

A Syntactic Method in Recognizing Unidentified Objects

N. I. Sidnyaev, Yu. I. Butenko, E. E. Bolotova

Department of Higher Mathematics,
Bauman Moscow State Technical University, Moscow, Russian Federation
E-mail: Sidnyaev@yandex.ru

Received: June 12, 2020

Abstract. The paper deals with the formation of contextual grammars in the methods of complex scene recognition. It proposes the use of multi-level grammar, which includes the task of syntactic analysis of image sequences and the task of syntactic analysis of a scene taking into account the multi-level movement of objects. It is shown that the formation of grammar, describing both the structural information of the image and the interaction of images, is associated with the need to develop an algorithm to output grammar on a given set of dynamic images, which represent a learning sample. As a result of training, structural descriptions of images and descriptions of their relations are formed and later used for syntactic analysis of complex structure events. It is postulated that for dynamic scenes with multi-level movement and complex structure, which is constantly changing in time, it is reasonable to apply context grammar rules, and in this connection arises the concept of multi-level context grammar. Some basic principles of the theory of formal grammars inherent in structural methods of recognition are described.

Keywords: recognition, formal grammar, syntactic method, sign, evaluation, solution, image

INTRODUCTION

Computer vision is one of the most demanded directions in the field of intellectual systems, including complex algorithms of recognition and complex mechanisms of deep machine learning. Computer vision is based on the reproduction of a complex human vision system, allowing to obtain significant information from images, video and other visual means, as well as take the necessary measures or make recommendations based on the information obtained.

Problems appeared in solving problems of image recognition led to the development of various recognition algorithms, one of which is a structural method, also called linguistic or syntactic method [1, 2]. Its peculiarity consists in the fact that the priori descriptions of classes are structural descriptions - formal constructions, in obtaining of which the principle of taking into account the hierarchy of the object structure and the relations existing between the individual elements of this hierarchy within and between the same levels is consistently carried out [3, 4]. The syntactic method of recognition allows to use for the object description the methods of combinatorial regularity of structures, which consist in the fact that operating with a very limited number of atomic (non-derivative) elements and a limited number of rules of

combination, it is possible to obtain a significant variety of descriptions with the help of unlimited (for example, recurrence) application of the rules of combination to the original elements. To build such rules it is possible to apply the theory of formal grammar. In the course of image recognition such elements of the object as subimages, which in turn are divided into more simple subimages and which are in some relations with other subimages, are allocated. This allows to present information about the object in the form of a hierarchical structure, most often with the help of a graph in the nodes of which there are subimages, and arcs indicate the relationship between the selected subimages of the recognized object, as shown in Fig. 1.

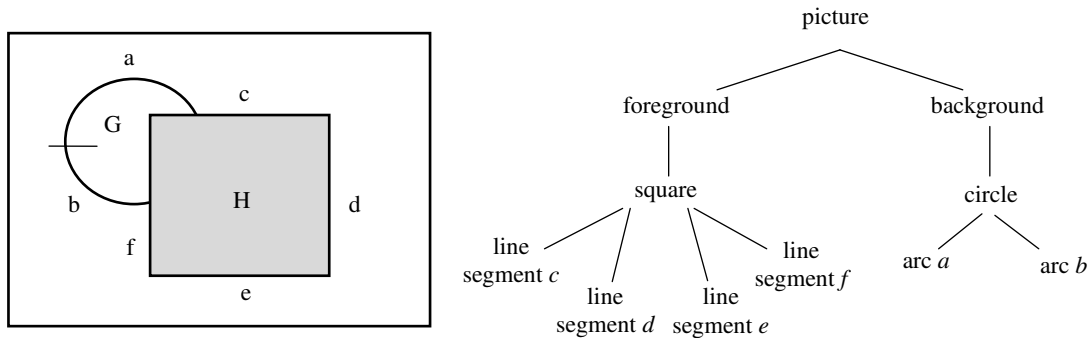


Figure 1. Syntactic image analysis

Within the context of the syntactic approach the decision of a recognition problem is reduced to use of the linguistic constructions containing some dictionary of the recognized object signs and grammar according to which rules elements of the given object are described. Once each subimage inside the object has been identified, the recognition task is realized by performing a syntactic analysis of a sentence describing the object as a tree structure in order to determine whether the sentence is syntactically correct with respect to the grammar indicated. Sentence templates are defined as sequences to be built from selected subimages in different ways, just as sentence phrases are built by concatenation of words and words by concatenation of morphemes. The syntactic approach to image recognition makes it possible to describe many subimages using small sets of simple formal grammar rules [3, 5].

BASIC CONCEPTS OF FORMAL GRAMMAR

Since the implementation of structural methods of pattern recognition is based on the apparatus of formal grammar, we consider some of its basic concepts needed to understand the essence of the structural approach.

