Automation of aerial and satellite image interpretation*

A. N. Solovyova[1], A. V. Kuchuganov[2]

Department of Automated Data Processing and Control Systems
Kalashnikov Izhevsk State Technical University
Izhevsk, Russian Federation
[1]alexandra-solovyova@yandex.ru, [2]aleks_kav@udm.ru

Abstract—**The paper describes a technique for the analysis and description of images and its application to the automated interpretation of aerial and satellite images. The image description is formed at the three levels of abstraction (simple, textural and significant objects). At each of the levels the quantitative and qualitative features of objects are determined using algorithms based on analytical geometry and fuzzy logics. The paper describes a software system for image semantic codification which implements the suggested technique. The results of the experiments on object detection are provided.**

Keywords—aerial and satellite imagery, semantic network, fuzzy attributed graph, fuzzy logics, texture analysis, automated interpretation of images

I. INTRODUCTION

One of the means for improving the performance of technical systems in various fields of activity is the automation of the processes of image analysis and the development of image-based decision support systems.

The main problem in the way of obtaining a high-quality description of the structure of the displayed objects is the semantic gap between the levels of information used by computers and humans during the image interpretation. The computer accesses the low-level features of the image such as pixels, color, texture and shape and the human for the most part uses high-level concepts such as object, feature, action.

The methods of image analysis widely used at the moment are based on low-level (raster points) and medium-level (geometric features) invariant characteristics: object features formed by distinguishing combinations of color and tone, more rarely – shape attributes of objects and geometric relations between them. These methods do not involve "actual" high-level image understanding. Invariant features of the image located by machine learning methods can be unpredictable and unobvious to an user, though fully proven in terms of the low-level and medium-level of analysis.

The need to synthesize a description of an image without human intervention is relevant for aerial and satellite image interpretation. Studying and mapping of a territory from its images implies detecting the objects and terrain features, determining their quantitative and qualitative characteristics and marking them on the map. The interpretation process is based on the use of specified relations between the properties of the objects which are divided into direct indicators and indirect indicators. The direct indicators are the properties perceived by the eye (size, shape, shadow, tone, color, structure). The indirect indicators define an object through the properties of the others (including their situation) and are used when the direct indicators are not sufficient or when detecting underground or seasonal objects [1].

When digitizing an aerial or a satellite image a specialist performs most of the operations himself using vectorization tools. Thus, the resulting image interpretation is largely defined by his subjective decisions and skills.

Currently in GIS systems (GRASS GIS, Spatial Analyst tools for ArcGIS) the tools for automated image interpretation are represented by recognition of certain categories of objects and supervised and unsupervised classification of image regions (based only on direct indicators). The obtained results are revised and corrected manually. As opposed to pixel-level image processing, GEOBIA (Geographic Object-Based Image Analysis) approach [2] performs region-level processing: classification and specifying production rules for classes of objects (eCognition, Feature Analyst, ENVI EX Feature Extraction module etc.). Such an approach makes it possible to process high resolution images in less detail and reduce classification errors caused by noise. A negative tendency observed for the GEOBIA tools consists in oversimplifying the user interface due to hiding the mechanisms of object detection [3].

Providing a full and correct interpretation of an image involves referring to a certain ontology of high-level concepts. Besides, it should remain possible to access the low-level elements that make up the structure of the displayed objects and relations between them. To systemize them the model of hyerarchically structured multi-level image description is developed, which allows to describe the image at the low, medium and high levels in terms of logics and natural language [4].

II. A MULTI-LEVEL IMAGE MODEL

In the task of aerial and satellite image interpretation we suggest the following levels of image description: 0) *raster level*; 1) *"simple objects"* – separate uniform regions of the vectorized image; 2) *"textural objects"* – composite objects formed by groups of uniform regions with similar features; 3) *"significant objects"* – objects relevant for the domain (e.g. roads, forests, bridges); 4) *general description of the image* (e.g. the ecological state of the area).

The raster level information (level 0) is not explicitly used in the resulting image description, as the human does not think in terms of a raster but uses geometric information. The description of an image at level 4 can be obtained through querying the significant object level without referring to the geometric data directly.

Concerning the transformation of the image features from the low level to the high level, levels 1–3 are of the main interest. The indicators used for the image interpretation correspond to different abstraction levels: size, tone and color can be formalized at the simple object level, structure and pattern – at the textural object level, falling shadows and most of the indirect indicators imply formalization at the significant objet level.

