

Krasnoyarsk : Institute of forest named after V. N. Sukhachev, Russian Academy of sciences, 2007.

6. Holmgren J., Persson A. Identifying species of individual trees using airborne laser scanner // Rem. Sens. Environ. 2004. Vol. 90. № 4. P. 415–423.

7. Laser scanning of forest resources: the Nordic experience / E. Nasset [et al.] // Scand. J. For. Res. 2004. Vol. 19. № 6. P. 482–499.

8. Soille P. Morphological Image Analysis: Principles and Applications. 2nd ed. Berlin : Springer-Verlag, 2003.

© Danilin I. M., Danilin A. I., Svisehev D. I., 2010

M. N. Favorskaya, A. M. Popov
Siberian State Aerospace University named after academician M. F. Reshetnev, Russia, Krasnoyarsk

CONTEXT-DEPENDENT GRAMMAR DESIGN FOR DESCRIBING COMPLEX SCENES WITH MULTI-LEVEL OBJECT MOTION

The problems of context-dependent grammars formation which describes structural information about patterns and pattern interaction in complex scenes are discussed in this article. The application of three-level grammar based on the task of an image sequences syntactic analysis (with extended contents of main and auxiliary dictionaries) and the task of scene syntactic analysis with multi-level object motion is suggested.

Keywords: context-dependent grammar, syntactic analysis, multi-level motion.

Initially, the structural or linguistic approach had been based upon using different linguistic structures, consisting of a dictionary and rules of sentence building from a specified dictionary. Such structural description of images permits to make an analogy between image structure and language syntax of formal grammars. Notice that this line of development appeared in the 1960's as one of the first approaches for image describing and recognition. Structural approaches not only permit the reference of supervise static objects to a definite pattern, but also describe some object properties that exclude its referring to another pattern.

Traditional methods of structural approach are based on syntax description of complex image sets with limited sets of primitives and grammatical rules. It is considered that these images are formed from elements which are connected in a variety of ways just as phrases and sentences of languages are built by connecting words, and the words are composed from letters. The simplest elements, from which words and then sentences are built, are called primitives. Designing rules of composing primitives are usually assigned with special grammars of images description. Grammar rules (rules of substitution) may be applied any number of times, which allows for a compact and sufficient definition of primary structural characteristics for a sentences infinite set. The language for image structural description in terms of primitives sets and designs such element compositions; it is called the image description language. During identification, the recognition of primitives and image description in terms of special language is realized. Essentially, pattern recognition consists of a syntax analysis (or grammar analysis) and a "sentence" which describes some image. Recognition maintains the syntax correspondence between the analyzable "sentence" or image description and special grammar [1].

The system of pattern syntax recognition includes three main modules: the pre-processing module, the description module, and the syntax analysis module. The pre-processing module realizes coding, approximation, filtration, reconstruction, and the improvement of the image. The description module includes the primitives' segmentation and allocation based on predetermined syntax operations. Each allocated part of the image is identified relatively to a special primitives set, and the whole image is characterized by a set of primitives' sequences as the structures of language types. The syntax analysis module checks the accuracy of the sets in the context of predetermined grammars. The predetermined grammar corresponds to each pattern, and if the description of analyzable image is syntactically correct in the context of such grammar, then the image is related to the pattern for which this grammar corresponds.

The development of grammar describing both structural pattern information and patterns interaction is connected to the necessity of designing grammar reconstruction (or conclusion) algorithms according to a defined set of dynamic images that present the learning sampling. Such algorithms accomplish the learning of the recognition system. In result, pattern structural descriptions and their relationship descriptions are formed; then they are used for a syntax analysis of events and the genre of a complex scene. Basically the learning process isn't executed; the choice of grammar and primitives set are realized by a tutor. Since the dynamic scene with multi-level motion has a very complicated and time-dependent structure, it's necessary to use context grammar rules, which form a multi-level context grammar.

