VILNIUS GEDIMINAS TECHNICAL UNIVERSITY

Jurgita MILIEŠKAITĖ

THE CLUSTERIC CORRELATION ANALYSIS OF LAND COVER DIGITAL RASTER IMAGES

SUMMARY OF DOCTORAL DISSERTATION

TECHNOLOGICAL SCIENCES,
MEASUREMENT ENGINEERING (10T)

Vilnius LEIDYKLA TECHNIKA 2012

Doctoral dissertation was prepared at Vilnius Gediminas Technical University in 2008–2012.

Scientific Supervisor

Prof Dr Habil Jonas SKEIVALAS (Vilnius Gediminas Technical University, Technological Sciences, Measurement Engineering – 10T).

The dissertation is being defended at the Council of Scientific Field of Measurement Engineering at Vilnius Gediminas Technical University:

Chairman

Prof Dr Eimuntas Kazimieras PARŠELIŪNAS (Vilnius Gediminas Technical University, Technological Sciences, Measurement Engineering – 10T).

Members:

Assoc Prof Dr Vladislovas Česlovas AKSAMITAUSKAS (Vilnius Gediminas Technical University, Technological Sciences, Measurement Engineering – 10T),

Assoc Prof Dr Mindaugas JUREVIČIUS (Vilnius Gediminas Technical University, Technological Sciences, Measurement Engineering – 10T),

Prof Dr Habil Kazys KAZLAUSKAS (Vilnius University, Physical Sciences, Informatics – 09P),

Dr Habil Saulius ŠLIAUPA (Nature Research Center, Physical Sciences, Geology – 05P).

Opponents:

Assoc Prof Dr Giedrė BECONYTĖ (Vilnius University, Physical Sciences, Physical Geography – 06P),

Prof Dr Habil Algimantas ZAKAREVIČIUS (Vilnius Gediminas Technical University, Technological Sciences, Measurement Engineering – 10T).

The dissertation will be defended at the public meeting of the Council of Scientific Field of Measurement Engineering in the Senate Hall of Vilnius Gediminas Technical University at 2 p. m. on 19 December 2012.

Address: Saulėtekio al. 11, LT-10223 Vilnius, Lithuania.
Tel.: +370 5 274 4952, +370 5 274 4956; fax +370 5 270 0112;
e-mail: doktor@vgtu.lt

The summary of the doctoral dissertation was distributed on 16 November 2012.

A copy of the doctoral dissertation is available for review at the Library of Vilnius Gediminas Technical University (Saulėtekio al. 14, LT-10223 Vilnius, Lithuania).

VILNIAUS GEDIMINO TECHNIKOS UNIVERSITETAS

Jurgita MILIEŠKAITĖ

ŽEMĖS DANGOS SKAITMENINIŲ RASTRINIŲ VAIZDŲ KLASTERINĖ KORELIACINĖ ANALIZĖ

DAKTARO DISERTACIJOS SANTRAUKA

TECHNOLOGIJOS MOKSLAI,
MATAVIMŲ INŽINERIJA (10T)

VGTU
LEIDYKLA
Vilnius TECHNIKA 2012

Disertacija rengta 2008–2012 metais Vilniaus Gedimino technikos universitete.
Mokslinis vadovas

prof. habil. dr. Jonas SKEIVALAS (Vilniaus Gedimino technikos universitetas, technologijos mokslai, matavimų inžinerija – 10T).

Disertacija ginama Vilniaus Gedimino technikos universiteto Matavimų inžinerijos mokslo krypties taryboje:

Pirmininkas

prof. dr. Eimuntas Kazimieras PARŠELIŪNAS (Vilniaus Gedimino technikos universitetas, technologijos mokslai, matavimų inžinerija – 10T).

Nariai:

doc. dr. Vladislovas Česlovas AKSAMITAUSKAS (Vilniaus Gedimino technikos universitetas, technologijos mokslai, matavimų inžinerija – 10T),

doc. dr. Mindaugas JUREVIČIUS (Vilniaus Gedimino technikos universitetas, technologijos mokslai, matavimų inžinerija – 10T),

prof. habil. dr. Kazys KAZLAUSKAS (Vilniaus universitetas, fiziniai mokslai, informatika – 09P),

habil. dr. Saulius ŠLIAUPA (Gamtos tyrimų centras, fiziniai mokslai, geologija – 05P).

Oponentai:

doc. dr. Giedrė BECONYTĖ (Vilniaus universitetas, fiziniai mokslai, fizinė geografija – 06P),

prof. habil. dr. Algimantas ZAKAREVIČIUS (Vilniaus Gedimino technikos universitetas, technologijos mokslai, matavimų inžinerija – 10T).

Disertacija bus ginama viešame Matavimų inžinerijos mokslo krypties tarybos posėdyje 2012 m. gruodžio 19 d. 14 val. Vilniaus Gedimino technikos universiteto senato posėdžių salėje.
Adresas: Saulėtekio al. 11, LT-10223 Vilnius, Lietuva.
Tel.: (8 5) 274 4952, (8 5) 274 4956; faksas (8 5) 270 0112;
el. paštas doktor@vgtu.lt
Disertacijos santrauka išsiuntinėta 2012 m. lapkričio 16 d.
Disertaciją galima peržiūrėti Vilniaus Gedimino technikos universiteto bibliotekoje (Saulėtekio al. 14, LT-10223 Vilnius, Lietuva).
VGTU leidyklos „Technika“ 2045-M mokslo literatūros knyga.

Introduction

Topicality of the problem. Land gradually turns into more and more important non-reproducible resource of any state. Because of development of cities, growing of road infrastructure and changing agricultural policy in Lithuania and the European Union, increasing risks menace both to the environment and the land structure and their management is possible through effective monitoring only. Accounting of the Republic of Lithuania land fund is one of the key tools for monitoring the changes of land structure, so its reliability and relevance are of a particular importance for users of statistical data on land. Land accounting is important for a number of spheres, such as environmental protection, agriculture, cadastre, geodesy, regional development and so on, thus monitoring of changes of the land fund and improvement of this process are required for systematic accomplishment of miscellaneous strategic targets set by the state.

At present, the data on the Earth's surface are mostly obtained by remote methods, i.e. aerial photos and space images that are processed and analyzed upon applying advanced GIS technologies. Development of algorithms for classification of digital raster images was started long ago and they are improved on a permanent basis, because no method for clustering among abundant ones is universal. Methods usable for classification are not complicated; however, some of them require abundant computing operations and computer resources. In order to ensure successful results of classification, it is proposed to choose digital images of a higher resolution that improve the accuracy of the image; however, they cause increasing of the classification expenses.