The model M of an image consists of several levels:

$$M = (Raster, Vector, Description),$$

where $Raster$ is the original raster image, $Vector$ is the vectorized image [5],

$Description$ is the image description at the three levels of the abstraction of qualitative and quantitative features represented in the form of a semantic network:

$$Description = ((PDescr, TDescr, SDescr), Links),$$

where $PDescr$, $TDescr$, $SDescr$ are the three layers of the image description corresponding to the levels of simple objects (level 1), textural objects (level 2) and significant objects (level 3) respectively;

$Links$ are the inter-level "Part-of" relations: between simple objects and textural objects formed from them; between simple or textural objects and significant objects that they represent (Fig. 1).

Each layer of the $Description$ semantic network is represented by a fuzzy attributed graph (V, R, A, L), where
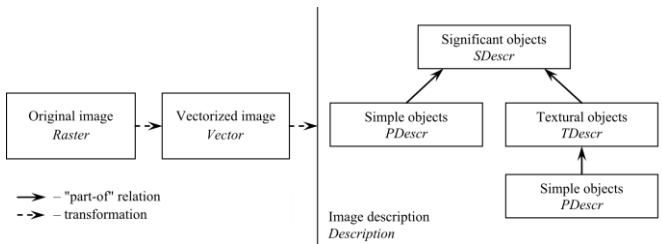

Fig. 1.   The components of the image model

vertices V correspond to the objects detected at the image, arcs R – to the relations between the objects: nesting, adjacency, for significant objects – "Is-a"; an arc can also be added at the request of the user. Each vertex and each arc of the graph are characterized by a subset of attributes from the set A. The attributes corresponding to direct and indirect indicators are assigned linguistic variables [6] of the set L.

III.   A TECHNIQUE FOR THE AUTOMATED INTERPRETATION OF AERIAL AND SATELLITE IMAGES

Within the discussed image model M we suggest using the following technique for the interpretation of aerial and satellite images:

1. Vectorize the raster image.

2. Construct the simple object layer $PDescr$ in the semantic network of the image.

    2.1. Construct the network of vertices and arcs of the $PDescr$ graph.

    2.2. Calculate and fuzzify the values of the indicators of simple objects and relations between them. Fill the attributes of the $PDescr$ vertices and edges with the obtained values.

3. Construct the textural object layer $TDescr$ in the semantic network of the image.

    3.1. Select and outline the textural objects.

    3.2. Construct the network of vertices and arcs of the $TDescr$ graph.

    3.3. Calculate and fuzzify the values of the indicators of textural objects and relations between them. Fill the attributes of the $TDescr$ vertices and edges with the obtained values.

4. Formalize the significant object detection rules in terms of the indicators of simple and textural objects. Define the membership functions for the verbal values of the qualitative features.

5. Construct the significant object layer $SDescr$ in the semantic network of the image.

    5.1. Select and outline the significant objects.

    5.2. Construct the network of vertices and arcs of the $SDescr$ graph.

    5.3. Calculate and fuzzify the values of the indicators of significant objects and relations between them. Fill the attributes of the $SDescr$ vertices and edges with the obtained values.
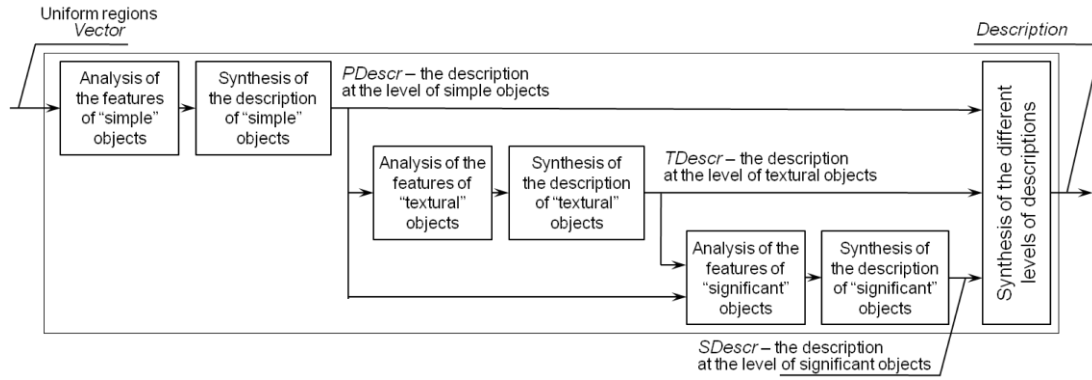
Fig. 2. The scheme of the image description process

Therefore, the image description is formed from the low level to the high level (Fig. 2). The result of the image interpretation is a set of thematic layers each of which corresponds to a certain significant object and is filled with simple and textural objects of the image.

