UDC 528.46

HYBRID CLASSIFICATION OF FORESTS BASED ON HIGH RESOLUTION SATELLITE IMAGES

Burshtynska Ch. V., Polishchuk B.V., Fikovska O. A.
Lviv Polytechnic National University

Key words: Earth remote sensing, satellite image, forest monitoring, controlled classification, uncontrolled classification, standard, maximum-likelihood method, Mahalanobis distance method, minimum distance method.

Problem statement and its connection to scientific and practical tasks

Solving significant problems of national importance concerning exploration of forest is possible by using aerospace information, which is prior for forest monitoring, processing methods of forecast, detection of changes and generally ensure effective forest management. For a long time Earth remote sensing data is used in the practice of forestry for forest inventory, phenological observations, monitoring of environmental state and its disturbances.

In 2002, the Government of Ukraine adopted the State Programme "Forests of Ukraine 2002-2015". The priorities of the National Strategy are as follows: 1) significance of forest protection and the need for sustainable use of forest resources and 2) increase of productivity and improvement the species composition of forest stands, and 3) improvement of the sustainability of forest ecosystems to the negative factors of anthropogenic disturbances and global climate change.

Analysis of research papers and publications

The expediency of using satellite imagery for studying the Earth's forest resources is determined by their advantages over traditional methods, and availability of infrared channels of modern surveying systems increases their informative value. [3, 8].

Based on the study of professional literature [1, 2, 6, 10], we defined the following forestry primary tasks:

- deforestation control;
- inventory of forests;
- forest-site mapping;
- forest resources monitoring.

Forest monitoring by using Earth observation data includes:

- acquisition, accumulation and preprocessing of specific space information;
- receiving and accumulation of ground-burst assured information containing forestry parameters;
- joint processing satellite and ground data using GIS technology and software such as ERDAS Imagine;
- computer modeling of ecosystems to predict the direction and intensity of ecological processes;
- decision-making and the development of measures to create optimal conditions for forest management [4, 12].

Survey frequency during monitoring varies accordingly to the objectives of research from several years to several hours. [1]

Defining many characteristics of forest state such as species, age of forest disease areas, ecological state is implemented in the world practice through the use of multi-and hyperspectral survey systems.

A few achievements of Scientific Centre for Aerospace Research of the Earth of NAS of Ukraine in this field should be mentioned. [1, 9, 13] The authors emphasize the importance of selecting a relatively small number of spectral channels that ensure a specified criterion of informativeness. From a mathematical point of view this is an aim of statistical optimization of hyperspectral images.

Since estimation of channel information value and further processing of images are carried out to solve specific thematic problems, libraries of spectral characteristics of objects are of primary importance.

Expediency of comprehensive approaches to image processing based on the principles of controlled classification, cluster analysis and spectral mixtures decomposition is pointed out in [1].

Per pixel image classification techniques allow to accurately allocate homogeneous forest stands. In [1, 13] the authors emphasize the importance of a priori knowledge about the species structure of forest, division level of species according to spectral characteristics. The authors indicate advantages of the method of spectral mixtures decomposition versus supervised classification method.

In ERDAS Imagine algorithms of classification by using the methods of controlled and uncontrolled classification are implemented [8, 10].

The process of forests classification is iterative. Measurement vector of each pixel is compared to each signature, according to the algorithm. For parametric methods the most common solutions are following:

- minimum distance;
- Mahalonobisa distance;
- maximum likelihood.

Basic dependencies of mathematical methods are presented in [1, 8].

The development of geographic information systems and remote sensing methods elevates forest

monitoring to a new level for solving various problems of forest health evaluation and forestry activities. Evaluation of illegal felling and design of clear-up lines gain urgent character..

Researc objectives

The goal of this research:

- to analyze the effectiveness of different methods for automatic forests classification based on satellite images in winter;
- determine the optimal number of classes by using uncontrolled classification;
- work out criterion for determining the optimal number of classes;
- develop a hybrid classification method and analyze its performance.

Description of raw materials. The raw material for classification of forests near Ivano-Frankove village, Jaworowski area, Lviv region served space image obtained from optoelectronic survey system of satellite Ikonos (March, 2002).