Relevance of the work. Images obtained upon applying the remote methods present one of sources of information on the objects on the Earth's surface; however, practical use of the said images is not an old tradition in Lithuania. Images provided in digital form enable using various (complicated) computer algorithms for singling out information of various types. The gap of using the said images in Lithuania includes identification (perception) of objects and application of the obtained data in solving practical tasks. Information perception is a result of the long and complicated data acquisition process.

The methods of classification proposed in the work provide an effective and fast-operating tool for identifying (processing) digital raster images. They do not require abundant computer resources and computing operations. In addition, they are successfully applicable for processing digital images of lower resolution as well.

Objective of the work. Identification of land cover digital raster images upon application of clustering correlation analysis and created classification algorithms and methods.

Aim of the work. To develop a new methodology for identification of land cover digital images upon application of clustering correlation analysis and to propose the methodology of application of covariance functions in digital image identification procedures.

Tasks of the work

1. To analyze upon methods for identification of digital raster images and algorithms applicable for image identification.
2. To carry out a theoretical and experimental research on clustering raster images upon application of the theory of covariance functions.
3. To develop a methodology for classification of digital images upon application of results of clustering correlation analysis (to develop a system for classification of raster images).

Methods of the work. Theoretical and experimental research works upon application of developed and applicable specialized GIS information technologies as well as methods of mathematical statistics for forecasting and assessment of their impact upon the results of classification digital images.

Scientific novelty

In the dissertation, the following new results for measurement engineering were achieved:

1. The method of clustering correlation analysis for identification of digital images upon simulation of parameters of the range of RGB colors and the theory of covariance functions were proposed.
2. The impact of relevant parameters (the frequencies of RGB colors, such as contrast) on the quality of the results of identification of digital images was established.

Practical value. The results of the research disclosed new opportunities for identification (analysis) of digital raster images.

Defended propositions

1. The automated identification of digital raster images enables considerable reducing the usual image processing expenses.

2. The method of application of covariance functions empowers to measure the impact of variations of the tensor's of RGB spectrum parameters upon the results of identification.

The scope of the scientific work. The dissertation consists of the following parts, namely the introduction, four chapters and summary of the results, two supplement is included as well.

The dissertation contains 146 pages, apart from the supplements; there were used 45 numbered equations, 57 figures and 22 table. There were used 164 literature sources when working on the dissertation.

1. The review of identification of digital raster images upon application of clustering analysis

In the chapter, the concept of clustering analysis, its structure as well as methods of clustering analysis and their classification are reviewed. The spheres of application of clustering analysis are specified. The usable most popular methods of clustering analysis and opportunities of application of digital images are reviewed. The properties of presently available space images are discussed upon. GIS methods applicable for analysis of digital raster images that are used in world practice are described. The problems of digital images' classification are identified and the tasks to be solved in the dissertation are formulated.

2. Application of correlation method for identification of digital images

In the chapter, the theoretical assumptions of correlation method and covariance functions as well as the opportunities of application of matlab software in digital image analysis are reviewed. Because images under classification often are affected by a noise to a certain extent, it was proposed to eliminate such noise upon application the Wavelet theory.

While analyzing identification of digital images we will base on the theory of co-variance. A correlation coefficient of two random variables (pixels) X and Y is calculated according to:

$$r_{xy} = \frac{K_{xy}}{m_x m_y}, \tag{1}$$

where K_{xy} – covariance influence, m_x, m_y – standard deviation influence. The values of correlation coefficient vary in the range $-1 \leq r_{xy} \leq 1$. A strong correlation between random variables (in this case, between digital images) takes place, when $r_{xy} = 1$.

Covariance influence under the available sample members is calculated:

$$K_{xy} = \frac{1}{n}\sum_{1}^{n} x_i y_i - \overline{xy}, \tag{2}$$

where x_i, y_i – two samples of the respective members, $\overline{x}, \overline{y}$ – sample averages, n – sample volume. Sample members are pixel color values.

Following available vector data of random variables the covariance matrix influence is calculated:

$$K_{\mathbf{XY}} = \frac{1}{(n-1)}\delta X^T \delta Y, \tag{3}$$

where δX and δY – are centered matrices of two random variables which equal:

$$\delta X = X - MX, \tag{4}$$

$$\delta Y = Y - MY, \tag{5}$$

where X, Y – is a pixel matrix (an array) of certain digital image, MX – is a matrix of vector averages in matrix X, MY – is a matrix of vector averages in matrix Y.

When the values X and Y are independent, the covariance K_{XY} = 0 (too r_{xy} = 0).

The estimates of covariance matrices are reduced into estimates R_{xy} of the correlation coefficients' matrix:

$$R_{xy} = D_X^{-1/2} K_{xy} D_Y^{-1/2}, \tag{6}$$

$$D_X = \frac{1}{n-1}\left\{\delta X^T \delta X\right\}_{diag}, \tag{7}$$

$$D_Y = \frac{1}{n-1}\left\{\delta Y^T \delta Y\right\}_{diag}, \tag{8}$$

where D_X, D_Y – the diagonal matrices of the principal diagonal members of the estimates of covariance matrices.

Stationary random functions are formed of digital images. Auto- and intercovariance functions of the said random functions were analyzed upon applying different quantization intervals for image pixels. For covering the whole surface of a digital image, a translational covariance function is applied.

When covariance function is applied in an analysis of digital images, the continuous covariance function $K_h(\tau)$ of two segments of one digital image or pixel arrays of segments of two images – $h_1(u)$ ir $h_2(u+\tau)$, considered realizations of random functions where pixels are arguments, will be expressed as follows:

$$K_h(\tau) = \frac{1}{T-\tau} \int_0^{T-\tau} \delta h_1(u) \delta h_2(u+\tau) du, \tag{9}$$

where $\delta h_1(u), \delta h_2(u+\tau)$ – the centered segments of pixel parameters, u – the pixel parameter of the segment, T – the length of the segment in relative units, $\tau = k \cdot \Delta$ – the varying quantization interval, Δ – the value of the pixel parameter, k– the number of pixels in the quantization interval.