Let us consider the image description at each of the levels.

A. Analysis and description of simple objects

In this stage for each of the uniform regions of the image the following quantitative and qualitative features corresponding to the direct indicators are determined: 1) *size*: the area of the region in absolute units or relative to other objects; 2) *shape*: elongation, tortuosity of the boundary, orientation, rectilinearity, roundness, density of boundary points, geometrical regularity of the boundary, proportion of holes and branches; 3) *tone*; 4) *color*: hue and saturation; 5) *contrast*: tonal and color.

The following relations between simple objects are assessed: 1) *situation*: relative horisontal and vertical position, distance, symmetry etc.; 2) *comparison*: relative intensity of features.

B. Analysis and description of textural objects

In the vectorized image the texture is formed by a (sufficient) number of uniform regions with similar features located within a certain boundary. The uniform regions with similar features can be treated as a cluster in the indicator space. In order to separate the similar objects into geometrically compact groups the objects within each indicator space cluster should be additionaly clustered by their geometric centers. Then the obtained geometric clusters can also be grouped into complex clusters by their geometric centers, which allows to model more complex textures.

The heuristic algorithms are applied to each of the textural objects to determine its quantitative and qualitative features corresponding to the direct indicators: 1) *structure*: the average, maximum and minimum values of the features of simple objects that form the textural one; 2) *shape*: for the boundary of the textural object the shape features are determined; 3) *texture*: coarseness, contrast, directionality, linearity, regularity, roughness. Also the relative position and the relative intensity of the features are assessed for the pairs of textural objects.

As each textural region is formed by a number of vector objects, the task of the description of its generalized shape is raised. Modeling the boundary of a textural object can be treated as the construction of a "concave" hull of a set of points made up of the boundaries or geometric centers of vector objects [7]. The shape of the obtained polygon is analyzed similarly to the shape of simple objects.

C. Analysis and description of significant objects

In this stage the objects relevant for the domain are detected and described, such as areal objects (bodies of water, gardens etc.), linear objects (roads, fences etc.) and point objects (wells, transmission towers etc.).
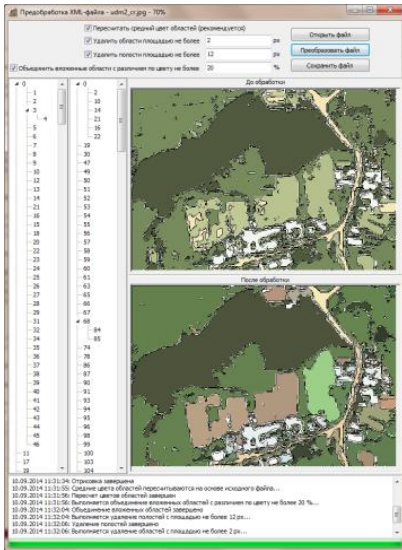
To detect the significant objects one should analyze the image description at the simple object and textural object levels. To implement the detection of the user-requested types of significant objects we suggest searching at the two lower levels of the semantic network $PDescr \cup TDescr$ for the subgraphs which attributes meet the constraints posed by the query graph. The query graph is used to locate a required combination of attribute values in the semantic net, so to draw a conclusion that the image contains significant objects of the specified type.

At the significant object level the relations of *adjacency* and *nesting* are assessed, as well as the *part-of* relations (e.g. between the "Tree" and "Forest" objects, "Building" and "Group of buildings" objects) and *is-a* relations (e.g. between "Unpaved road" and "Road" objects).
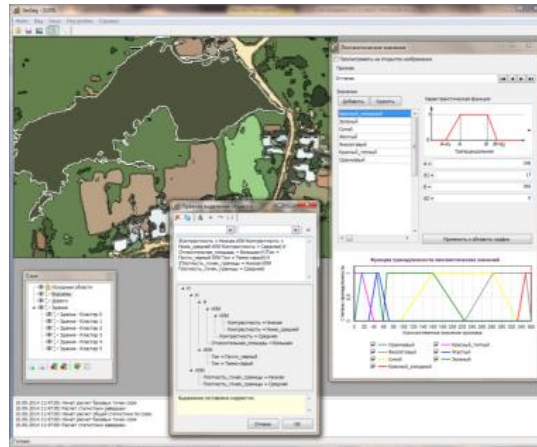
The list of indicators used at the significant object level is defined by the type of the object. For example, the attributes of a "Forest" object can include *Tree species*, *Average height*, *Average distance between the trees*, the objects of a "Road" object – *Roadway width*, *Paving material*.

