Methods of Automated Processing of Multispectral and Hyperspectral Data of Earth Remote Sensing for Solving Applied Problems of Natural and Anthropogenic Objects Assessment

Andrey V. Markov[1], Olga V. Grigorieva[1], Maxim O. Ivanets[1], Denis V. Zhukov[1]

[1] Mozhaisky Military Space Academy, St.Petersburg, alenka12003@mail.ru

Abstract. The paper gives approaches to the development of methodologies and programs for automated thematic processing of multi- and hyperspectral data of Earth remote sensing designed in Mozhaisky Military Space Academy. There are some examples of their practical use in classification of woody vegetation based on species composition, detection of burning areas and soil contamination by petroleum products.

Keywords: Earth remote sensing; multispectral and hyperspectral data; thematic processing; spectral features; assessment of the state of natural and anthropogenic objects.

1 Introduction

Analysis of domestic and foreign experience in Earth remote sensing (ERS) shows that at the present moment there is a steady interest in the development and implementation of new methods and means of image recording utilizing multi- and hyperspectral (MHS) aerial and space equipment as well. The use of these innovative technologies allows applying a new and more informative system of identification of objects features to thematic processing of aerospace imaging, which makes it possible to solve a number of socio-economic, scientific and defense tasks in automated mode.

At the same time, difficulties in large information flows analysis and existing limitations of standard remote sensing data processing programs of ERS (ENVI, Erdas Imagine, ScanEx Image Processor, etc.) impede practical implementation of MHS technologies. With all the variety of functions provided, these programs do not contain clear actions recommendations for an operator when solving specific thematic problems. The multiplicity of the proposed processing algorithms in some cases does not facilitate, but complicates a decryption process, since an operator often does not know which algorithm is the most effective for object recognition or its state identification.

These problems cause the urgency to create specialized methodologies and programs for MHS data processing adapted to applied tasks.

2 General approach to the development of thematic processing methodologies of MHS data of ERS

The technological development cycle of methodologies and programs for thematic processing of MHS data of ERS in a general case involves the following main steps:

- analysis of the optical properties of objects and media of wave propagation considered in the studied problem in order to identify patterns of signal formation recorded from objects of interest and related backgrounds in different parts of the electromagnetic spectrum;
- obtaining experimental measurements to study or refine (including testing theoretical hypotheses) spectral characteristics (SC) of objects and backgrounds;
- determination of informative spectral features of studied objects and backgrounds on the basis of their SC analysis;
- choosing the best classifying methods for MHS data (in terms of the best reliability of results and calculation speed);
- formation of a general methodology for thematic processing which unites all justified single methods and creation of a computer program that implements this methodology;
- verification of the designed methodology on independent data of aerial or space MHS shooting (not used in determining informative spectral features).

Spectral features may reflect both the differences in brightness of the object of interest and the background (or the differences in brightness of different classes of objects) in certain spectrum and the shape of spectral signature of the object of interest. In the first case, they are traditionally resulted either in the form of brightness contrasts of the object and its background in a fixed spectrum band or in the form of various indices (for example, vegetational) formed by mathematical relations between the brightness values in several informative spectral channels. In the second case, the features supply information about the local part of spectrum of the object, where the observed changes (extremes, kinks, etc.) are unique for it. The gradient method, the scale-space filtering, the sequential analysis, etc. are used for the calculation of form features.

Parametric (maximum likelihood, minimum distance, parallelepiped, spectral angle, etc.) and non-parametric (artificial neural networks, SVM, Fuzzy, etc.) image classification methods can be applied to MHS data analysis, as well as clustering algorithms and morphological segmentation (high-frequency filtering methods, spatial differentiation, watershed method). In some cases ensemble algorithms of various methods are developed in order to achieve the desired result. For example, for the classification of forests by species (see Figure 1), it is recommended to use an algorithm which takes parametric classification methods and creates a training sample based on the results of segmentation or uncontrolled clustering [2].