The estimate $K'_h(\tau)$ of the covariance function $K_h(\tau)$ was computed on the base of the available measurement data as follows:

$$K'_h(\tau) = K'_h(k) = \frac{1}{n-k} \sum_{i=1}^{n-k} \delta h_1(u_i) \delta h_2(u_{i+k}), \tag{10}$$

where n – the total number of discrete intervals.

The formula (10) is applicable in a form of auto- or intercovariance function. When the function is an autocovariance function, the segments $h_1(u)$ ir $h_2(u+\tau)$ are segments of a single digital image and when the function is an intercovariance function, the segments belong to two different images.

The estimate of a normed covariance function is equal:

$$R'_h(k) = \frac{K'_h(k)}{K'_h(0)} = \frac{K'_h(k)}{\sigma'^2_h}, \tag{11}$$

where σ'_h – the estimate of the standard deviation of random function.

For eliminating the column trend of the i-th measurement array, the following formulas are used:

$$\delta H_i = H_i - e \cdot \bar{h}_i^T = (\delta h_{i1}, \delta h_{i2}, \ldots, \delta h_{im}), \tag{12}$$

where δH_i – the i-th array of reduced pixels of a digital image where the column trend is eliminated; H_i – the array of pixels of i-th image; e – the unit vector with the sizes (n×1), n – the number of lines of the i-th array, $\bar{h}_i$ – the mean vector of columns of the i-th pixel array, δh_{ij} – the j-th column (vector) of the i-th array of reduced pixels.

The mean vector of columns of the i-th pixel array is computed according to the following formula:

$$\bar{h}_i^T = \frac{1}{n} e^T \cdot H_i = \frac{1}{n} H_i^T \cdot e, \qquad (13)$$

A realization of the random function of the *j*-th column in the digital image *i*-th pixel array is expressed in vectorial form as follows:

$$\delta h_{ij} = \left(\delta h_{ij,1}, \ldots, \delta h_{ij,m}\right), \qquad (14)$$

The estimate of covariance matrix for the digital image i-th pixel array is:

$$K'(\delta H_i) = \frac{1}{n-1} \delta H_i^T \delta H_i, \qquad (15)$$

The estimate of covariance matrix for two pixel arrays of one digital image or a pixel array of two digital images shall be written as follows

$$K'\left(\delta H_i, \delta H_j\right) = \frac{1}{n-1} \delta H_i^T \delta H_j, \qquad (16)$$

where the sizes of arrays $\delta H_i, \delta H_j$ should be the same.

Upon applying the theory of covariance functions, the impact of components of *RGB* colours spectrum upon the expressions of the covariance functions of digital images was examined; in addition, the expressions of the covariance functions of digital images were assessed according to the *RGB* colours continuous spectrum in the meaning of tensor of colours. The changes of the values of correlation coefficients in matrices are shown as a comparative graph.

The estimates $K'(\delta H_i)$ and $K'\left(\delta H_i, \delta H_j\right)$ of covariance matrices are reduced into estimates $R'(\delta H_i)$ and $R'\left(\delta H_i, \delta H_j\right)$ of matrices of correlation coefficients:

$$R'(\delta H_i) = D_i^{-1/2} K'(\delta H_i) D_i^{-1/2}, \qquad (17)$$

$$R'\left(\delta H_i, \delta H_j\right) = D_{ij}^{-1/2} K'\left(\delta H_i, \delta H_j\right) D_{ij}^{-1/2}, \qquad (18)$$

where D_i, D_{ij} – the diagonal matrices of the principal diagonal members of the estimates of covariance matrices $K'(\delta H_i)$ and $K'\left(\delta H_i, \delta H_j\right)$, respectively.

The accuracy of computed correlation coefficients is defined by the standard deviation σ_r, that’s value is found according to the following formula:

$$\sigma_r = \frac{1}{\sqrt{n}}\left(1 - r^2\right) \tag{19}$$

where n – the number of pixels, r – the computed correlation coefficient. The maximum value of the standard deviation is obtained, when the value of r is close to zero.

3. The experimental procedures for identifying digital raster images

Identification of changes of the Earth's surface caused by various factors (such as human activities, erosion and so on) in a locality takes much time. Supplemental devices for fixing and assessing such changes of the Earth's surface are required. In this chapter, an opportunity to identify a similarity (dependence) of digital images (for example, formed in different moments) in an automated way is provided.

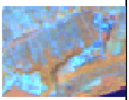
In the procedures related to digital images, couples of cutouts of digital images from orthographic M1:10 000 maps (in *TIFF* file) and *LANDSAT* wave spectrum images were used. Two computer programs were offered for the computations. Upon application of the said programs, identity of the digital images under comparison is assessed (established) according to the computed values of correlation coefficients (*Pavkor.m*) and normed expressions of covariance functions (*Pkor.m*).

The analysis of the digital images of the Earth's surface was performed in *RGB* colours spectrum using its single components, such as red (*R*), green (*G*), blue (*B*) and the whole *RGB* spectrum in the meaning of its tensor. On the base of the obtained results (a part of them is provided in the Table 1), it was found that in analysis of digital images, the values of correlation coefficients were different upon using single components and the whole tensor of *RGB* colours. The low value of the correlation coefficient for a single component of the spectrum (for example, the green (*G*) obtained between c1 and c2 from the table 1) shows that pixels of the said colours are not abundant in the digital images under analysis.

On establishing correlation (identity) of digital images, application of the whole tensor of *RGB* spectrum is purposeful.

The obtained expressions of the normed autocovariance functions upon applying single components of *RGB* colours are different as well (Fig. 1, a, b, c). The expressions of the normed autocovariance functions according to the blue (Fig. 1, c) and the green (Fig. 1, b) components of *RGB* colours are close, i.e. the values of the said functions vary from 1,0, when the quantization interval $k=1$, up to the value close to zero, when $k=300$.

Table 1. The values of correlation coefficients upon applying the whole tensor of *RGB* colours and its single components between the cutouts a1 and a2 of orthographic image (two different forest's cutouts) and cutouts b1 and b2 (cutouts of arable and built-up territories) and between cutouts c1 and c2 of *LANDSAT* photo (cutouts of a leafy forest and built-up territory).