The satellite Ikonos system provides images in five spectral ranges (panchromatic, blue, red, green, near-infrared). System resolution is 1 m in panchromatic mode and 4 m in the spectral ranges [3].

The area is dominated by coniferous and deciduous forests, in the northern part village Lepehivka is located. Near the village slice of territory is occupied by farming lands. Territory is flat, that allows to ignore terrain effect on the image.

Assured information. Data obtained through field research, with the division of the image into providing grounds and the corresponding description with temporal interpolation changes served as assured information.

Data collected on the following parameters:

- general description of the providing ground (eg, mature coniferous forest, uncontrolled overgrow etc.);
- percentage of dominant species;
- average height of trees;
- average thickness of stems;
- average distance between trees.

Main material presentation. The process of distribution of pixels into classes is carried out in spectral dimention. If a pixel satisfies certain conditions, it is credited to the class that match criterion. The difficulties of classification are mainly associated with the variability of features - reflectivity varies depending on time of day, season, etc., as the result rules of classification for different images may not be the same. Classification is often uncertain and inaccurate because according to the values of the spectral brightness pixels may belong to several classes at once, and the pixel raster can be an integral value of objects of different classes - so-called mixed pixels. However, the uncertainty is ignored in the process of classification, and each pixel is considered to belong to one of the classes [5, 8].

According to the area of object recognition, space of features for classification should be divided into closed area, each of which contains a value attributes specific to one of the classes of objects, and classify each pixel of the image to the class with its feature vector . The method of classifying the image pixels to object classes defines the so-called decision rule - the rule of classification, implementation of which is provided by appropriate computer algorithm. Classification rule is based on the characteristics of typical objects which are known to belong to a certain class (for example, features of standard objects on test areas). Algorithms of computer analysis that implement different rules for classification are divided into two types: algorithms of controlled and uncontrolled classification.

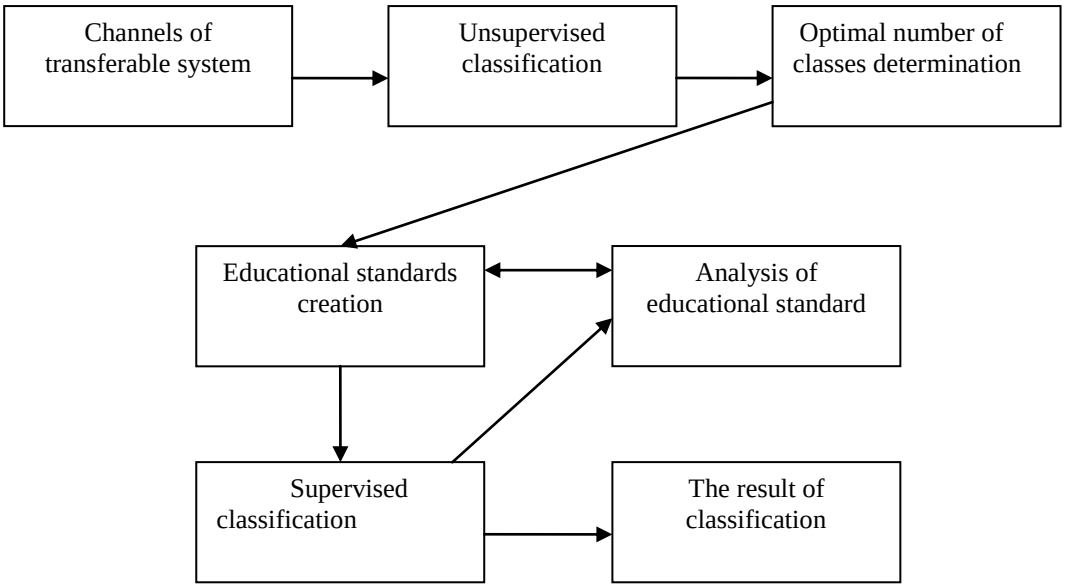
The fact that minimum distance, Mahalonobis distance and maximum likelihood leads to different results is highlighted in [1, 8]. Maximum likelihood method allows most accurately identify species of trees, Mahalonobisa and minimum distances have nearly identical accuracy, allow to identify 1-3 species in area.