The theory of formal languages as a separate discipline usually originates from the studies of the famous American linguist Noam Chomsky, when in the 1950s he attempted to give an exact and unique characteristic of the structure of natural languages. His aim was to define the syntax of languages using simple and precise mathematical rules. It was later discovered that the syntax of programming languages could be described using one of Homsky's grammar models called free-context grammar [6].

In formal language theory, grammar G is a set of rules for generating strings in the formal language. The rules describe how to generate strings from the language alphabet that are valid according to the language syntax. The alphabet is a finite non-empty character set. It is assumed that the characters are indivisible. The word (string) in the alphabet Σ is a finite se-

quence of elements. The generating grammar G is defined as a production system consisting of an ordered set of four elements $G = \langle T, N, I, P \rangle$, where T is a set of terminal characters of the alphabet Σ , i.e. source elements of the dictionary, N is a finite set of nonterminal (auxiliary) characters, I is the initial character of the grammar and P is the finite set of rules.

The construction of grammar begins with the initial symbol I . By applying the rules first to I and then recursively to the result of the previous transformation, it is possible to generate a correct “sentence” of the formal language defined by grammar. Grammar is built using special substitution rules - expressions of the form “ $x \rightarrow y$ ”, which means “substitute x for y ” or “substitute x instead of y ”, where x and y are chains containing any terminal or non-terminal characters. The transformation algorithm stops when the expression no longer contains any nonterminal characters of the alphabet. Thus, according to the rules of grammar, the object is represented by a sentence in this language. To further describe the process of generation it is necessary to introduce such concepts as direct deducibility, deducibility and language generated by grammar [6, 7].

It is important to note that a properly constructed grammar should not contain useless and unattainable symbols.

Definition. The symbol X is called useful in grammar $G = \{T, N, I, P\}$, if there is an output $A \Rightarrow^* \alpha X \beta \Rightarrow^* w$, where $w \in N^*$.

It should be noted that X can be both a variable and a terminal, and the output chain $\alpha X \beta$ is the first or the last in the output.

Definition. If the symbol X is not useful, it is called useless. Obviously, the exclusion of useless characters does not change the grammar generated language, so all such characters can be removed.

Definition. The symbol X is called generating in the grammar $G = \{T, N, I, P\}$, if there is the output $X \Rightarrow^* w$, where $w \in N^*$.

It should be noted that each terminal is a generating symbol, as it is output from itself in 0 steps.

Definition. The symbol X is called achievable in grammar $G = \{T, N, I, P\}$, if there is output $A \Rightarrow^* \alpha X \beta$ for some α, β .

A useful symbol, as the definition suggests, is both generating and achievable. If you remove the non-generating and then unachievable symbols from the grammar first, only useful symbols will remain.

IMPLEMENTATION OF THE RECOGNITION PROCESS BASED ON STRUCTURAL METHODS

In order to recognize an unidentified object on the basis of structural methods, it is necessary first to find its non-derivative elements and the relations between them, and then to determine with the help of syntactic analysis (grammatical parsing) whether the description of the image is consistent with the grammar, which, presumably, could have generated it.

To form the appropriate grammar it is possible to use either a priori information about recognized objects, or the results of the study of the final selective set of “typical” in some sense objects. In the first case they speak about the grammar task on the basis of heuristic considerations, in the second case – about the grammar output.

The practical use of the structural method of recognition requires solving the following main problems:

- 1) construction of the appropriate description of recognized objects;
- 2) the choice of grammar;
- 3) implementation of the recognition process through syntactic analysis procedures;

- 4) the use of training procedures for grammar output;
- 5) the application of procedures from other recognition methods (e.g. statistical to account for random interference and distortion, cluster analysis, etc.) within a structural approach.

Consider the basic techniques used in the practical application of structural recognition methods.

Methods of mathematical morphology

An important part of image recognition process based on structural methods is the preliminary processing of the analyzed image. At the pre-processing stage the object submitted for recognition is subjected to encoding and filtering, restoring and improving the image quality. Such filtration methods are considered in the framework of mathematical morphology, which provides an effective approach to the analysis of digital images [8, 9]. Morphological filtering does not use analytical parameters of signals, but their geometrical characteristics. The object is encoded or approximated in such a way that it is convenient to work with it further. For example, a black and white image can be encoded using a grid (or matrix) of zeros and ones. To improve the efficiency of processing at subsequent stages of work, at this stage often resort to some kind of "data compression". Then, using filtering and restoration methods, distortion is eliminated in order to improve image quality. It is assumed that at the end of the pre-processing stage, images of sufficiently good quality are reproduced.

Selection of non-derivative elements

The first step in building a structural description of the recognition object is to define a set of non-derivative elements with which to describe the considered object. This choice depends on the nature of the object, the type of source data, the application area and the way the recognition system is implemented in practice. At present, there is no universal solution to the problem of selecting non-derivative elements.