Couples of cutouts of digital images	The values of correlation coefficients			
	Tensor of *RGB* colours	Red *(R)*	Green (*G*)	Blue (*B*)
 a1  a2	**0,43**	0,26	0,14	0,18
 b1  b2	**0,16**	0,12	0,10	0,12
 c1  c2	–**0,26**	0,04	–0,18	–0,04

The expressions of the normed autocovariance and intercovariance functions upon applying the whole tensor of RGB colours are shown in Fig. 2. The expressions of the normed autocovariance functions for cutouts of *LANDSAT* images (Fig. 2, the curves a and b) are similar, i.e. the values of these covariance functions vary from 1,0 (when the initial interval of quantization $k=1$) to the value close to zero. The normed intercovariance function of two image cutouts varies between -0,17 and 0,2 „chaotically“ (Fig. 2, the curve c). The correlation between two cutouts of satellite digital images is very weak, and this fact attests that the Earth's surface is different.

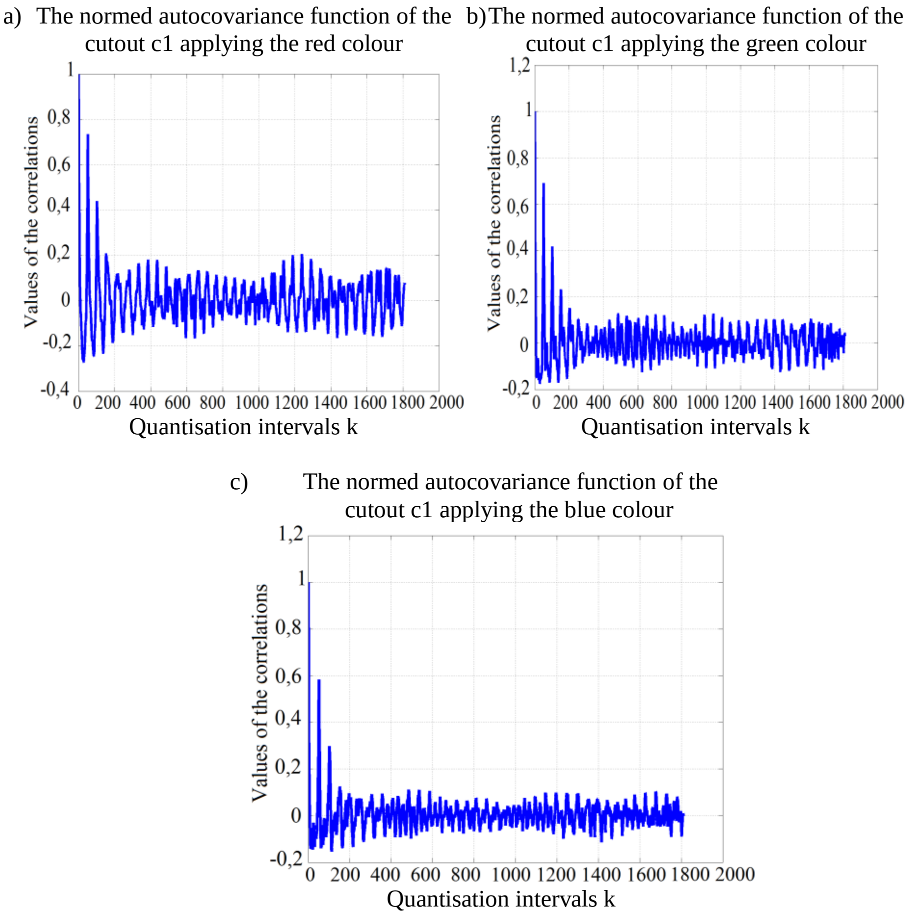


Fig. 1. The normed autocovariance functions of the cutout c1 from *LANDSAT* image (Table 1) upon applying: a – the red colour of *RGB* spectrum; b – the green colour of *RGB* spectrum; a – the blue colour of *RGB* spectrum

The curve of changes of correlation coefficients between pixel columns for two cutouts of *LANDSAT* image (c1 and c2 in the Table 1) is shown in Fig. 3. In the curve, the obtained correlation coefficients between the pixel columns vary as follows: the 1st pixel column of the cutout c1 with other columns of the cutout c2; the 2st pixel column of the cutout c1 with other columns of the cutout c2; the 3st pixel column of the cutout c1 with other columns of the cutout c2.

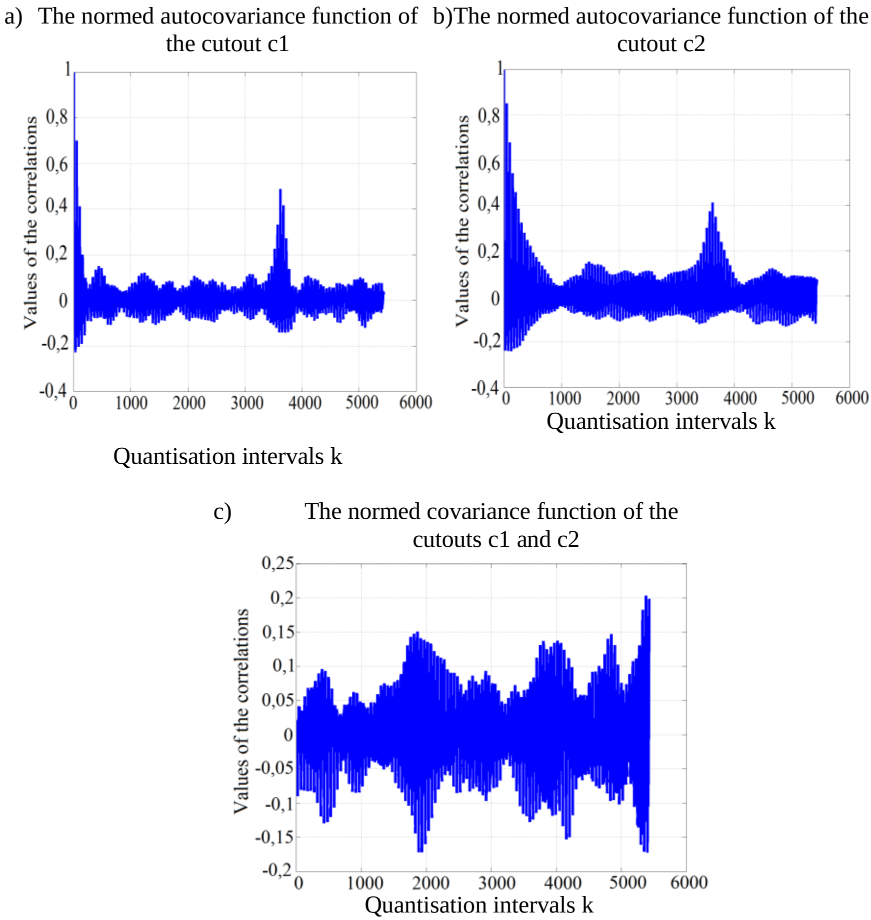


Fig. 2. The normed autocovariance and covariance functions of the cutouts of *LANDSAT* images (c1 and c2 in the Table 1) upon applying the whole tensor of *RGB* colours: a – the normed autocovariance function of the cutout c1; b – the normed autocovariance function of the cutout c2; c – the normed covariance function of the cutouts c1 and c2

Scanning equipment of *LANDSAT* satellites registers the reflection of the Earth's surface in seven ranges. Information channels may be adjusted and shown upon using three principal colours (blue, green and red). The data on each channel are provided as one of the principal colours, and different colours are presented by various proportions of the principal colours according to the relative luminosity (i.e. the digital values of the image elements of the respective channel.

