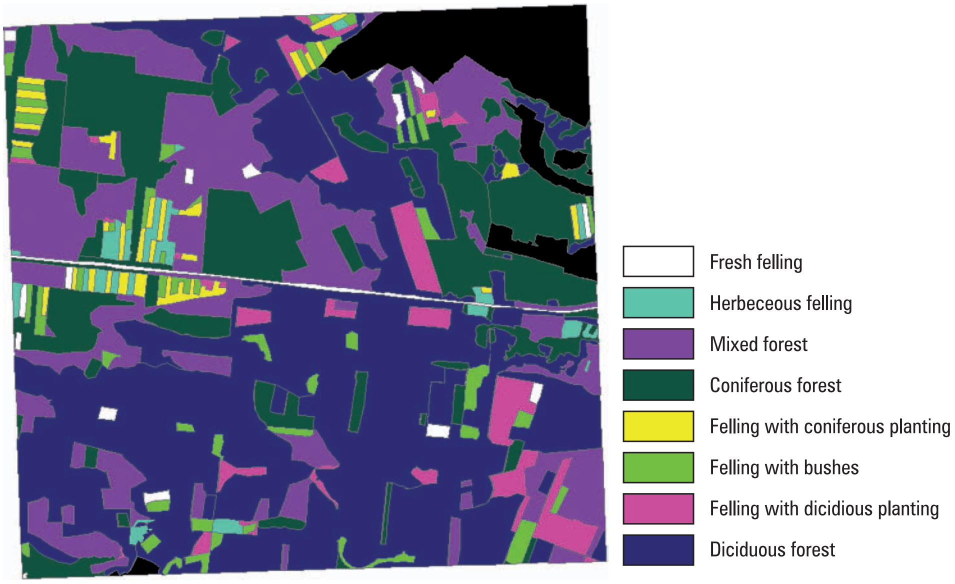
It has been established that the felled-area flora is very different. This indicates that these areas were cut over at different times. On the basis of the verification data and visual interpretation, we can distinguish the following main types of objects: fresh felled-area not overgrown with vegetation; felled-area covered with grassy vegetation; and young plantings of coniferous forests, deciduous trees and mature forests: coniferous, mixed and deciduous forests (Figure 2).

The picture is referenced in the cartographical system WGS 84. The ENVI 4.5 software environment has been used for the processing of satellite images.

2.5. Methods of the classification and determining the objects area

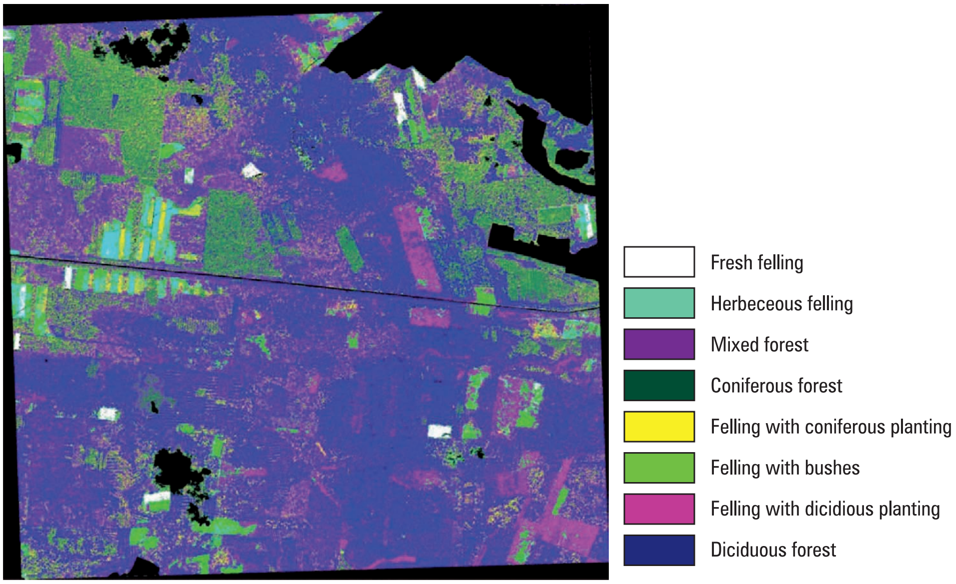
The preliminary processing of the image included channels composite and "masking" clouds and gardens (Figure 2). To interpret forests, a combination of channels corresponding to the objects' reflectivity, such as infrared, red and green channels, has been used.