Criterion 1. Non-derivative elements shall serve as the main elements of the image so as to provide a compact and relevant description of the source data by means of well-defined structural relations (e.g., connection relations).

Criterion 2. Non-derivative elements shall be easily distinguishable or recognizable by known nonstructural techniques, as they are assumed to be simple and compact images, which "internal" structural information is not taken into account.

Example. Consider that it is necessary to be able to distinguish rectangles of different sizes from other shapes. The following set of non-derivative elements is selected:



Figure 2. Example for describing a rectangle

$a' - 0^\circ$ – horizontal line segment; $b' - 90^\circ$ – vertical line segment; $c' - 180^\circ$ – horizontal line segment; $d' - 270^\circ$ – vertical line segment. A set of all possible rectangles (different sizes) is defined using a single sentence – a chain $a'b'c'd'$ (Fig. 2). If it is necessary to distinguish between rectangles of different sizes, the given description is inappropriate. In this case it is necessary to use segments of unit length as non-derivative elements. Many rectangles of different sizes can be described using the language:

$$L = \{a^n b^m c^n d^m, n, m = 1, 2, \dots\}.$$

Criterion 2 for the selection of non-derivative elements may in some cases conflict with criterion 1 because non-derivative elements selected under the latter may be difficult to recognize using currently known methods. However, criterion 2 permits the selection of sufficiently complex non-derivative elements as long as they can be recognized. The complexity of the non-derivative elements allows for simplified structural descriptions, i.e. less complex grammar. Finding a compromise solution is essential when building real recognition systems based on the use of structural methods.

Use of grammars and languages for structural description of objects

As already mentioned, grammar can be used to generate proposals that represent an object and to conduct grammatical analyses of proposals to determine whether their structure is acceptable from the point of view of the grammar concerned.

Example. Consider the grammar $G = \{T, N, I, P\}$, where $T = \{a, b, c, d\}$; $N = \{I, A, B, C, D\}$;

$$P: \quad I \rightarrow aA \quad A \rightarrow b \quad I \rightarrow cC \quad C \rightarrow d \\ I \rightarrow bB \quad B \rightarrow c \quad I \rightarrow dD \quad D \rightarrow a$$

In this case four sentences can be displayed (\Rightarrow – output symbol):

$$I \Rightarrow aA \Rightarrow ab; \quad I \Rightarrow cC \Rightarrow cd; \\ I \Rightarrow bA \Rightarrow bc; \quad I \Rightarrow dD \Rightarrow Da.$$

If we consider the oriented segments shown in Fig. 2 as non-derivative elements, we can use this grammar to obtain four descriptions of segments forming a right angle. Here we use a technique that allows to describe two-dimensional objects using chain grammar. This method consists in joining the structures only in special points. One of the ways to implement this requirement is that in each structure only two points are selected. In this case, the selected points are interpreted as the beginning and end of the arrow.

Output trees

In the field of computer science the tree is understood as a widely used abstract data type, similar to the hierarchical structure of the tree, with a root value and subtrees - branches with a parent node represented as a set of related nodes [10, 11]. Each tree node has a certain degree that characterizes the number of node subtrees. A node with zero degree is called a leaf. Leaves are nodes from which no branch comes out.

Recognition procedures related to the use of decision trees belong to the group of structural methods, although in the strict sense of the word decision trees are an instrument of hierarchical method of division [11]. The latter is used in cases when the solution tree “contains” many attributes. In each node of the tree one attribute is studied and, depending on the results of this study, the next tree branch is determined. The result of classification is determined at the lower tier of the tree. Such a recognition scheme is very convenient for taking into account a priori information about objects, but it lacks optimal training procedures.

IMPLEMENTATION OF THE RECOGNITION PROCESS USING GRAMMATICAL PARSING

The importance of formal grammars for recognition involves, in particular, the fact that they allow us to speak about the syntactic correctness or incorrectness in relation to a particular grammar representation of the studied image using given non-derivative elements and relations. The answer to this question allows to obtain a grammatical parse procedure [12]. Two

main types of grammatical parse are used: top-down parse and bottom-up parse. The top-down parse procedure consists of consecutive attempts to obtain a given terminal sentence based on the initial symbol by using the rules of corresponding grammar. When using the bottom-up parse procedure, it is necessary to restore the output tree starting from the elements of the main dictionary and applying inverted substitution rules. This procedure starts with a specific sentence and ends when the initial character is obtained.

It is necessary to form a grammar such that when introducing into the grammar analyzer any of the given objects and parsing it according to the received grammar one or several structural descriptions of the object are reproduced at the analyzer output. Then the grammar conversion procedure is performed:

1. All excessive rule entries are eliminated if a rule enters the transformable G grammar several times.

2. Such a pair of auxiliary dictionary elements are found, that identification of these elements (replacing one with another in the whole grammar) will lead to excessive rule occurrence. If there are different options for selecting such pairs, the pair leading to the largest number of redundant occurrences should be selected. After identification of the rules it is necessary to return to step 1.