In some works, in particular [8, 10] is mentioned that so-called hybrid classification of forests may be prior, which combines methods of uncontrolled and controlled classification.

Having applied uncontrolled classification, which is implemented in ERDAS Imagine following algorithm Isodata, we can determine on first stage the optimal number of classes based on the spectral characteristics of objects, and on the second stage carry out controlled classification by specific method: minimum distance, Mahalanobis distance or maximum likelihood.

Structure chart of proven methodology is shown in Fig. 1.

Figure 1.



Structure chart of hybrid classification

For the good performance of uncontrolled classification the right channel combination is to be selected at first, whereas uncontrolled classification is carried out automatically and depends on the image data/ space survey findings only.

The first step is the adjustment of channels combination. According to the reflectivity of objects such channels as infra-red, red and green were selected to interpret forest.

Uncontrolled classification. Uncontrolled classification enables to distinguish numbers of classes (groups of pixels) and can be useful for thematically defined set of classes creation using the following controlled classification.

Uncontrolled classification algorithms implement frequently used in various multidimensional studies clustering methods, which are based on the so-called limiting procedures. Their usage provides the sufficient amount of initial information for classes separation and consequently samples are not used. During the image processing algorithms use various measures of classes resolution in the space of mathematical objects and constitute the basis of their formal clustering.

The ISODATA classification algorithm is based on the use of the minimum spectral distance formula: dependence (4). This method of classification is employed in case of the similarity of the spectral features of different classes and ranges of classes values overlapping.

Thus, the pixel belongs to the class with minimum distance to the aggregate average values. The process resultes in several stages - iterations. During each iteration, all pixels are reclassified and new averages for clusters, to which distances are defined, selected. As a result all the pixels of the image are distributed between classes.

The criterion for choosing the optimal number of classes may be the expression

$$\frac{\Delta M_{(i+1)-i}}{\Delta M_{1-n}} \le \varepsilon, \quad (1)$$

where $\Delta M_{(i+1)-i}$ – the average differences of average spectral brightness between classes of channels;

n - number of classes ;

ΔM_{1-n} – average difference of spectral brightness. The choice of criterion ε set value depends on the classification task : integrated arrays of tree species (eg , pine , foliar , mixed), more structured (birch, oak , etc.).

In general, the choice of ε depends on the difference in spectral brightness ΔM_{1-n} and the average difference between the classes $\Delta M_{(i+1)-i}$.

That is:

$$\Delta M_{(i+1)-i} = \frac{\sum_{l=1}^{L} \Delta M_{(i+1)-i}}{L}, \quad (2)$$

$$\Delta M_{1-n} = \frac{\sum_{l=1}^{L} \Delta M_{1-n}}{L}, \quad (3)$$

where L – the number of channels.

To investigate the optimal number of classes in the region and to obtain statistical characteristics such as the expectation of the spectral brightness classes and histograms of images, the division into 3, 5, 8, 10, 12, 16 classes was made.

Figure 2 shows the average values of the spectral brightness of classes in three channels.

Formula (2) calculates the average difference in spectral brightness, which are made in Table 1. In the last row of the table the average value of the total

spectral brightness difference calculated by the formula (3) is entered.

Table 2 presents the results of calculations by formula (1). If the set value is , the optimal number of classes will be 8 classes. Similar in results is the choice of 10 classes, only one value exceeds the tolerate.