Let's consider some main regulations which are peculiar to the structural methods of scene describing or recognition. The generative grammar is a well-ordered set of parameters $\mathbf{GR} = (V_T, V_N, P, S)$, where V_T – is a finite

alphabet, defining the set of terminal symbols; V_N – is an alphabet, defining the set of nonterminal symbols; P – is a finite set of conclusion rules, i. e. a set of following pairs $u \rightarrow v$, where $u, v \in (V_T \cup V_N)^*$; S – is an initial symbol (grammar axiom), $S \in V_N$. The sequences of grammar generative language consist of terminal symbols. The symbol in left part of the first rule of grammar conclusion is an axiom. In grammar \mathbf{GR} the sequence x directly generates the sequence y , if $x = \alpha u \beta$, $y = \alpha v \beta$, and $u \rightarrow v \in P$, i. e. the sequence y directly concludes from the sequence x , that denotes $x \Rightarrow y$. The language which is generated from grammar $\mathbf{GR} = (V_T, V_N, P, S)$ is called the set of terminal sequences, concluding in grammar \mathbf{GR} from axiom: $L(\mathbf{GR}) = \{x \mid x \in V_T^*; S \Rightarrow *x\}$, where symbol \Rightarrow^* – is deducibility.

Rules generated by grammars permit string transformations. Constraints on the rule type determine grammar classes. The classification which was proposed by N. Kholmskiy, defines four grammar types:

- grammars type 0 – grammars, which don't have any constraints on the conclusion rules;
- grammars type 1 (context grammars) – grammars, the rules of which have the following view: $xAy \rightarrow x\varphi y$, where $A \in V_N, x, y, \varphi \in (V_N \cup V_T)^+$;
- grammars type 2 (context-free grammars – CF-grammars). Conclusion rules in these grammars have the following view: $A \rightarrow \varphi$, where $A \in V_N, \varphi \in (V_N \cup V_T)^*$;
- grammars type 3 – finite state grammars which are divided into two types:
 - a) left linear (left recursive) grammars, conclusion rules for which have the following view: $A \rightarrow Aa \mid a$, где $A \in V_N$;
 - b) right linear (right recursive) grammars, conclusion rules for which have the following view: $A \rightarrow Aa \mid a$.

Language L is called i type language if the grammar of type i exists and generates language L . A conclusion tree often called the tree of grammar analysis or syntax tree, and the building process of the conclusion tree called – the grammatical analysis (syntax analysis). For one language sequence more than one tree can correspond, because this sequence can have different conclusions which are generated by various trees. For example, CF-grammar $\mathbf{GR} = (V_T, V_N, P, S)$ is called an ambiguous (indefinite) grammar, if the sequence $x \in L(\mathbf{GR})$ exists, and has two or more conclusion trees. It should be remembered that the tree of syntax analysis isn't grammar in the form of graph. Grammar graphs contain sentential forms (any sequences which are generated from an axiom) as nodes.

The principal disadvantages of the mentioned grammars are connected to their suitability in a greater extent for description scenes than for their recognition. This disadvantage has been avoided due to investigations which had been carried out under M. I. Shlezinger who used two-dimensional programming method. The suggested two-dimensional grammar \mathbf{GR}_S in [2] is a function of six parameters:

$$\mathbf{GR}_S = \langle V_O, \mathbf{S}, \mathbf{T}_V, \mathbf{T}_S, R, \{Z, Z(t, t'); (t, t') \in \mathbf{N}\} \rangle.$$

Let object recognizing images be situated in images \mathbf{T}_V . Each of these images is a function given in images \mathbf{T}_V , possessing values of the object alphabet V_O , which corresponds to the primary alphabet in one-dimensional grammars. The elementary images of this alphabet are used for the composition of more complex images. Besides the signal alphabet there is the alphabet of \mathbf{S} structural elements which corresponds to auxiliary alphabets in one-dimensional grammars. On one hand, the structural elements define possible values of corresponding signals. On the other hand, they maintain constraints on image structures as local constraints. Structural elements make up an image description which is defined as function \bar{S} on finite set \mathbf{T}_S (description) and possessing values from set \mathbf{S} . Generally these descriptions are not clear isomorphic images. An element of sets $Z = V_O \cup \mathbf{S}$ called “a symbol”, and denoted as z . The set \mathbf{T} is a combination of image and description. This element is called “a cell”, and is denoted as t . Two cells t and t' are regarded as adjacent cells if some fixed for this grammar symmetric predicate $R(t, t')$ is equal “1”. At that, \mathbf{N} is the set of adjacent cells.