As it has been already mentioned, an important stage in the controlled classification is the formation of signatures. In this regard, the verification data obtained during field observations has been applied.



Source: authors' study

Fig. 2. Image with dividing into 8 classes using results of field observations



Source: authors' study

Fig. 3. Image, classified by method of supervised classification by rule of maximum probability

The present paper uses polygonal method of forming training samples. The dimensions of the standards ranged from 100 to 400 pixels for different objects.

The controlled classification has been performed using the method of the maximum likelihood. The method has been selected on the basis of the processed literature [Lyalko et al. 2006, Myklush et al. 2012] and previous studies [Burshtynska et al. 2014, 2015]. Classification results are presented in Figure 3.

It had been established that young coniferous plantings correlate to coniferous forests, young deciduous plantings – to deciduous forests. Accordingly, the area determination error, calculated on the basis of the verification data and using controlled classification method for some classes reaches 35–50%. Only fresh felled-areas are identified with a high accuracy.

In reality, analysis of classes resolution is frequently carried out using the statistical values as the measure of resolution. Table 1 presents the transformed divergence values calculated by the formula 4. If the value of the classes resolution by the transformed divergence method is less than 1 ($D < 1$), they are inseparable, if D is defined within (1, 1.5) – slightly separable; if D (1.5; 1.8) – separable classes; if D (1.8; 2) – classes are well separable.

Consequently, the matrix elements values characterizing the classes resolution using the transformed divergence method for the “young coniferous plantings” and “coniferous forests” classes amount to 0.28, which means that these classes are inseparable together. The “mixed forest” and “young coniferous plantings” have the value of 0.82; “deciduous forest” and “young deciduous plantings” – 0.78, which indicates their low resolution.

Usually, to determine whether the thematic classes on the image correspond to the classes of real objects a confusion matrix is used (Table 2).

Classification accuracy is a proportion of the correctly classified pixels to the total number of pixels checked. For the controlled classification, the classification accuracy amounts to $P = 81.6\%$, while the coefficient amounts to $\kappa = 0.79$. The classification accuracy of the mixed forests is the lowest and amounts to 62.8%. This is due to the fact that this class contains pixels that by their spectral characteristics are similar to the coniferous and deciduous forests ones.

The next stage involves the post-classification processing, which will improve the classification results and measuring the forest cover.

3. Post-classification image processing

The classified image may contain misclassified pixels and plots. Objects areas are determined on the basis of the number of pixels with similar spectral characteristics. Isolated pixels or groups of pixels within a complex in terms of the number of objects image significantly distort the results of the area determinations. In order to improve the results, pixels reclassification shall be carried out. We have used the Majority Analysis tool. It works on the principle that the central pixel in the kernel is replaced with the class value that the majority of the pixels in the kernel has. First, a study, which involved

Table 1. Matrix of separation of classes by method of transformed divergence

Class	Fresh felling	Felling covered with grass	Felling with bushes	Young coniferous planting	Young deciduous planting	Coniferous forest	Mixed forest	Deciduous forest
Fresh felling	2.00	1.62	1.73	1.96	1.96	1.99	1.92	1.99
Felling covered with grass	1.62	2.00	1.38	1.82	1.94	1.97	1.82	1.97
Felling with bushes	1.72	1.38	2.00	1.79	1.35	1.97	1.59	1.83
Young coniferous planting	1.96	1.82	1.79	2.00	1.92	**0.28**	**0.82**	1.72
Young deciduous planting	1.96	1.94	1.35	1.92	2.00	1.99	1.71	**0.78**
Coniferous forest	1.99	1.97	1.97	**0.28**	1.99	2.00	1.13	1.66
Mixed forest	1.92	1.82	1.59	**0.82**	1.71	1.13	2.00	1.16
Deciduous forest	1.99	1.97	1.83	1.72	**0.78**	1.66	1.16	2.00