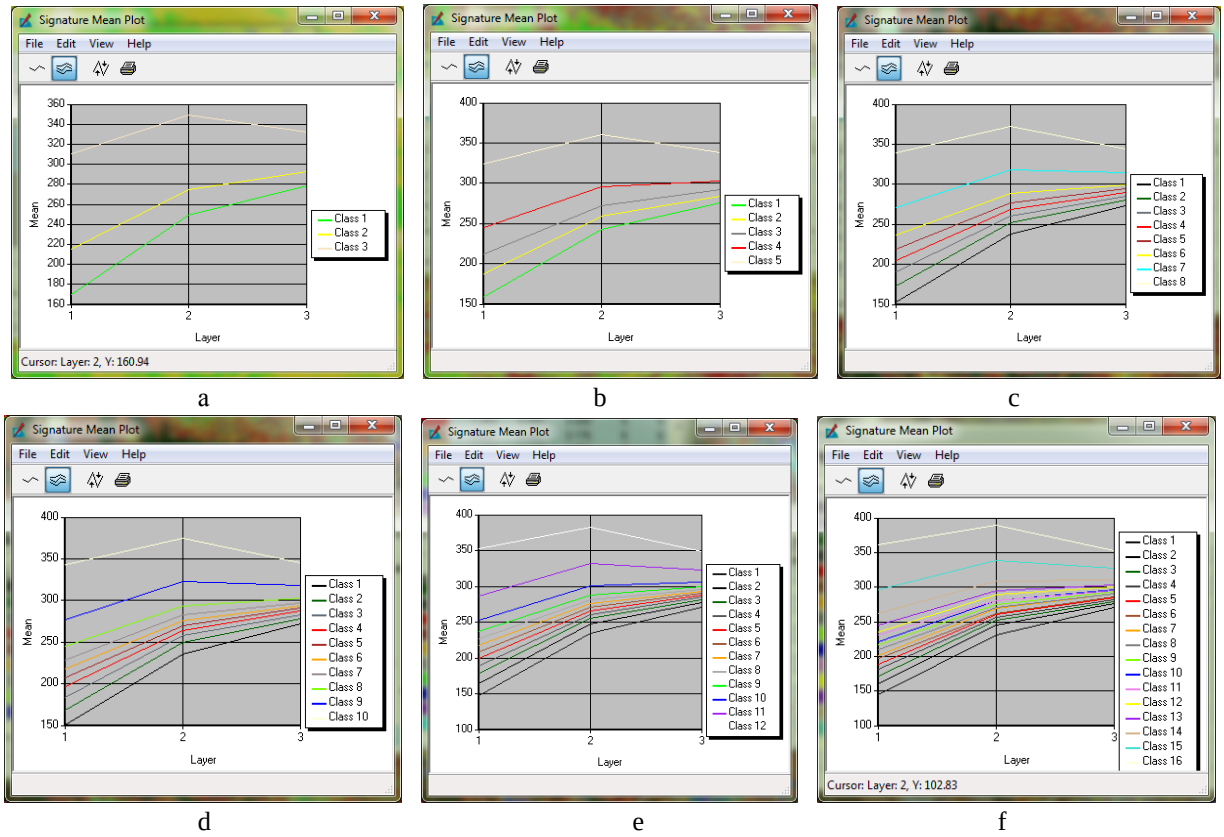


Fig . 2. Graph plotting for average values of the spectral brightness classes: a) 3 classes , b) 5 classes , c) 8 classes, d) 10 classes , e) class 12, f) 16 classes.

Table 1

Average differences in spectral brightness

Class №	3	5	8	10	12	16
1-2	27	18	13	15	10	11
2-3	71	17	11	9	9	7
3-4		23	9	9	7	5
4-5		59	10	6	8	6
5-6			10	7	6	2
6-7			28	7	7	6
7-8			50	11	7	6
8-9				26	7	5
9-10				49	9	5
10-11					27	1
11-12					47	6
12-13						4
13-14						11
14-15						33
15-16						44
1-n	97	117	131	140	144	151

Table 2

Selection of the optimal number of classes

Class №	3 кл	5 кл	8кл	10 кл	12 кл	16 кл
1-2	0,28	0,15	0,10	0,11	0,07	0,07
2-3	0,73	0,15	0,08	0,06	0,06	0,05

3-4		0,20	0,07	0,06	0,05	0,03
4-5		0,50	0,08	0,04	0,06	0,04
5-6			0,08	0,05	0,04	0,01
6-7			0,21	0,05	0,05	0,04
7-8			0,38	0,06	0,05	0,04
8-9				0,19	0,05	0,03
9-10				0,34	0,06	0,03
10-11					0,19	0,00
11-12					0,33	0,04
12-13						0,03
13-14						0,07
14-15						0,22
15-16						0,29

Classified image with division into 3 , 5, 8 , 10, 12, 16 classes shown in Fig. 3.