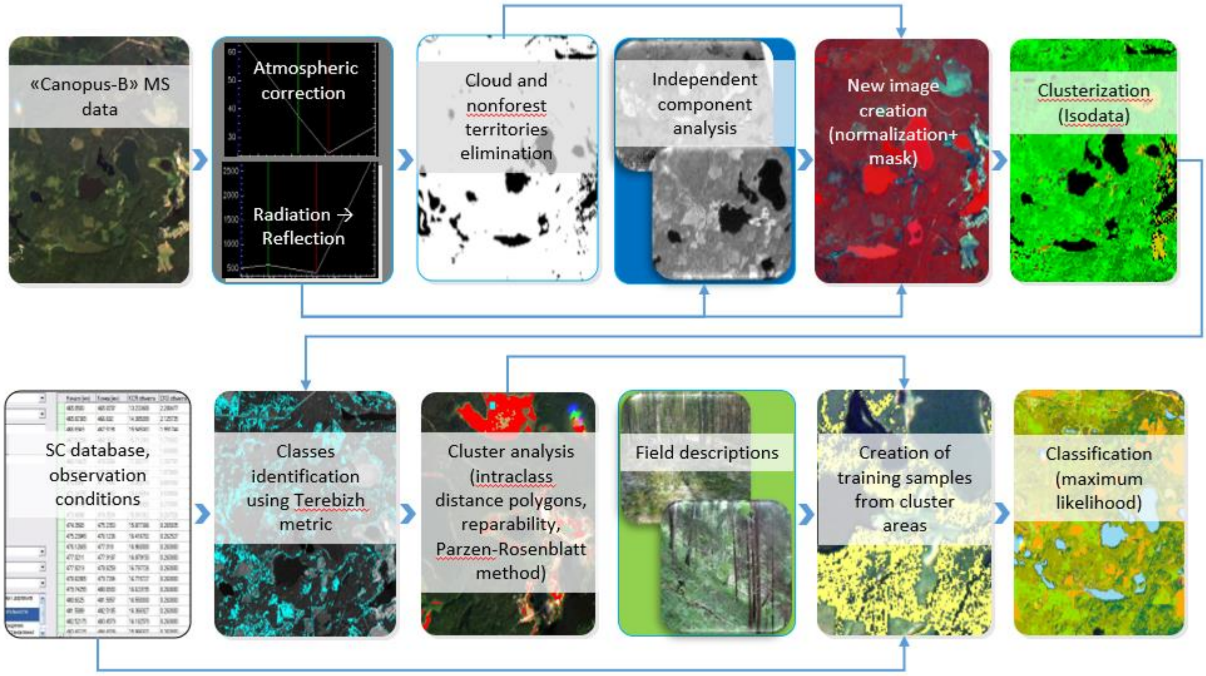


Figure 1. Ensemble algorithm of MS data processing for the classification of forest species.

To identify objects in hyperspectral (HS) images of medium and low spatial resolution there are algorithms of spectral separation of mixture. In this case, depending on the object of study, linear or non-linear models of formation of a mixed spectrum pixel are used. To determine the proportion (relative prevalence) of objects that make up a pixel the algorithms analyze spectral components (LSMA, linear spectral decomposition, etc.) taking into account the limitations of positivity and additivity of relative prevalence. At the same time, a set of reference spectra of objects can be obtained from SC specialized databases (DB) or initialized by the methods that determine spectral signatures composing images, such as: minimal volume simplex methods, dependent spectral components analysis, etc. [3]

In accordance with the described approach scientists of Mozhaisky Military Space Academy created a series of methodologies and computer programs for automated thematic processing of MHS data in order to solve applied problems concerning the state of natural and anthropogenic objects. The following represents the examples of work and briefly discusses methods underlying them.

3 Classification of woody vegetation by species

The developed program that determines the wood species codes during processing of MHS data, at its initial stage, separates forest and non-forest lands. It creates a forest mask with the help of the relaxation marking algorithm using the vegetation index NDVI as a pattern. This allows you to exclude the influence of sufficiently high values of metric indices in a feature range from further classification of the remaining fragments.

There are two types of treatment that are used for the tree classification.

The first algorithm works if some training samples are present. It applies the method of parametric classification of images, where recognition implements the Bayesian approach based on the discriminant functions. At the same time, HS data undergo preliminary processing aimed at optimizing spectral channels in order to minimize object identification errors using the maximum likelihood method. Optimization eliminates correlation between the channels by transforming the fragments of analyzed scene using the main component method. It allows determining the weight factor of each spectral channel (the elements of the own vector of covariance matrix), which indicates channel informativeness.

The second variant of classification of species composition involves the use of DB of the basic forest-forming species and their statistical estimates (coefficients of variation or standard deviations) incorporated in the program. Each spectral signature in the DB is accompanied by information about a projective cover (density), a phenological observation period (8 phenophases are provided), predominant species or composition of stand for mixed forests. Additionally, the forest type is characterized by common forest conditions (a complex of climatic, hydrological and soil factors) and bonitet (an indicator of growth and potential productivity of plantations for the given conditions of sprouting).