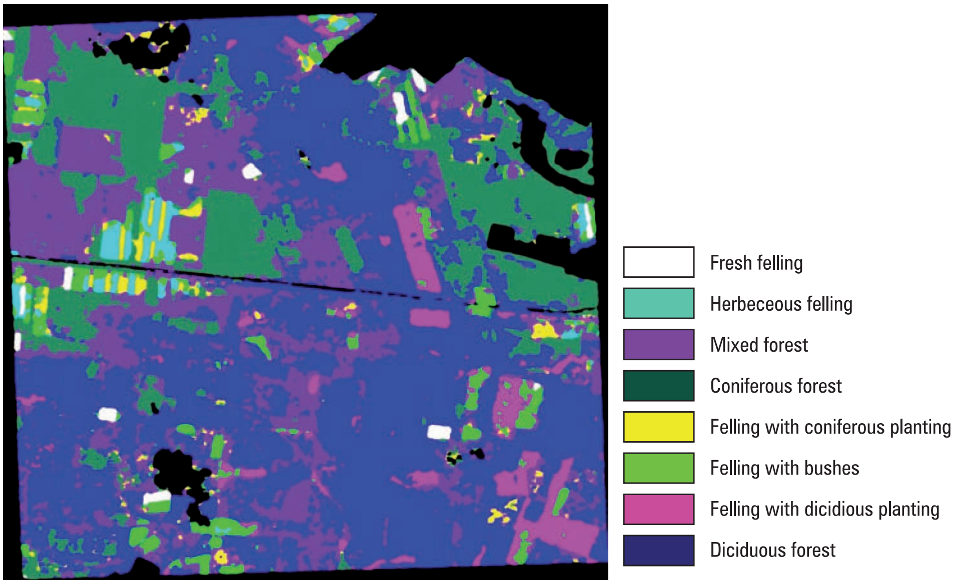
Source: authors' study

Table 2. Error's matrix of classification of forest [%]

Class	Felling with bushes	Coniferous forest	Deciduous forest	Mixed forest	Fresh felling	Felling covered with grass	Young deciduous planting	Young coniferous planting
Felling with bushes	79.5	0.0	0.2	1.5	3.0	8.6	6.8	1.3
Coniferous forest	0.0	77.6	2.2	12.5	0.0	0.0	0.0	13.0
Deciduous forest	0.8	2.1	82.2	8.7	0.0	0.0	7.4	1.2
Mixed forest	1.2	6.2	8.1	62.8	0.00	0.3	1.6	12.2
Fresh felling	0.5	0.0	0.0	0.1	92.2	0.9	0.00	0.00
Felling covered with grass	7.2	0.1	0.0	0.5	3.1	89.1	0.1	1.1
Young deciduous planting	9.4	0.0	6.7	1.3	0.0	0.1	84.0	0.7
Young coniferous planting	1.1	14.0	0.5	12.5	0.0	1.0	0.1	70.5

Source: authors' study

the selection of the kernel, within which pixels reclassification shall take place, has been undertaken. In accordance with the results of the comparison of forest areas obtained from the verification data and calculated from the pixels reclassification, the optimum reclassification kernel has been established. In general, kernels of size 15 × 15, 17 × 17, 19 × 19 were analysed. This made it possible to establish that the kernel of size 17 × 17 gives the smallest error in determining areas. Figure 4 presents the post-processing results.



Source: authors' study

Fig. 4. Results of post-processing with using the tool Majority Analysis

Differences between the forest areas received during the post-classification image processing and testing areas identified by the results of areas digitizing taking into account the verification data are presented in Table 3.

The post-classified image contains pixels that have not been classified and are coloured in black, and, accordingly, their areas have not been taken into account when calculating the forest cover. To get rid of this defect, the Clump Classes tool based on morphological operators was applied. Since, as a result of the classification, pixels that are not related to any of the classes and which are coloured in black were not taken into account when calculating the forest cover, the Clump Classes tool shall be used.

Table 3. Differences of areas of forests on the results of post-processing

Type of forest	Areas calculated by the decryption of the image [ha]	Areas calculated on the basis of reclassification [ha]	Differences of areas [ha]	Differences of areas [%]	The reliability of determination of areas [%]
Fresh felling	19.8	18.5	1.3	6.5	**93.5**
Felling with bushes	84.3	86.0	1.7	2.0	**98.0**
Felling covered with grass	35.7	32.9	2.8	7.8	**92.2**
Young deciduous planting	106.3	112.6	6.3	5.9	**94.1**
Young coniferous planting	42.6	38.5	4.1	9.6	**90.4**
Mixed forest	498.5	494.5	4.0	0.8	**99.2**
Coniferous forest	487.4	445.0	39.7	8.1	**91.9**
Deciduous forest	1012.8	1014.8	2.0	0.2	**99.8**