A pair (\bar{V}, \bar{S}) image – description is called a variant \bar{Z} .

This means that the variant is a function, specified on a set $\mathbf{T} = \mathbf{T}_V \cup \mathbf{T}_S$ and assumed values from set Z , such that $Z(t) \in V_O$, if $t \in \mathbf{T}_V$, and $Z(t) \in \mathbf{S}$, if $t \in \mathbf{T}_S$. Sets $Z, Z(t, t')$ of allowable pairs (Z, Z') of symbols $Z, Z' \in Z$ are defined for each pair of adjacent cells t and t' . Variant \bar{Z} is called an allowable variant, if for each pair $(t, t') \in \mathbf{N}$ ratio $(Z(t), Z(t')) \in Z, Z(t, t')$ is executed. An image \bar{V}^* is called an allowable image, if the allowable variant $\bar{Z}(\bar{V}^*, \mathbf{S})$ exists. If variant $\bar{Z}(\bar{V}^*, \bar{S})$ is an allowable variant, then description \bar{S} is called a possible description of image V^* .

Let the assign with two-dimensional grammar \mathbf{GR}_S not be a whole set of images \mathbf{X}^* , which are concerned to one visual pattern, but if it's a small part, it is called a set $V^*(\mathbf{GR}_S)$ of ideal or pattern images. Any pattern image $\bar{V} \in V^*(\mathbf{GR}_S)$ corresponds to some set of real images which are similar to pattern image \bar{V} . The membership function $f_V(\mathbf{X})$ of recognizing image to set $V^*(\mathbf{GR}_S)$, is called a similarity, possessing variable values (not only equal “0” or “1”). The syntax analysis task of image \mathbf{X} is in the pattern definition of the image, which is generated by grammar \mathbf{GR}_S and maximized by a similar function:

$$V^*(\mathbf{X}) = \arg \max_{\bar{V} \in V^*(\mathbf{GR}_S)} f_V(\mathbf{X}). \quad (1)$$

In paper [2] a solution of this task was proposed: it was called a method of two-dimensional programming. This method permits to simultaneously receive a description with an optimal image $V^*(\mathbf{X}) S^*$ which corresponds to this image, i.e. to find a possible optimal variant $B^* = (\bar{V}^*, S^*)$. The most essential peculiarities of the two-dimensional grammars are their universality (any set of images can be specified to the corresponding two-

dimensional grammar) and construction (effective algorithms for finding possible optimal variant B^*). Another advantage of such algorithms is that they work directly with visual signals; that are why they permit the use of different methods for image pre-processing. Note that elementary images of grammars, describing complex pattern images can have constant size (that decreases possibilities of two-dimensional programming) and a different size. In the latter case, such elementary images are related to so-called block two-dimensional grammars.

The solution for equation (1) with a larger noise level requires considerable computer calculations for practical problems. However the possibility of computer process paralleling usually exists during the realization of algorithms of two-dimensional programming. With noise reduction, such algorithms are not more complex than another algorithms of images analysis, but they provide a higher result reliability.

However, the two-dimensional grammar of M. I. Shlezinger is meant for the recognition of simplest binary graphical primitives in static scenes. For dynamic scenes with a multi-level motion the system of syntax pattern recognition is more complex. Temporal relationships appear between objects for which the describing of the patterns relationships recognition system design is required. The patterns relationships recognition system realizes four main principles of dynamic object recognition: the recognition aim on initial stages of video sequences processing; the recognition of behavioral situations for dynamic objects; the prehistory estimation of dynamic objects; changeable supervising object numbers in complex scenes.

A context grammar of complex scenes recognition with multi-level objects motion realizes the following procedures:

1. Pre-segmentation of the scene.
2. Description of regions with local features of motion.
3. Grouping of regions with local features of motion according to neighborhood.
4. Video objects recognition.
5. Grouping of video objects with global features of motion according to levels.
6. Description of multi-level motion in scenes.
7. Temporal events recognition in scenes.
8. Scene genre recognition (for multi-media libraries).