IV. A SYSTEM FOR IMAGE SEMANTIC CODIFICATION

The proposed formalized image model and the technique for automated image interpretation have been implemented in the semantic codification software system TexSeg. The system makes it possible to considerably reduce the amount of manual operations when detecting the significant objects and terrain features, analyzing their characteristics and digitizing their boundaries to mark them on the map (Fig. 3).

Fig. 3. The interface of the semantic codification system TexSeg: (a) the module of vectorized image preprocessing; (b) extracting the bodies of water from the satellite image (the layers window, the production rule interpreter, the verbal values editor are displayed)

Fig. 4. The verbal value editor window

The system database stores the geometric data of the image, membership function parameters for the verbal values of the indicators and the significant object detection rules. The quantitative values of the indicators are calculated by means of SQL select queries.

The user sets the membership functions for the verbal values of the indicators using the editor interface represented at Fig. 4. The system provides the highlighting of the regions that possess the specified verbal value of an indicator.

The user specifies the constraints on the simple and textural object indicators by means of the embedded interpreter in the form of production rules expressed by verbal statements. Each rule is represented by a logical combination of *Feature = Value* assertions according to the following grammar:

<Object> ::= Field | Forest | River | Building | ...

<Feature> ::= Holes_count | Elongation | Tortuosity | Tone ...

<Value> ::= Continuous_area | Many_holes | White | Near_to_white | Gray | ...

<Premise> ::= <Feature> = <Value>

<Premise> ::= NOT (<Premise>)

<Premise> ::= (<Premise>) AND (<Premise>)

<Premise>) ::= (<Premise>) OR (<Premise>)

<Rule> ::= IF <Premise> THEN <Object>

For example, the rule for the "Road" objects detection can be of the following form:

IF (Elongation = Very_high) OR (Elongation = High)
AND NOT Branch_count = No_skeleton
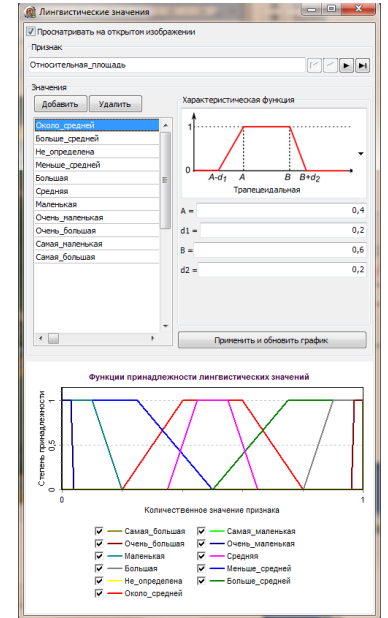*AND Point_density = Low*
AND Tonal_contrast = Average

AND (Hole_count = Continuous_area
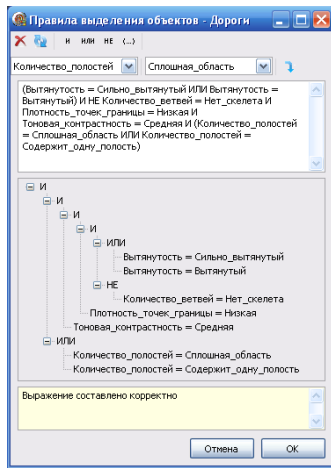*OR Hole_count = Has_one_hole) THEN Road*

The production rule interpreter window and the examples of significant object detection are shown in Fig. 5.

If the user is planning to further analyze only a certain separate group of the significant objects, the system provides the tools for dividing the thematic cluster into geometric clusters. The clustering can be based on the *geometric centers* and/or the *skeleton endpoints* of the objects. For example, Fig. 6 is showing the result of the *Trees* thematic layer clustering – the obtained geometric clusters corresponding to the separate groups of the trees.
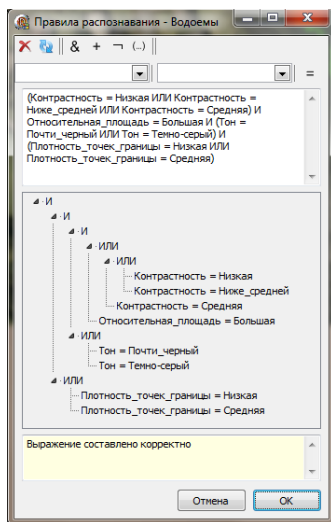
The operability of the system has been tested on 150 Landsat-7 images and 50 Quickbird images obtained by means of Google Earth program.