Source: authors' study

4. Conclusions

The supervised classification of the Precarpathian region forest was implemented using QuickBird-2 high resolution satellite image. The following conclusions can be drawn from the research results:

1. It is established that for the satellite images with high resolution significant obstacle to separability of classes and determination of their areas is a phenomenon of variability of spectral brightness, in the literature called"salt-pepper".
2. The analysis of separability of classes by the confusion matrix and statistical resolution demonstrated that some of them, for example the "coniferous plantings" and "coniferous forests", "deciduous planting" and "deciduous forest" classes are low separable classes.
3. Using post-processing with the previous definition of optimal neighbourhood significantly improves accuracy of classification.
4. The result of the definition of the area of objects in the area of investigation using method of maximum likelihood with post-classification processing ranged from 92 to 99%.

References

Burshtynska K., Madyar J., Polishchuk B. 2015. Deforestation monitoring at different periods by satellite imagery. ISPRS WG IV/2 Workshop, Novosibirsk, 114–127.

Burshtynska K., Polishchuk B., Madyar J. 2014. The definition of the area of felling forests by high resolution satellite images. Geomat. Landmanag. Landsc. (GLL), 3, 43–54.

Ekologia, https://ecology.unian.ua.

Foody G.M. 2008. Harshness in image classification accuracy assessment. Int.J. Remote Sens., 29(11), 3137–3158.

Gloabal Forest Watch, https://www.globalforestwatch.org

Lu D., Weng Q. 2007. A survey of image classification methods and techniques for improving classification performance. Int. J. Remote Sens., 28(5), 823–870.

Lyalko V.I. 2006. Bagatospektralni metody dystancijnogo zonduvannya Zemli v zadachah pryrodokorystuvannya, V.I. Lyalko, M.O. Popova (eds). Nauk. dumka, 306.

Myklush S.I., Gavryljuk S.A., Chaskovskiy O.G. 2012. Dystancijne zonduvannja Zemli v lisovomu gospodarstvi: navch. posib. Lviv, ZUKC, 324.

Peng G., Jie W., Le Y., Yongchao Z., Yuanyuan Z., Lu L., Zhenguo N., Xiaomeng H., Haohuan F., Shuang L., Congcong L., Xueyan L., Wei F., Caixia L., Yue X., Xiaoyi W., Qu Ch., Luanyun H., Wenbo Y., Han Z., Peng Z., Ziying Z., Haiying Z., Yaomin Z., Luyan J., Yawen Z., Han Ch., An Y., Jianhong G., Liang Y., Lei W., Xiaojun .L., Tingting S., Menghua Z., Yanlei Ch., Guangwen Y., Ping T., Bing X., Chandra G., Nicholas C., Zhiliang Z., Jin Ch., Jun Ch. 2013. Finer resolution observation and monitoring of global land cover: first mapping results with Landsat TM and ETM+ data. Int. J. Remote Sens., 34(7), 2607–2654.

Sakhatsky A.I., McCallun J., Khodorovsky A.Ja., Bujanova I.Ja. 2002. Classification of space image for forest state identification within the Siberia region. Pt. 1 IIASA, Laxenburg, Austria, IR-02-09, April, 45.

Sesin V.A. 2003. Geoinformacijnyj pidhid do kartografuvannja lisovogo gospodarstva. Visnyk geodezii i kartografii, 3, 27–32.

Slobodjanyk M. 2014. Vykorystannya metodiv DZZ ta GIS-technologij dlya monitoryngu lisovyh resursiv. Visnyk geodezii i kartografii, 1(88), 27–31.

Swain P.H., Davis S.M. (eds). 1978. Remotesensing – the quantitative approach. McGraw-Hill, New York.

Viewegh J., Kusbach A., Mikeska M. 2008. Czech forest ecosystem classification. J. Forest Sci., 49, 74–82.

Andriy Babushka
Lviv Polytechnic National University
Department of Photogrammetry and Geoinphormatics
79013 Lviv, S.Bandery str. 12
e-mail: andriy.babushka@gmail.com

Dr. Khrystyna Burshtynska
Lviv Polytechnic National University
Department of Photogrammetry and Geoinphormatics
79013 Lviv, S.Bandery str. 12
e-mail: bursht@polynet.lviv.ua

PhD student Yuliya Denys
Lviv Polytechnic National University
Department of Photogrammetry and Geoinphormatics
79013 Lviv, S.Bandery str. 12
e-mail: yuliya-d@mail.ua