Analysis of these procedures shows that in case of recognition complex dynamic scenes with multi-level motion should use the following tree-level grammar:

$$\mathbf{GR}_D = \langle \mathbf{V}_{O,E,G}, \mathbf{S}_{S,LM,GM}, \mathbf{T}_V, \mathbf{T}_S, \mathbf{T}_E, R_E, \\ \{\{\mathbf{E}, \mathbf{E}(a, a'); (a, a') \in \mathbf{M}\}, R_O, \\ \{R_R, \{\mathbf{Z}, \mathbf{Z}(t, t'); (t, t') \in \mathbf{N}\}\}\rangle.$$

where $\mathbf{V}_{O,E,G}$ – is the main vocabulary of the objects, temporal events, scene genres; $\mathbf{S}_{S,LM,GM}$ – is the additional vocabulary of structural elements, local features of motion and global features of motion; R_R – is the predicate of region building; R_O – is the predicate of object building; R_E – is the predicate of temporal events. A set element $\mathbf{E} = \mathbf{V}_{O,E,G} \cup \mathbf{S}_{S,LM,GM}$ is called the event. A set \mathbf{T}_E describes

event sequence. A set $\mathbf{T} = \mathbf{T}_V \cup \mathbf{T}_S \cup \mathbf{T}_E$ in this case is the association of the event and description.

A context grammar of complex scenes recognition with multi-level objects motion realizes two tasks: the task of syntax analysis of image sequence \mathbf{X} (with extended contents of main $\mathbf{V}_{O,E,G}$ and additional $\mathbf{S}_{S,LM,GM}$ vocabularies) according to equation (1), and the task of scene syntax analysis \mathbf{SC} . Let's consider them in detail.

The aim of the syntax analysis of the image sequence \mathbf{X} is the recognition of dynamic objects, which are classified into two large groups:

- objects originating from regions with continual colors, texture features during determined light conditions, and having a fixed set of projections in the frontal plane; contours of region changes in compliance with affine or projective groups of transformations (man-made items);
- objects originating from regions with continual colors, texture features during determined light conditions, and having a random set of projections in the frontal plane; contours of region changes arbitrarily (anthropometrical items). These regions are characterized by constant relative directions and speed values in some temporal interval.

These groups are characterized by a few different features, at which projection scattering of anthropometrical items is compensated by local motion features of separate statistically homogeneous regions. The methodology of the object recognition with restricted possible numbers of projections is well designed; we can propose the following formal scheme for the recognition of such objects. Let's assume that each pattern is represented by only one image. We shall call it the initial pattern template and state it as v^j . Let's also set the possible transformations G_b of initial template parameterized by blending parameter b as given. The result of using the transformation G_b for template v^j is a transformed template:

$$V(j, b) = G_b v^j.$$

The set of values at which template $V(j, b)$ accepts the fixed value j and possible values $b \in B$, is assumed an area of pattern template j . Observable images are the realization of multi-dimensional random quantity with known probability distribution $P(\mathbf{X}/V(j, b))$, depending on the transformed template $V(j, b)$ as well as the multi-dimensional parameter. Value $V(j, b)$ is the expectation value or mode of this distribution.

Such formal schemes permit solving the recognition task for the known distribution $P(\mathbf{X}/V)$ and known functional dependence of transformed template from parameters j and b . Here we can use the method of maximum likelihood. For solving the value parameter j it's necessary to identify the maximum of the likelihood function for parameters j and b , and accept for a solution d such value j for which this maximum is achieved:

$$\mathbf{V}^*(\mathbf{X}) = \arg \max_j \max_{b \in B} P(\mathbf{X}/V(j, b)) \quad . \quad (2)$$

Solution d doesn't change if the likelihood function will be replace by any other function of parameter, values,

which are connected with values $P(\mathbf{X}/V(j,b))$ by steadily increasing function, i. e. if $g(\mathbf{X}, V)$ – any function of parameter \mathbf{X} and V , and satisfies condition:

$$P(\mathbf{X}/V) = f(g(\mathbf{X}, V)), \quad (3)$$

where $f(\cdot)$ – is a steadily increasing function, then rule (2) may be changed by the following expression:

$$\mathbf{V}^*(\mathbf{X}) = \arg \max_j \max_{b \in B} g(\mathbf{X}, V(j, b)). \quad (4)$$

The state would not in be principally changed, if function $f(\cdot)$ will be a monotonically decreasing function. However, the maximization in expressions (2) and (4) ought to be replaced by minimization. Because we can understand (4) as the solution rule for finding such values j and b for which similarity of observed image \mathbf{X} with transformed template $V(j, b)$ is the maximum, then values of any function $g(\mathbf{X}, V)$, satisfactory condition (3), will be a measure of similarity template V with image \mathbf{X} . Value j , found from expression (4), will be our solution.