The processing algorithm in this case considers a sequential use of classification and creation of a training sample according to the results of clustering. Depending on the type of data methods like Isodata or fuzzy clustering serve as the methods of clustering and maximum likelihood estimation or neural networks as the methods of classification. Training samples are assigned by cluster. For this, the cluster compactness is estimated by analyzing the indices of intra-class distance and the number of local maxima of the density distribution function which is described with the multidimensional Parzen-Rosenblatt estimator. The sample size is determined by the spatial resolution of data and the number of spectral channels participating in classification (for example, for HS data of Resurs-P satellite with resolution of 30m the sample size must be at least 4×4 pixels with the number of spectral channels not more than 16) .

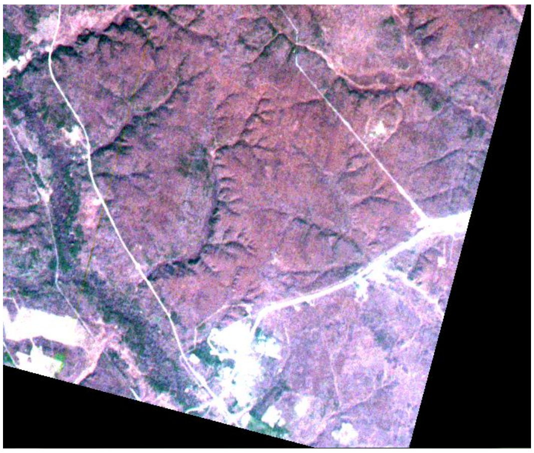
The correspondence between the class pixels and the species composition is established by the special identity criteria, such as: the credibility of spectra identification, special metrics (Mahalanobis and Teribizh) and a correlation factor. If there is a crown density and if the spectral reflectance coefficient (SRC) for an underlying surface is available, a sub-pixel method of mixture separation is additionally applied to the HS images or a reference signature is modeled. It is done so, since in this case several crowns of trees and a forest litter or an upper soil surface are included in a pixel [4].

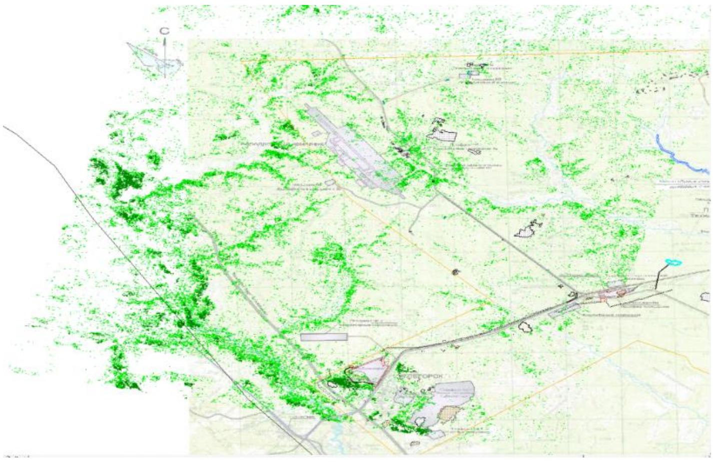
The frequency-contrast characteristics of the spatial brightness profile of a forest in an image are taken to determine a crown density. They are built using the direct Fourier transform and the calculation of the autocovariance function (ACF). The phase component of the Fourier transform contains important information about the mutual location of elements in an image and its spatial structure. It significantly differs for the areas with various densities and crown diameters due to the presence of shadows in the inter-crown space and illuminated areas of grass and shrub vegetation in light forests. ACF indicators (function maximum and frequency range at which ACF becomes less than a zero) effectively reflect the location and size of structural elements.

To each tree species, identified as the result of classification, is assigned a code (a letter symbol, for example, B - birch). The formula of species composition is recorded in the attribute data of electronic layer of the unit. For example, for the forest with 40% of pine and 60% of birch birch is recognized as a predominant species and the composition formula makes 4P6B.

In accordance with the established species composition of the units, polygons of an electronic layer of the map are painted using the color scale of forest strata given in the “Guidelines for the State Forest Inventory” approved by the order of Rosleskhoz of November 10, 2011 No. 472.