V. EXPERIMENTS

The accuracy and completeness of the significant object detection at the aerial and satellite images have been assessed. During the experiment the results of the automated interpretation in the developed system have been compared to the results of the manual interpretation performed by experts for 20 images. The *accuracy* value is defined as the ratio of the count of the objects correctly detected by the system to the total count of the objects detected by the system. The *completeness* value is defined as the ratio of the count of the objects correctly detected by the system to the count of the objects detected by the expert. According to the experiment results, the accuracy and completeness of the significant object detection amount to 84% and 89% respectively (Table I). The accuracy of the significant object outlining is defined as a ratio of the count of the objects which contours don't reqiure editing to the total count of the detected significant objects and averages to 96%.

(a)

(b)

Fig. 5. Specifying the rules for the significant object detection in the interpreter window: (a) extracting the roads; (b) extracting the bodies of water

Fig. 6. Clustering of a thematic layer

TABLE I. ASSESSING THE ACCURACY AND COMPLETENESS OF SIGNIFICANT OBJECT DETECTION WITH THE SEMANTIC CODIFICATION SYSTEM

Significant objects	Accuracy	Completeness
A set of the satellite images of the Udmurt Republic (10 items)		
Roads	86%	94%
Buildings	83%	85%
Bodies of water	84%	90%
Fields	83%	86%
Artificial objects	85%	88%
A set of the satellite images of the Republic of Sakha (Yakutia) (10 items)		
Rivers	84%	93%
Alas lakes	87%	95%
Taiga	84%	90%
Burned-out forests	79%	87%
Meadows	82%	86%
Residential areas	85%	87%
Average	*84%*	*89%*

The reliability of the obtained image descriptions is proven by their matching to the manual interpretation results as well as by using reference data on indicator values to specify the significant object detection rules.

VI. Conclusion

The proposed image model and the technique for the automated interpretation of aerial and satellite images makes it possible to use in addition to the low-level features (pixels, color, texture, shape) the quantitative and qualitative features of separate uniform regions (simple objects), textural objects and spatial relations between them, as well as features relevant for the domain. The technique does not require the considerable amount of training examples.

Unlike the current approaches to the aerial and satellite image description, the image model represented in the form of a multi-level semantic network makes it possible to interpret the image in terms of each of the levels separately and all of the levels as a whole, as well as to analyze both the immediate properties of the objects and the spatial relations between them. As the queries to the semantic network are executed, a hyerarchically structured description is formed.

REFERENCES

[1] "GKINP-02-121-79. The manual for aerial image interpretation for plane surveying and updating the 1:2000 and 1:5000 surface plans (Geodesic and cartographic instructions, norms and regulations)", 1980. (In Russian).

[2] T. Blaschke, S. Lang, G.J. Hay (Eds.), "Object-Based Image Analysis: Spatial Concepts for Knowledge-Driven Remote Sensing Applications (Lecture Notes in Geoinformation and Cartography)," 2008.

[3] P. Aplin. (2014, May 24). "GEOBIA state of the art, science or technology" [Online]. Available: http://geobia2014.web.auth.gr/geobia14/sites/default/files/pictures/aplin.pdf

[4] A. V. Kuchuganov, A. N. Solovyova, "Multilevel description of images as applied to the interpretation of aerospace images," in Intellektual'nye sistemy v proizvodstve (The Intelligent Systems in Manufacturing), vol. 2(24), 2014, pp. 164-166. (In Russian).

[5] A. V. Kuchuganov, "Recursions in tasks of images analysis," Pattern Recognition and Image Analysis: New Information Technologies (PRIA-9-2008), vol. 1, pp. 365-367.

[6] L. A. Zadeh, "The Concept of a Linguistic Variable and its Application to Approximate Reasoning," in Information Science, vol. 8, 1975, pp. 199-249.

[7] A. N. Solovyova, "Construction of polygons representing the boundaries of textural regions of an aerospace image," in Intellektual'nye sistemy v proizvodstve (The Intelligent Systems in Manufacturing), vol. 2(24), 2014, pp. 167-168. (In Russian)