Syntax analysis of the image sequences with the objects having a random set of projections is a more complex process. It is impossible to predetermine the set of available transformations; it requires a recurrent procedure for tracing regions with local features of motion for their following grouping in a common video object. In this case a template structure $\mathbf{V}^*(\mathbf{X})$ consisting of region sets, each of which has its own set of local motion features. Hence, a recurrent procedure of finding image template $\mathbf{V}_i^*(\mathbf{X})$ at step i , generating by grammar \mathbf{GR}_D , maximizes the similarity function the following way:

$$\mathbf{V}_i^*(\mathbf{X}) = \mathbf{V}_{i-1}^*(\mathbf{X}) + \gamma_i f_V(\mathbf{X}_{i-1}, \Delta_i), \quad (5)$$

where $i, i-1$ – are approximation steps; γ_i – is some function depending from approximation step (for example, a sequence of positive numbers); Δ_i – is changing quantity during the process of image sequence analysis; $\mathbf{V}_i^*(\mathbf{X}) = \arg \max_j P(\mathbf{X}/V(j))$ – is the similarity function on step i .

Function (5) is a variation of the stochastic approximation method for solving the task of pattern recognition learning. It is important to choose a function loss for the organization of the recurrent procedure (5). For example, we can use the following rule: if an image is classified correctly by some separating function, then the penalty equals “0”. If the classification was realized incorrectly, then the penalty is assigned to the value of proportional distance between the vector corresponding to recognizing images and separating hyperplane.

The target of scene **SC** syntax analysis is the recognition of events produced by single objects, interactive objects, and also the definition of the dynamic scene genre. These questions are covered for the problems of image understanding and scene analysis. In the case of complex scenes, before events recognition, it is essential to create a model of multi-level motion, i. e. to define the number of significant levels (in the simplest case to reach a decision on the existence of two levels – foreground and background), and to relate each

recognized video object to one level or another. Such a task is most essential for virtual 3D-reconstruction in cartography, navigation system, and in cases when the video sensor is maintained on a moving platform, and the relative motion of all scene objects occurs. It gives an impression that the objects which are nearer to the video camera “move” faster than remote objects. In this case the model of global motion is similar to the model of multi-level motion which is defined by the set of various but internal similar motion levels, associating with solids located on various distances from the moving camera and k segments (image [3]). Let’s propose that the motion levels assign in a parametric form, and there are h motion levels. For image sequence it is necessary to define: a) the motion level for which each video object is associated; b) the parameter values of each level. For a known motion level, parameter values of the level are determined, and vice versa if we know the parameter values then we may determine to which motion level the video object is associated.

During the scene syntax analysis temporal events accumulating may be classified into motion classes and their interpretation on a conceptual level. The motion in image sequences, accounting their repeatability in time and space, can be classified into three classes: temporal textures, active actions and events. Temporal textures are determined as statistical regularities in space and time (sea waves, movement of clouds, leaves, birds, etc.). Active actions are interpreted as some are repeatable in time (but not in space) structures (walking, dances, separate movements of animals, insects, etc.). Events consist from isolated simple movements but do not repeat in time and in space (expressions, coming into a room, ball casting, etc.).

For dynamic changing scenes there are additional characteristics for objects such as the prehistory of the object motion and the procedure of associations, used for the final concept of forming and interpretation in terms of them. A prehistory of the object’s motion as a coordinate function from time to time may be roughly approximated because it is required to define the properties of motion but not its concrete characteristics. Then the prehistory is interpreted since some event of the object’s motion at the concept level – an analysis of temporal relationships between the objects play an essential role (handshake, discussion, aggressive action, etc.) [4].

For scene interpretation, let’s use a procedure of association which is described by two characteristics [5]:

- association value as measures of similarity calculating the nearness of the vector’s features for the scene objects, a nearness of relative transformations of these objects, and the objects’ significance;
- association similarity is subset of the objects spanned by the global motion event.