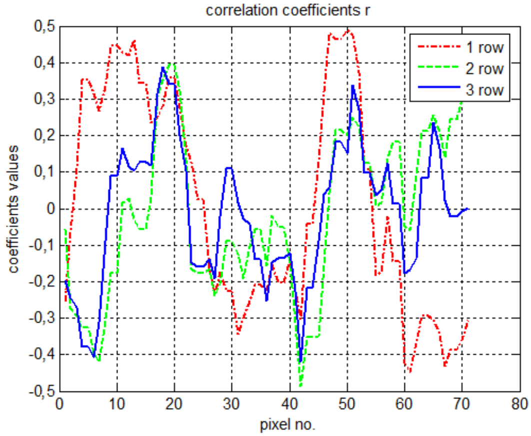


Fig. 3. The changes of the correlation coefficients between pixel columns of two cutouts of *LANDSAT* image (c1 and c2 in the Table 1).

LANDSAT satellite images used for the investigation were formed of the following combinations of the spectrum ranges: the ranges 2, 3 and 4; the ranges 4, 5 and 3; the ranges 3, 4 and 5; the ranges 5, 4 and 3; the ranges 1, 2 and 3.

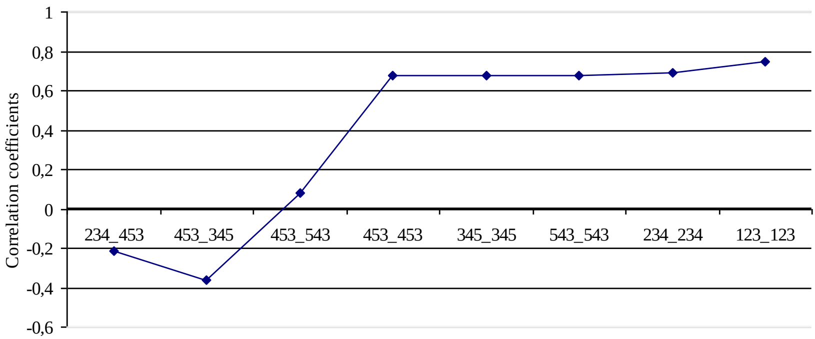


Fig. 4. The values of the correlation coefficients obtained after the analysis of *LANDSAT* images formed of combinations of different spectrum ranges

It was found that the minimum values of correlation coefficients (-0,22, -0,36, 0,08) were obtained when couplings of images under the analysis were formed of combinations of different spectrum ranges (Fig. 4), and the maximum values of correlation coefficients (0,68, 0,69, 0,75) were obtained when couplings of images under the analysis were formed of combinations of same spectrum ranges. Consequently, for establishing correlation (similarity) of multispectral digital images, couples of images (or cutouts) formed of combinations of same spectrum ranges should be used.

In this chapter, the visual impact of the quality of the digital image upon the results of image identification was discussed upon. The impact of contrast and luminosity was assessed on comparison of the results obtained on analysis of cutouts of digital images of the same contrast with the results obtained on analysis of cutouts of digital images of the different contrasts. Taking into account the obtained results, it was found that the quality of orthophotographic images and *LANDSAT* satellite images (their contrast and luminosity) does not cause a considerable impact upon the values of correlation coefficients. The values of correlation coefficients on changing the contrast or luminosity of orthophotographic images varied between 0,01 and 0,08, and the values of correlation coefficients o *LANDSAT* images – between 0,01 and 0,02.

4. Application of clustering correlation analysis in procedures of automated land cadastre accounting

A classification of a multispectral digital image upon applying supervised classification according to two *SMAP* (Sequential Maximum a Posteriori) and *MCL* (Maximum Likelihood Classifier) algorithms was provided.

The objective of the experiment included automated (upon using supervised classification) grouping of pixels of satellite image into the given the Earth's surface classes (standards) and establishing the accuracy of such classification. The accuracy of the obtained results depends on correctly decoded zones of the Earth's surface and the number of the chosen classes.

The experimental tasks (classification of digital images according to the chosen classes) were performed upon applying *Quantum GIS (QGIS)* software together with geographic information system *GRASS*. *LANDSAT* digital raster images dated to the year 1994 were used. For visual identification (singling out) the landed property, *LANDSAT* image formed of the spectrum ranges 4, 5 and 3 was chosen.

In the environment of *QGIS* software, *LANDSAT* image the Earth's surface areas (according to the colours) were attributed to the relevant class (marshes, leafy forests, water, bushes, peatbogs, coniferous woods, grasslands, abandoned fields, surfaces without flora and so on), i.e. the preset meanings of the standard pixels.

Then, after choosing *SMAP* algorithm in *GRASS* information system, the classification of *LANDSAT* image was performed, i.e. the values of the image pixel were attributed to one of the above-mentioned classes. In another case, classification of *LANDSAT* image was performed upon choosing *MLC* algorithm in the said information system. So, two images classified according to *SMAP* algorithm and according to *MLC* algorithm, respectively, were obtained.

In the Table 2, the areas of the classes computed from the classified images are provided.

Table 2. The areas of the classes (in hectares) computed from the classified images

Classes	*MLC*	*SMAP*
	The areas of the classes (in hectares) computed from the classified images	
Marshland	48,04	31,90
Leafy forest	586,.38	604,41
Water	69,43	69,40
Scrubland	593,00	582,86
Peatland	277,49	243,88
Coniferous forests	336,66	334,06
Grassland	194,29	124,73
Abandoned field	840,52	899,70
Surface without vegetation	2571,63	2626,50
Total:	**5517,44**	**5517,44**

It may be seen from the Table 2 that the areas of individual classes (lands) in the classified images are different; however, the total area in both images is the same and equals to 5517,44 hectares.