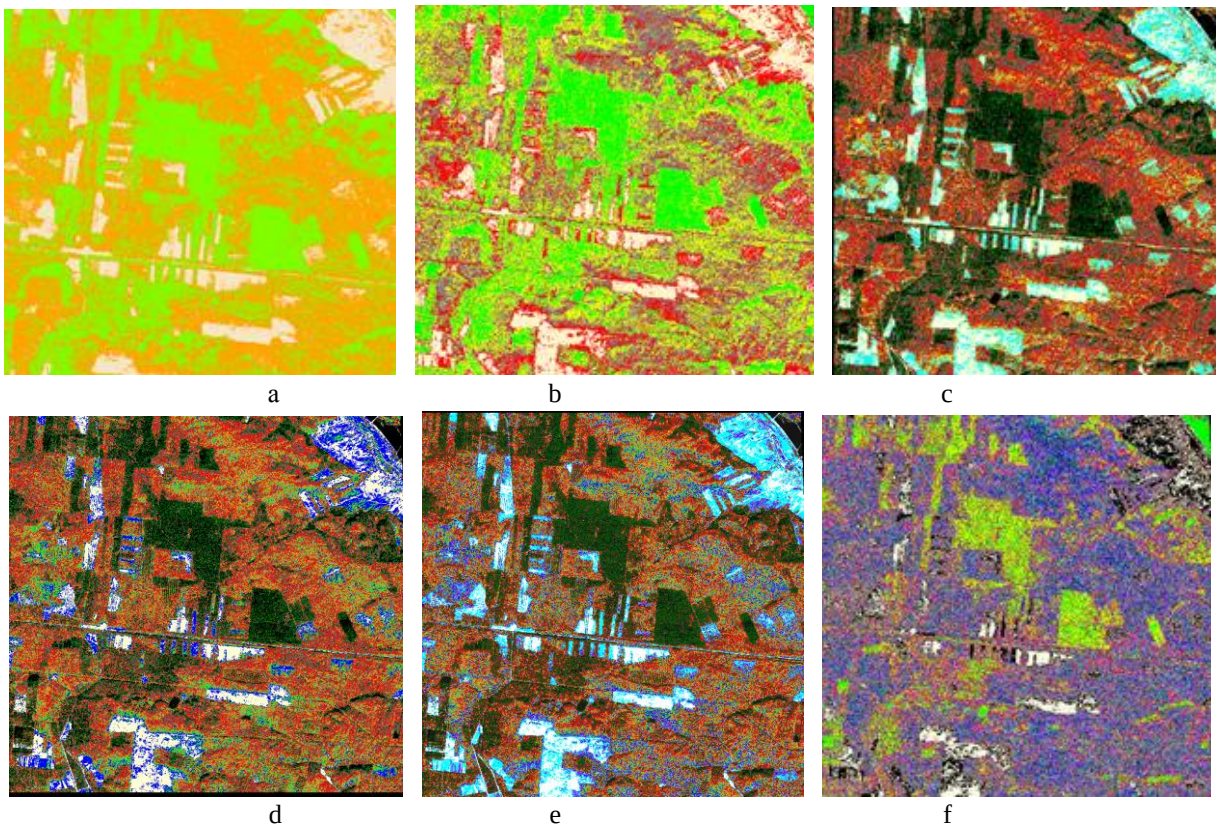


Fig . 3. Classified space Ikonos satellite image divided into a) 3 classes , b) 5 classes , c) 8 classes, d) 10 classes , e) class 12, f) 16 classes

Further elaboration of the supervised classification method is carried out with image and with its division into 8 classes (Fig. 3 B).

Key dependencies of supervised classification. The following main ways (rules) are distinguished in the supervised classification: minimum distance, Mahalanobis distance and maximum likelihood [1, 8].

Classification by minimum distance rule is the calculation of the spectral Euclid distance :

$$SD_{xyl} = \sqrt{\sum_{l=1}^{L} (\mu_{li} - X_{xyl})^2}, \quad (4)$$

where L – the number of channels; l – channel number; μ_{li} – average value of the i class in the l channel; X_{xyl} – number of the class значення l; SD_{xyl} – spectral distance of the pixel at coordinates x, y.