For association forming, each object of the knowledge base O_j , besides its direct description, has a set of additional characteristics the values of which are calculated in accordance with common scene dynamics:

- nearness $r_j^t = r(O_j, O^t)$ to interest object determining membership of the trace;

- nearness $q_k^t = q(V_k, V^t)$ connections with adjacent (on relations) objects O_k ;
- value of association $a_j^t = \max_k(0, a_j^{t-1} - c, q_k^t \cdot a_k^t + r_j^t)$, ($c \ll 1$), defining significance of the association trace.

Concept forming is based on proportional association values increasing weights of objects belong to the associations' traces. Concepts are built as often occurring substructures consisting of significant objects.

Consequently in this paper, the basics of formal grammar designs in the context of the structural approach with pattern recognition are considered. The system structure of the syntax pattern recognition, which includes the pre-processing module, the description module, and the syntax analysis module, is proposed. The two-dimensional grammar of M. I. Shlezinger for recognizing the simplest binary graphical primitives in static scenes has been studied in detail. It has shown that for the recognition of complex scenes with multi-level motion objects we can apply a three-level grammar including the main vocabulary of objects, temporal events, scene genres, additional vocabulary of structural elements, local features of motion and global features of motion, predicates of regions building, predicates of objects building, and predicates of temporal events. Procedures of object recognition, based on the possible transformations, and recurrent procedure of statistical approximation, depending on the number of possible video object projections on the frontal plane, have been proposed. The association procedure,

calculating the nearness of the vector features for scene objects, has been designed for the interpretation of complex scenes with a multi-level motion.

References

1. Favorskaya M. N. To the problem of applying formal grammars in indentifying objects in complex scenes // In the materials of the XIII international science conference "Reshetnev Conference". Part. 2. Krasnoyarsk, 2009. P. 540–541.
2. Shlezinger M. I. The syntactic analysis of two-dimensional visual signals in interference conditions // Cybernetics. № 4. 1976. P. 76–82.
3. Favorskaya M. N. Possible methods of video-flow segmentation as a problem lwith missing data // Vestnik. Scientific Journal of the Siberian State Aerospace University named after M. F. Reshetnev. 2010. Ed. 3 (16). 2009. P. 4–8.
4. Video Event Classification using Bag of Words and String Kernels / L. Ballan [et al.] // ICIAP09. 2009. P. 170–178.
5. Favorskaya M. N. Local time-space signs of events in video sequences // In the materials of the X international science conference "Theoretical and practical problems of modern information technologies", Part. II. Ulan-Ude, 2009. P. 461–466.

© Favorskaya M. N., Popov A. M., 2010

E. A. Furmanova, O. G. Boiko, L. G. Shaimardanov
Siberian State Aerospace University named after academician M. F. Reshetnev, Russia, Krasnoyarsk

THE POSSIBILITIES FOR OPTIMIZING THE FUNCTIONAL SYSTEM STRUCTURE OF CIVIL AVIATION AIRCRAFT

This is an analysis of the traditional approach to systems with individual reserving reliability calculation usage. An alternative calculation method for these systems with an individual reserve system has been developed. Its application is demonstrated.

Keywords: functional systems, analysis of the complicated systems, system structure optimization.

The airplanes functional systems execute many important functions: make the planer's steering surfaces drive by mechanization means, provide the aviation engines fuel supply, provide the cabin air pressure and air conditioning in them, provide all the consumers with electricity, protect the airplane from ice, provide fire extinguishing functions, provide automatic piloting, and air navigation.

The functional systems of the same type on all the route airplanes execute the same functions. At the same time, systems with the same name from different developers, or one developer, which is seen more often, on different types of aircraft, may have a different functional systems structure. With the same development in machine building, which provides a similar level of aggregate systems reliability, the reservation level is

different in individual and typical aircraft system types with the same name.

Such a position is connected with the absence of research in the field of the system's optimization structure.

The proposed study shows the possibilities of common and individual reserving in the reliability securing systems.

Let's look through the individual reserving system, which contains n units that are successively linked. Each of the units includes $m = 2$ aggregates connected parallel. The structural scheme of such system is shown in the fig. 1.

Let's think that all the aggregates have the same breakdown stream parameters ω .

Let's consider for a mathematical model of the aggregate's breakdown probabilities the distribution with