The results of the classification were assessed upon using the layer of test points developed in *QGIS* software. Upon applying segmenting algorithms *SMAP* and *MCL*, it was strived to assess whether pixels of a space image were duly grouped into the set classes of the Earth's surface (standards). For each test point (the total number of the chosen test points was 500), the relevant class was manually attributed, taking into account the location of the points in *LANDSAT* digital image. The said procedure requires a particular experience of the operator (specialist) as well as the skills in identifying the Earth's surface from *LANDSAT* digital image. Then *QGIS* software attributed the relevant class to the same test points in the automated way, taking into account their location in the classified images (i.e. the class they „lay on"). According to the idea, the locations of the test points established manually from *LANDSAT* digital raster image and established from the classified images in the automated way should coincide (for example, if a test point of *LANDSAT* image is located in a marsh, the said point in the images classified upon applying *SMAP* and *MCL* should „lie" to the class of marsh). It was found that a higher percentage of coincidence of the test points (77,40 %) was computed in the image classified upon applying *SMAP* segmenting algorithm.

The dissertation presented in this chapter showed that remote observance and *GIS* technique could be excellent tools for identifying the types of the

Earth’s surface and computing the areas. However, the results of the research should be assessed on the site as well. In addition, the research requires good skills in decoding.

The research should be practically applicable in a case when it is strived to identify areas of specific zones of lands (ha) from satellite images and to make conclusions about the changes of the Earth’s surface on their base. The methodology for image identification according to the computed values of correlation coefficients and the normed expressions of covariance functions provided in the chapter Three does not require skills in decoding. The said methodology is an effective and rapid tool for identifying changes of the Earth’s surface (such as various lands, shores of water bodies and so on) and their fluctuations in course of time. For this purpose, digital images of different periods and duly specified arrays of the digital images (that’s interdependence should be established) are required only.

General conclusions

1. The method of clustering correlation analysis was proposed where a similarity of digital images is assessed according to the computed values of correlation coefficients upon formation the summarized estimate of a matrix of correlation coefficients. When a positive value of the correlation coefficient is higher, the level of similarity of digital images is higher (the interdependence is stronger) and vice versa). In the computing procedures, large arrays of digital images (≈ 40 000 pixels) were used, so the accuracy of establishment of correlation coefficients was high enough: $\sigma_r \approx 0{,}004$.
2. The variation of the level of interdependence of digital images was assessed according to the variation of the values of their autocovariance and intercovariance functions upon applying different quantization intervals. The expressions of the autocovariance and intercovariance functions show the rate of reducing of covariance between pixel columns of the cutout on increasing the quantization intervals between the columns. The value of the normed autocovariance function in *LANDSAT* satellite images is close to zero, when the quantization interval $k \rightarrow 100$. The expressions of the normed intercovariance functions are irregular („chaotical“) and their values are small: $r \rightarrow 0{,}1–0{,}2$.
3. The impact of the components of the spectrum of digital images upon the estimates of correlation parameters established out as well. When the tensor of *RGB* and single components of it – red, green and blue are used in procedures for identification of digital images, different values of correlation coefficients are obtained. When single colour

components are applied, the differences of the values of correlation coefficients vary between 0 and 0,12 (analyzing orthophotographic images), and between 0 and 0,24 (analyzing *LANDSAT* satellite images). A program for presenting the scatter of values of the digital image correlation matrix (in percent) in the digital and graphic form was developed. Similarity of digital images was classified to a certain number of classes according to the accepted width of the interval of correlation coefficients.

4. The impact of the principal parameters of the quality of the digital images, i.e. their contrast and luminosity, upon the results of identification was assessed. It was found that fluctuations of values of the said parameters do not affect the final results of the identification.
5. The results of the researches of multispectral digital images showed that seeking to establish an identity of multispectral digital images (*LANDSAT)* according to the values of correlation coefficients, couples of images (or cutouts) formed of combinations of same bands of the spectrum should be used.
6. The results of image classification show that a particularly reliable identification of homogenous zones of the Earth's surface is ensured upon applying segmenting algorithms *SMAP* and *MLC*, i.e. waters (100%), peat bogs (96%) and coniferous woods (92%) are particularly well identified upon applying *SMAP* segmenting algorithm. The average value of the correlation coefficient upon applying *SMAP* segmenting algorithm was equal to 0,73 and upon applying *MLC* segmenting algorithm was equal to 0,69.

List of Published Works on the Topic of the dissertation
In the reviewed scientific periodical publications

Milieškaitė, J. 2010. The application of covariation method to the analysis (comparable) of digital images in cartography, *Geodesy and Cartography*, 36(1): 30–34 (in Lithuanian). ISSN 1392-1541 (Compendex)

Milieškaitė, J.; Vaitkus, G. 2011. Landsat TM satellite image segmentation comparison of methods in land cover classification, *Geodesy and Cartography*, 37(3): 135–142 (in Lithuanian). ISSN 2029-7009 (Compendex)

Milieškaitė, J. 2011. The application of covariance method while analysing the digital images of land surface, *Geodesy and Cartography*, 37(3): 105–110 (in Lithuanian). ISSN 2029-7009 (Compendex)

In the other editions

Milieškaitė, J. 2011. The application of covariance method while analysing the digital images of land surface, In *Proceedings of the 8th International Conference*

„Environmental Engineering“ Selected papers, Vol. III, 19–20May [8-osios tarptautinės konferencijos „Aplinkos inžinerija“, įvykusios Vilniuje 2011 m. gegužės 19–20 d., medžiaga], 1389–1393. ISBN 978-9955-28-827-5.

About the author

Jurgita Milieškaitė acquired the bachelor‘s qualification degree in measurement engineering in Vilnius Gediminas Technical University in 2004, and in 2006 she acquired master‘s qualification degree in measurement engineering. From 2008 to 2012 is a doctor student at Vilnius Gediminas Technical University.

Leter of thanks

The author expresses sincere thanks for the original scientific guidance, remarks and suggestions on the dissertation to the supervisor of the thesis Prof Dr Habil Jonas Skeivalas. The author is grateful the Geodesy and Cadastre Department of Vilnius Gediminas Technical University for the support.