The equation used to classify the Mahalanobis distance method is:

$$D_M = (X - M_i)^T K_i^{-1} (X - M_i), \quad (5)$$

where D_M – Mahalanobis distance; i – defined class; X – measurement vector of classified pixel; M_i – average value of class i brightness; K_i^{-1} – covariance pixel matrix of class i brightness; K_i^{-1} – inverse matrix for K_i.

Classification by the rule of maximum likelihood realized by the following algorithm:

$$D = \ln a_i - \frac{1}{2}\ln|K_i| - \frac{1}{2}(x - m_i)^T K_i^{-1}(x - m_i), \quad (6)$$

where D – weight distance (probability); a_i – percentage probability of the classified pixel belonging to the class i (equal to 1.0 or is introduced on the basis of a priori data) - $|K_i|$ – determinant of K_i.

In professional literature, there are no clear reccomendtions to size of model samples. In [8] pointed out, that selection can be started from the 1 pixel, increasing its size. The following sample sizes 2 × 2, 5 × 5 and close to 100 pixels are selected for the research.

If to compare area each of the methods with test information, there can be found which one the most accurately determine the area of each class. In this experiment, not only the methods are compared together, but also the optimal sample size for effective classification is defined.

Areas differences presented as the research results calculated by the method of maximum likelihood, Mahalanobis and minimum distances by 2 × 2 pixels (Table 3), 5 × 5 pixels (Table 4) and 100 pixels (Table 5). The area of each class is calculated as a percentage of the total plot area of research. Area differences between relevant method and test information are given in percentage.

Area differences, defined by the size (2 × 2 pixels) pointed at sensible differences with test information, identifying young deciduous forest by all means, and a significant error is present at soil identification. Felling is identified most accurately.

Earth surface object classification by the standards of larger size (5x5 pixels) by the maximum likelihood method gives better results for some classes (mature coniferous forest, felling). Mixed forests are identified worse; discrepancy by the Mahalanobis' method was 22%. In classes (soils, mixed forest, young coniferous planting) the differences percentage of test information is also big. But in general, classification by higher samples are better in accuracy.

The following experiment, which involves the increasing of training set to 100 pixels, is conducted. Results are included in Table 5.

Table 3

Controlled classification of area differences, defined by samples 2×2 pixels

Classes	Mahalanobis' method	Minimum distance	Maximum likelihood
Coniferous mature	9,8	0,2	8,9
Coniferous planting	6,9	4,8	5,5
Young coniferous planting	0,3	5,6	0,2
Mixed forest	15,0	1,9	0,4
Deciduous mature	10,5	2,1	3,5
Deciduous young	44,9	13,2	33,3
Felling	3,1	0,2	2,8
Soils	13,4	14,4	13,4

Table 4

Area differences, defined by samples 5×5 pixels

Classes	Mahalanobis' method	Minimum distance	Maximum likelihood
Coniferous mature	0,2	1,1	0,1
Coniferous planting	3,6	8,2	0,6
Young coniferous planting	14,9	0,8	8,8
Mixed forest	22,0	1,2	9,3
Deciduous mature	9,1	2,4	1,7
Deciduous young	7,2	3,4	5,2
Felling	2,6	5,1	0,6
Soils	14,6	12,5	14,4

Table 5

Area differences, defined by samples close to 100 pixels

Classes	Mahalanobis' method	Minimum distance	Maximum likelihood
Coniferous mature	2,3	5,1	2,4
Coniferous planting	0,4	5,5	3,7
Young coniferous planting	9,5	11,1	3,2

Mixed forest	13,2	1,1	16,3
Deciduous mature	0,7	6,5	9,5
Deciduous young	7,1	16,3	14,0
Felling	3,1	1,8	2,9
Soils	11,4	11,3	11,1

In general, a great improvement showed classification result by Mahalanobis method. Sensible differences are between test information and evaluation according to the minimum distance method revealed for almost all classes, except for mixed forest and felling.

Accuracy sated results obtained by the method of maximum likelihood under the following samples for classes: coniferous mature forests, coniferous planting and young coniferous planting (accuracy of identification to 4%) and felling (2.9 %). Generally, for such a type of division into classes best results were obtained by using samples 5x5 pixels.

Analysis of training samples histograms with the submitted classification indicates of some overlapping class spectral brightness: coniferous planting and young coniferous planting, geological substate has the great impact on the identification of the last and two types of deciduous forests classified not clearly. Note that the gardens are often identified with felling.