An example of the program work that implements the described algorithm is shown in Fig.2





a) initial HS image

b) processing result

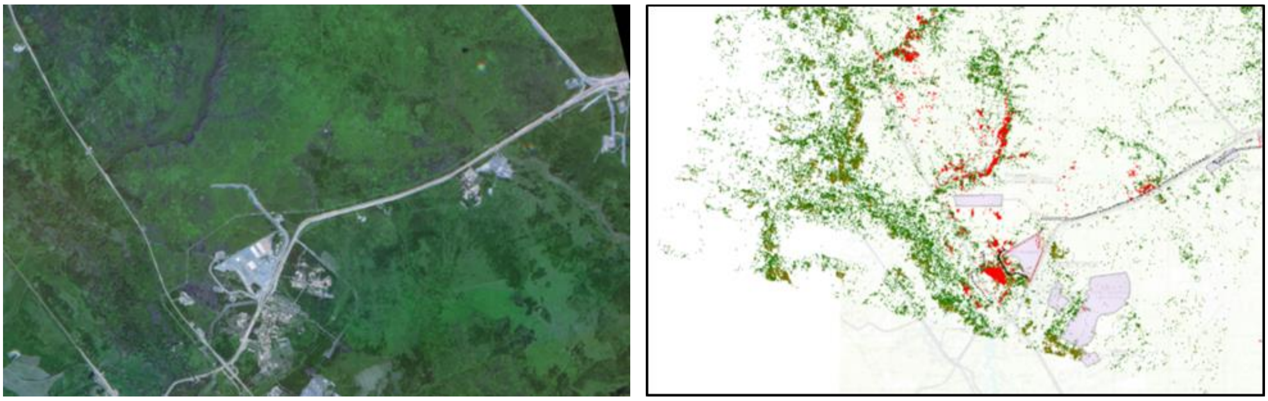
Figure 2. HS data –based identification of forest species composition (coniferous and mixed forests).

4 Burnt area detection

Detection of burnt areas in processing of MHS data is based on operations with vegetation indices. Their composition depends on the spectral range of MS or HS images. In the presence of data in the spectral bands of

1.55-1.75 or 2.10 2.35 μm it is sufficient to calculate only two indices - NDVI and SWVI [5]. If the right border of the spectral range does not exceed 1.0 μm, three vegetation indices are needed: NDVI, PSRI, TCHVI. They reflect changes in the physiological state of vegetation in the form of a decrease in chlorophyll concentration [5]. For HS data it is also necessary to take into account the indices characterizing the moisture content in the vegetation cover (for example, WBI). In addition, the detection algorithm itself is getting complicated. It is a decision tree which includes threshold processing for specified vegetation indices and values of the SRC in spectral ranges of 540-600; 620-680; 720-800 and 800-900 nm. The use of these features allows eliminating the fragments (cuttings, swamps, peat lands, light forests) that have spectral characteristics similar to burnings and increase the accuracy of classification.

An example of the proposed methodology is shown in Figure 3.



a) initial MS image

b) processing result

Figure 3. MS data – based identification of burnt areas.

5 Detection of soil contamination by petroleum products

The methodology of detection of soil contamination by petroleum products (PP) consists of the following steps:

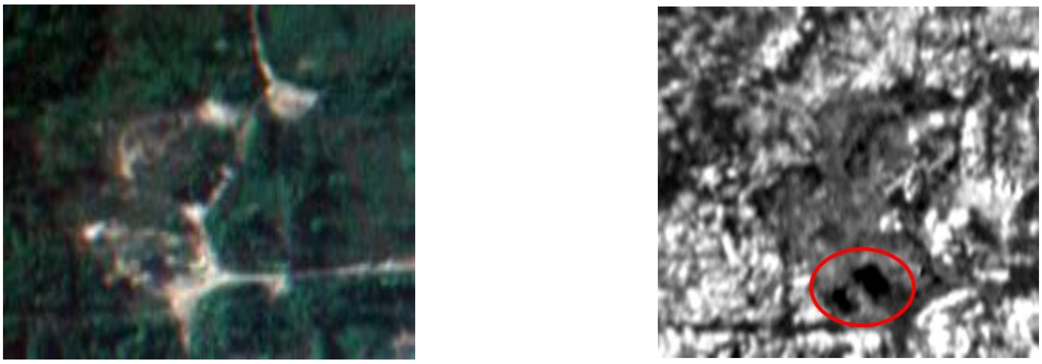
1) Uncontrolled clustering of three NDVI index images obtained by the fuzzy set method. The Euclidean distance that divides the analyzed area into three clusters is taken as a cluster separation norm. NDVI index images are calculated in three spectral intervals: 700-740 nm, 720-800 nm, and 800-900 nm.