ŽEMĖS DANGOS SKAITMENINIŲ RASTRINIŲ VAIZDŲ KLASTERINĖ KORELIACINĖ ANALIZĖ

Problemos formulavimas. Žemė tampa vis svarbesniu kiekvienos valstybės neatkuriamuoju ištekliu. Dėl vykstančios miestų plėtros, kelių infrastruktūros augimo bei besikeičiančios Lietuvos ir Europos Sąjungos žemės ūkio politikos tiek visai aplinkai, tiek žemės struktūrai iškyla vis daugiau grėsmių, kurias galima valdyti tik efektyvios stebėsenos dėka. Lietuvos Respublikos žemės fondo apskaita yra vienas iš esminių įrankių žemės struktūros pokyčių kaitai stebėti, todėl jos patikimumas bei aktualumas yra itin svarbus statistinių duomenų apie žemę naudotojams. Žemės apskaita svarbi tokioms sritims kaip aplinkosauga, žemės ūkis, kadastras, geodezija regioninė plėtra bei daugeliui kitų sričių, todėl žemės fondo dangos kitimų stebėjimas ir šio proceso tobulinimas yra būtinas, siekiant sistemiškai įgyvendinti valstybės užsibrėžiamus įvairialypius strateginius tikslus.

Duomenų apie žemės dangą šiuo metu daugiausia gaunama, naudojant distancinius metodus – aerofotonuotraukas ir kosminius vaizdus, kurie apdorojami ir analizuojami, naudojant pažangias GIS technologijas. Skaitmeninių rastrinių vaizdų klasifikavimo algoritmai pradėti kurti gana seniai ir jie kuriami, tobulinami nuolat, kadangi tarp daugelio klasterizavimui sukurtų metodų nėra nė vieno universalaus. Kai kurie klasifikavimui naudojami metodai reikalauja daug skaičiavimo operacijų bei kompiuterinių resursų. Siekiant sėkmingų klasifikavimo rezultatų, siūloma pasirinkti didesnės skiriamosios gebos skaitmeninius vaizdus, kurie padidintų ne tik vaizdo klasifikavimo tikslumą, bet ir klasifikavimo išlaidas.

Darbo aktualumas. Distanciniais metodais gaunami vaizdai yra vienas iš informacijos šaltinių apie žemės paviršiuje esančius objektus, tačiau praktinis šių vaizdų naudojimas Lietuvoje yra dar nesenas reiškinys. Skaitmenine forma pateikti vaizdai leidžia naudoti įvairius (sudėtingus) kompiuterinius algoritmus daugelio tipų informacijai išskirti. Šių vaizdų panaudojimas Lietuvoje yra nepakankamas, todėl aktuali problema objektų identifikavimas (jų suvokimas) bei gautų žinių pritaikymas sprendžiant praktinius uždavinius. Informacijos suvokimas yra ilgo ir sudėtingo duomenų gavybos proceso rezultatas.

Šiame darbe pasiūlyta klasifikavimo metodika yra efektyvi ir greita priemonė skaitmeniniams rastriniams vaizdams identifikuoti (apdoroti). Ji nereikalauja didelių kompiuterinių resursų, skaičiavimo operacijų ir puikiausiai taikoma mažos rezoliucijos skaitmeniniams vaizdams apdoroti.

Tyrimų objektas. Žemės dangos skaitmeninių rastrinių vaizdų identifikavimas, taikant klasterinę koreliacinę analizę, ir sukurtus klasifikavimo algoritmus bei metodus.

Darbo tikslas. Sukurti metodus žemės dangos skaitmeniniams vaizdams identifikuoti, taikant klasterinę koreliacinę analizę, pasiūlyti kovariacinių funkcijų taikymo metodiką skaitmeninių vaizdų identifikavimo procedūrose.

Darbo uždaviniai

1. Išnagrinėti skaitmeninių rastrinių vaizdų identifikavimo metodus ir taikomus algoritmus.
2. Atlikti rastrinių vaizdų klasterizavimo teorinį ir eksperimentinį tyrimą, taikant kovariacinių funkcijų teoriją.
3. Sudaryti skaitmeninių vaizdų klasifikavimo metodiką, panaudojant klasterinės koreliacinės analizės rezultatus (parengti rastrinių vaizdų klasifikavimo sistemą).

Tyrimų metodika. Disertacijoje naudojami teoriniai ir eksperimentiniai tyrimai, taikant specializuotas geoinformacines technologijas, matematinius statistinius prognozavimo metodus ir įvertinant jų įtaką skaitmeninių vaizdų klasifikavimo rezultatams.

Darbo mokslinis naujumas

Disertaciniame darbe buvo gauti šie matavimų inžinerijos mokslui nauji rezultatai:

1. Pasiūlytas klasterinės koreliacinės analizės metodas skaitmeniniams vaizdams identifikuoti, taikant RGB spalvų spektro parametrų modeliavimą bei kovariacinių funkcijų teoriją.

2. Įvertinta skaitmeninių vaizdų atitinkamų parametrų (RGB spalvų dažniai, kontrastas) įtaka identifikavimo rezultatų kokybei.

Darbo rezultatų praktinė reikšmė. Tyrimų rezultatai atskleidė naujas žemės dangos skaitmeninių rastrinių vaizdų identifikavimo (analizės) galimybes.

Ginamieji teiginiai

1. Automatinis skaitmeninių rastrinių vaizdų identifikavimas leidžia ženkliai sumažinti įprastines vaizdų apdorojimo sąnaudas.
2. Kovariacinių funkcijų panaudojimo metodas įgalina įvertinti skaitmeninių vaizdų RGB spalvų tenzoriaus parametrų įtaką identifikavimo rezultatams.

Darbo apimtis

Disertaciją sudaro įvadas, 4 skyriai ir bendrosios išvados. Taip pat yra 2 priedai. Darbo apimtis yra 146 puslapiai, neskaitant priedų, tekste panaudotos 45 numeruotos formulės, 57 paveikslai ir 22 lentelių. Rašant disertaciją buvo panaudoti 165 literatūros šaltiniai.

Pirmasis skyrius skirtas mokslinės literatūros analizei. Jame nagrinėjami esami klasterizavimo metodai ir jų taikymas. Apžvelgta skaitmeninių rastrinių vaizdų struktūra ir jų analizės metodai, esami *GIS* metodai ir jų taikymas skaitmeninių vaizdų analizėje, ir suformuluoti uždaviniai, kuriuos disertaciniame darbe tikslinga išspręsti.