The result of this analysis is to clarify the following classes: coniferous mature, coniferous planting, mixed forest, deciduous forest, felling, soils, gardens.

Table 6 presented the results of classification as the area differences between the test information and defined samples close to 100 points for the three methods of supervised classification.

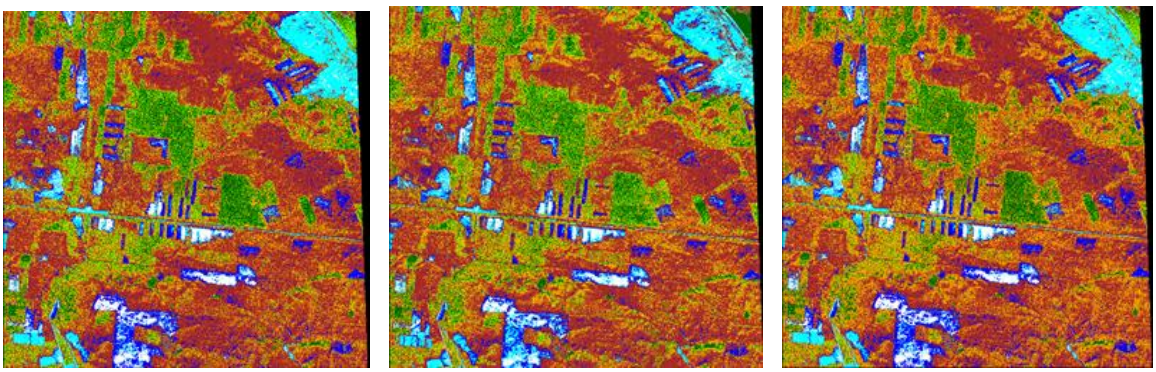
The results point at high accuracy of coniferous forests coverage and felling by the maximum likelihood method (0.5-1.2%), greatly improved the identification of deciduous and mixed forests by all methods (1,9-7%).

Three methods classified figure using samples close to 100 pixels with unspecified number of classes is presented in Fig. 4

Table 6

Area differences according to updated number of classes, obtained by samples close to 100 pixels

Classes	Mahalanobis' method	Minimum distance	Maximum likelihood
Coniferous mature	2,0	1,7	1,2
Coniferous planting	3,3	1,3	0,6
Mixed forest	4,4	1,9	3,4
Deciduous	2,7	7,0	6,5
Felling	1,2	1,0	0,5
Soils	2,1	1,5	2,1
Gardens	5,2	5,4	5,1



a) b) c)

Fig.4. Three methods classified figure using samples close to 100 pixels:
a) maximum likelihood b) minimum distance c)Mahalanobis

Row	Histogram	Color	Red	Green	Blue	Opacity	Class_Names	Area
0	27562		0	0	0	0	Unclassified	44.0992
1	174016		0	0.39	0	1	хвоя1	278.426
2	140457		0.5	1	0	1	посадки	224.731
3	340266		1	0.65	0	1	мышаний	544.426
4	773484		0.65	0.16	0.16	1	листяний_зрыли	1237.57
5	127321		0	0	1	1	вирубки	203.714
6	43806		1	1	0.98	1	грунти	70.0896
7	94311		0	1	1	1	городи	150.898

Fig.5. Attribute data of images

In figure 5 represented major attributes which are used for notation of particular objects.

Conclusions

1. The hybrid classification method is processed and criteria for determining the optimal number of classes is proposed.
2. Ikonos space survey findings show that 7-8 classes are the optimal number for the woodland classification in winter.
3. Controlled classification is realized for comparison of accuracy by methods of maximum likelihood, Mahalanobis and minimum distance.
4. Classification results point out that winter Ikonos space survey findings clearly identified felling , mature coniferous planting in rows and mature coniferous forest, determined by different methods and different samples.
5. Area differences between maximum likelihood method, Mahalanobis and minimum distance, using samples close to 100pixel with refinement of classes in comparison to test information ranging 1-7 %.
6. It is found that the choice of samples is of great importance. The small samples (2x2 pixels) overlap related classes, so that causes poor accuracy in comparison to test information. The classification is made according to samples 5x5 pixels, allows to identify felling areas and softwoods with the accuracy close to 1%, more badly identified mixed forest and soils ,at the margin of 9-14 % and Mahalanobis is up to 22 %.