2) Clusters corresponding to the vegetation cover are excluded from the zone of further search considering the values of vegetation index.

3) For the remaining clusters, Isodata re-clustering is performed on all initial MHS data and the clusters that correspond in their spectral characteristics to petroleum polluted soils are selected. The following measures of convergence of spectral signatures are used as the criteria for assigning a cluster to a site of petroleum pollution: the Euclidean distance which takes into account the difference in average values of the SRC of clusters in a spectral channel and a correlation coefficient of the reference spectra of PP and the analyzed spectra of clusters.

4) Identification of the selected sites of petroleum pollution according to their type (bituminized kir, liquid phase PP, oil film and water-oil emulsion) and measurement of the soil pollution level (complete destruction, high, medium, low) based on the reference spectral characteristics of ground and petroleum products. The level of soil pollution is determined by the index of petroleum contamination calculated in the near infrared and blue spectral channels.

Figure 4 demonstrates an example of detection of petroleum contamination by the proposed methodology.



a) initial HS image synthesized in natural colors 517, 569 and 677 nm

b) processing result

Figure 4. HS data –based detection of petroleum contaminated soil.

6 Conclusion

Practical implementation of the automated methods of processing of MHS data of ERS presented in this article demonstrates their high potential of use in the fields related to environment and information on SC of natural and anthropogenic objects. At present, the range of thematic tasks which can be solved by using MHS data is constantly expanding.

In particular, the results of studies of Mozhaisky Military Space Academy show that it is possible to use them effectively for identifying contaminants of water bodies [7], determining the depths of coastal waters [8], assessing the trafficability of terrain [9], analyzing vegetation [10], and others.

References

[1] Grigorieva O.V., Markov A.V., Ivanez M.O., Zhukov D.V. Methods of preparation of formal etalon features for target detection using hyperspectral remote sensing data // Vth All-Russian Scientific and Technical Conference with International Participation FOCAL PROBLEMS OF SPACE-ROCKET HARDWARE (Vth Kozlov Readings). Samara: JSC SRC Progress, 2017. Vol. 1. P. 281-286.

[2] Grigoreva O.V., Mochalov V.F., Rogachev S.A. Ecological Monitoring of Specially Protected Natural Territories on the basis of Imagery Taken by Domestic Satellites. // II All-Russian K.J. Kondratiev Scientific Conference ECOLOGY AND SPACE. SPb.: Mozhaisky MAA, 2015. P. 104–109.

[3] Grigoreva O.V. Subpixel identification of objects by multi- and hyperspectral data applying sequential quadratic programming and a method of spectral components analyses // Digital Signal Processing, 2018. No. 3. P. 26-32.

[4] Markov A.V., Grigoreva O.V., Chapurskiy L.I., Mochalov V.F. Methods of Preparation of Reference Information for Processing of Multi- and Hyperspectral Aerial Imagery Aimed at Environmental Monitoring. // All-Russian Scientific Conference "Problems of Military-Applied Geophysics and Environmental Control" SPb.: Mozhaisky MAA, 2014. P. 525-534.

[5] http://www.indexdatabase.de

[6] Grigoreva O.V., Saidov A.G., Panin A.V. Soil State Indicator in the Areas of Circulation of Oil and Petroleum Products Obtained by Non-contact Means of Observation. // Ecology and Industry of Russia, 2010. No. 10. P. 2-5.

[7] Grigorieva O.V., Zhukov D.V., Markov A.V., Mochalov V.F. The estimation technique of ecological-economic damage to water objects according to remote sensing data // Current problems in remote sensing of the Earth from space, 2018, Vol. 15, № 1. pp. 233-244.

[8] Grigorieva O.V., Zhukov D.V., Markov A.V. and Mochalov V. F. The Retrieval of the Coastal Water Depths from Data of Multi- and Hyperspectral Remote Sensing Imagery. // Atmospheric and Oceanic Optics, 2017, V. 30. No. 01. pp. 7–12.

[9] Saidov A.G., Grigoreva O.V., Markov A.V. The Use of Geospatial Data in Assessing Terrain Trafficability of Border Areas. // Materials of I interdepartmental Conference "Special geo-information technologies and aerospace sensing of the Earth." Kaliningrad: Kaliningrad Border Institute of the Federal Security Service of Russia, 2016. C. 33-36.

[10] Grigoryeva O. V. Observation of Forest Degradation Using Hyperspectral Data Aerial and Satellite Sensing // Earth Observation and Remote Sensing. No 1. 2014. P. 43–48.