Antrajame disertacijos skyriuje pateiktas teorinis tyrimas. Aptariamos koreliacijos metodų teorinės prielaidos, kovariacinių funkcijų taikymas skaitmeninių vaizdų analizėje. Apžvelgtas Matlab programinės įrangos taikymas identifikavimo procedūrose, siūlomų programų sudarymo ypatumai.

Trečiajame disertacijos skyriuje pateikti skaitmeninių rastrinių vaizdų identifikavimo tyrimai, įvertinama skaitmeninių vaizdų spektro skirtingų juostų įtaka skaičiavimo rezultatams. Nustatyti skaičiavimų ypatumai keičiant skaitmeninių vaizdų parametrų reikšmes.

Ketvirtajame disertacijos skyriuje aprašomas klasterinės koreliacinės analizės taikymas žemės kadastro automatizuotos apskaitos procedūrose.

Bendrosios išvados

Remiantis disertaciniame darbe atliktais teoriniais ir eksperimentiniais tyrimais, gautos šios apibendrintos išvados:

1. Pasiūlytas klasterinės koreliacinės analizės metodas, kai žemės dangos skaitmeninių vaizdų panašumas įvertinamas pagal apskaičiuotas koreliacijos koeficientų reikšmes, sudarant apibendrintąjį koreliacijos

koeficientų matricos įvertį. Esant didesnei koreliacijos koeficiento teigiamai reikšmei skaitmeninių vaizdų tarpusavio panašumo lygmuo yra didesnis (priklausomybė stipresnė) ir atvirkščiai. Skaičiavimų procedūrose buvo panaudoti dideli skaitmeninių vaizdų masyvai (≈ 40 000 pikselių), todėl koreliacijos koeficientų nustatymo tikslumas buvo pakankamai aukštas: $\sigma_r \approx 0{,}004$.

2. Skaitmeninių vaizdų tarpusavio priklausomybės lygmens kaita buvo įvertinta pagal jų autokovariacinių ir tarpusavio kovariacinių funkcijų reikšmių kaitą, kintant šių funkcijų kvantavimo intervalams. Autokovariacinės ir tarpusavio kovariacinės funkcijos išraiškos parodo kokiu greičiu mažėja kovariacija tarp iškarpos pikselių stulpelių, didėjant kvantavimo intervalui tarp stulpelių. *LANDSAT* palydoviniuose vaizduose normuotos autokovariacinės funkcijos reikšmė artima nuliui, kai kvantavimo intervalas $k \to 100$. Tarpusavio normuotų kovariacijų funkcijų išraiškos yra nereguliarios („netvarkingos“) ir jų reikšmės nedidelės: $r \to 0{,}1–0{,}2$.
3. Nustatyta skaitmeninių vaizdų spektro spalvų dedamųjų įtaka koreliacijos parametrų įverčiams. Skaitmeninių vaizdų identifikavimo procedūrose naudojant *RGB* spalvų tenzorių bei pavienes *RGB* spektro spalvas – raudoną, žalią ir mėlyną gauname skirtingas koreliacijos koeficientų reikšmes. Naudojant pavienes spalvų dedamąsias, koreliacijos koeficientų reikšmių skirtumai gauti intervale nuo 0 iki 0,12 (analizuojant ortofotografinius vaizdus), ir nuo 0 iki 0,24 (analizuojant *LANDSAT* palydovinius vaizdus). Sudaryta programa skaitmeninių vaizdų koreliacijos matricų reikšmių sklaidai vaizduoti procentais skaitmeniniu ir grafiniu pavidalais. Skaitmeninių vaizdų tarpusavio panašumas klasifikuotas į tam tikrą klasių skaičių pagal priimtą koreliacijos koeficientų intervalo plotį.
4. Nagrinėta skaitmeninių vaizdų kokybės pagrindinių parametrų – ryškumo ir kontrasto įtaka skaitmeninių vaizdų identifikavimo rezultatams. Nustatyta, kad minėtų parametrų reikšmių kaita neturi reikšmingos įtakos galutiniams skaičiavimų rezultatams.
5. Daugiaspektrinių vaizdų tyrimų rezultatai parodė, kad siekiant nustatyti daugiaspektrinių skaitmeninių vaizdų (*LANDSAT)* tapatumą, pagal koreliacijos koeficientų reikšmes, būtina naudoti vaizdų ar iškarpų poras sudarytas iš tų pačių spektro juostų kombinacijų.
6. Vaizdų klasifikavimo rezultatai parodė, kad taikant *SMAP* ir *MLC* segmentavimo algoritmus ypač patikimai identifikuojami homogeniški žemės paviršiaus plotai, t. y. vandenų (100 %), durpių (96 %) bei spygliuočių miškų (92 %) plotai. Vidutinė koreliacijos koeficiento

reikšmė taikant *SMAP* segmentavimo algoritmą gauta 0,73, o taikant *MLC* segmentavimo algoritmą gauta 0,69.

Trumpos žinios apie autorių

Jurgita Milieškaitė 2004 metais Vilniaus Gedimino technikos universitete įgijo matavimų inžinerijos bakalauro kvalifikacinį laipsnį, o 2006 metais – matavimų inžinerijos mokslo magistro laipsnį. 2008–2012 metais – Vilniaus Gedimino technikos universiteto doktorantė.

Padėkos

Autorė nuoširdžiai dėkoja darbo vadovui prof. habil. dr. Jonui Skeivalui už suteiktus originalius mokslinius patarimus, pagalbą bei pastabas ir pasiūlymus rengiant disertaciją. Padėka visam Geodezijos ir kadastro katedros kolektyvui už palaikymą.

Jurgita MILIEŠKAITĖ

THE CLUSTERIC CORRELATION ANALYSIS OF LAND COVER DIGITAL RASTER IMAGES

Summary of Doctoral Dissertation
Technological Sciences, Measurement Engineering (10T)

Jurgita MILIEŠKAITĖ

ŽEMĖS DANGOS SKAITMENINIŲ RASTRINIŲ VAIZDŲ KLASTERINĖ KORELIACINĖ ANALIZĖ

Daktaro disertacijos santrauka
Technologijos mokslai, matavimų inžinerija (10T)

2012 11 16. 1,5 sp. l. Tiražas 70 egz.
Vilniaus Gedimino technikos universiteto
leidykla „Technika“,
Saulėtekio al. 11, LT-10223 Vilnius,
http://leidykla.vgtu.lt
Spausdino UAB „Ciklonas“,
J. Jasinskio g. 15, 01111 Vilnius