References

1. V.I. Lyalko and M.O. Popov. Multispectral remote sensing techniques for land-use.– V.I. Lyalko and M.O. Popov eds., Kiev, “Naukova Dumka”, 2006. – 360 c.
2. S.A.Bartalev. Actual tasks, abilities and future direction of development methods satellites monitoring of boreal methods / S.A Bartalev, A.S. Isaev, D.V Ershov / / IV Internat. conf. "Space methods and technologies in geoinformational technologies agroforestry and forest management. - M., 2007. - P. 20-22.
3. Kh.V.Burshtynska. Aerospace shooting system: study guide. Kh.V. Burshtynska, S.A.Stankevich - Lviv: Lviv Polytechnic National University Publishing House, 2010. - 292 p.
4. Kh.V.Burshtynska. Research methods of forests classification by using satellite images of high resolution / Kh.V. Burshtynska B.V. Polishchuk, O.U.Kovalchuk / / Geodesy, Cartography and Aerial Photography: Coll. sciences. etc. - Lviv, 2013, vol. 78. - P. 101-110.
5. Remote sensing - Ed. F. Swain and S. Davis. Transl from English. - Moscow: Nedra, 1983. - 396 p.
6. G.M. Zholobak. Domestic experience in satellite monitor Ukrainian forests / Zholobak G.M. / / Space Science and Technology, 2010. T. 16. Number 3. P. 46-54.
7. A.S. Isaev. The use of satellite data for monitoring forest biodiversity / Isaev, SV Knyazev, M. Yu . Puzachenko. Study the Earth from space. - 2009. - № 2. - P.1-12.
8. S.S.Cohan. Remote sensing: theoretical foundations: a textbook / S.S. Cohan, A.B. Vostokov. -: Higher School, 2009. - 511 p.
9. V.I. Lyalko. Integration of multispectral satellite images of different spatial resolution to increase the efficiency of forest research (for example, the Chernobyl exclusion zone, and parts of Siberia) / Lyalko V.I., Sakhatskaya A.I., Jodorowsky A.Y. / / Space technology and science. - 2002. - 8, № 2/3. - P. 239-246.
10. S.I. Myklush. Remote sensing in forestry: teach. guidances. / S.I. Myklush, S. A. Hawryluk, O.G. Chaskovskyy. - Lviv: ZUKTS, 2012. - 324 p.
11. B.V.Polishchuk. Recent advances and challenges in the research and development of forests / Polishchuk B. / / Geodesy, Cartography and Aerial Photography, 2008. - Vol. 70. - P. 38-45.
12. R.F. Treyfeld. Value of remote sensing methods in forest inventory / Treyfeld R.F., Berezin V.I. / / Aerospace methods and GIS technology in forestry. Reports of the IV International Conference (Moscow, 17-19 April 2007). - P. 140-144.
13. Sakhatsky A. I. Classification of space image for forest state identification within the Siberia region / Sakhatsky A. I., McCallun J., Khodorovsky A. Ja., Bujanova I. Ja. // Pt. 1 IIASA, Laxenburg, Austria, IR-02-09, April 2002. – 45 p.
14. http://www.lvivlis.com.ua/file/LL_proekt.pdf.
15. http://ukrref.com.ua/?id=MTU3MDc%3D.
16. http://download.intergraph.com/?ProductName =ERDAS%20IMAGINE&current=true&legacy=true.

Hybrid classification of forests on high resolution satellite images

Burshtynskaya K. V., Polishchuk B. V.,
Fikovskaya A. A.

Hybrid classification technique, which is based on determining the optimal number of classes using the method of not control classification with subsequent image processing by the methods for controlled classification is presented in the work.

Criterion for determining the optimal number of classes is proposed. To investigate the effectiveness of identifying different types of forests in the Carpathian region it was used satellite image obtained from the satellite Ikonos. Influence of the size of training samples and classification methods: maximum likelihood, Mahalanobis, minimum distance on the of accuracy of forests identification was analyzed