3. Such a pair N, n is found, which includes an element of the auxiliary dictionary N and a non-derivative element n , that the inclusion of the rule $N \rightarrow n$ in the grammar G (if it is not included in it) and the selective substitution of N by n in the grammar G result in the reduction of the number of the rules after the elimination of the multiple occurrences. If there are different variants of choice of such pairs, it is necessary to choose the pair providing the greatest reduction of the number of rules in grammar. Then it is necessary to return to step 1.

4. The thresholds are entered according to the number of rules to be identified in steps 1-3: replacement is performed if the number of occurrences or the rules to be identified are not less than the threshold value. It is done in order not to decrease essentially dividing force of grammar.

5. A threshold is introduced for the number of rules that are identical in everything (the exception is the type of the non-derivative element): if the number of such rules is greater than or equal to the value of the threshold, the corresponding rules are replaced by one containing a non-derivative element of any type. We should then return to step 1.

If further simplifications are not possible, the procedure is terminated. The result is one grammar. After the specified procedure is performed for each of the grammar formed in the first step, it is necessary to choose the “best” grammar. For example, the ratio of the separating force of the grammar to the square of the number of its rules can be used as the criterion of optimality of the grammar, and the separating force of the grammar is defined as the sum of separating forces of each of its rules; the latter is defined as the sum of the number of non-derivative elements and the number of terms representing a relation or a property.

CONCLUSION

In this paper the syntactic method for recognition of multilevel and poorly structured objects is considered. It is postulated that in order to recognize an unidentified object on the basis of structural methods, it is necessary first to find its non-derivative elements and relations between them, and then to determine with the help of syntactic analysis whether the description of the image is consistent with the grammar, which, presumably, could have generated it. Each object is treated as a chain or sentence because it is composed of elements of the main dictionary. The grammars presented can be used to classify objects, since the presented unidentified object can be attributed to a certain class if it is a sentence of a language. Otherwise,

the object is attributed to another class. As a rule, the object is enrolled to the class in whose language it appears to be a grammatically correct sentence. If the last sentence is not executed, it is obvious that the object does not belong to any of the specified classes and, therefore, one more grammar is required, etc. The object to be recognized belongs to the class of interest in that and only if it is a grammatically correct sentence of the language.

As a criterion of optimality of grammar can be used, for example, the ratio of the separating force. Words can be combined into more complex structures – sentences. Here language is considered as a set of sentences. Suggestions are built from words and simpler sentences according to syntax rules. The syntax of the language is a description of the correct sentences. Alphabet, vocabulary and syntax fully define a set of acceptable language constructs and internal relationships between constructs. The set of syntax rules forms the grammar of the language. Syntax rules can describe either the procedure for receiving correct sentences or the procedure for recognizing the “correctness” of sentences (i.e. their belonging to the given language). In the first case the grammar will be generating, in the second case – recognizing. Criterion for the choice of method may be the simplicity of determining the measure of proximity, the complexity of writing off the boundaries of classes and images, resolution, etc. Thus, it is very important to consider the individual physical features of recognizable objects, informativity of the chosen attributes, quantity and quality of the a priori and current information, possibility of entering of adaptation (weight) factors, etc.

REFERENCES

1. Vasendina, I. S., Aleshko, R. A., Gur'ev, A. T., Karlova, T. V., & Shoshina, K. V. (2017). Razrabotka metodiki raspoznavaniya obrazov na izobrazhenii na osnove strukturnogo podhoda [Procedure development for image identification on picture based on structural approach]. *Bulletin of the Bryansk State Technical University*, (1), 171-177. doi: 10.12737/24908. (in Russian).
2. Prokof'eva, E. V., & Prokof'eva, O. Yu. (2018). Klasternyj, korrelyacionnyj i strukturno-lingvisticheskiy metody v raspoznavanii obrazov [Cluster, correlation, and structural-linguistic methods in pattern recognition]. In *Sovremennye issledovaniya v sfere estestvennyh, tekhnicheskikh i fiziko-matematicheskikh nauk [Modern research in the field of natural, technical and physical and mathematical sciences]* (pp. 679-685). Kirov, Russia : Interregional center for innovative technologies in education. (in Russian).
3. Haustov, P. A. (2017). Algoritmy raspoznavaniya rukopisnyh simvolov na osnove postroeniya strukturnyh modelej [Algorithms for handwritten character recognition based on constructing structural models]. *Komp'yuternaya optika [Computer Optics]*, 41(1), 67-78. (in Russian).
4. Sidnyaev, N. I., Butenko, Yu. I., & Bolotova, E. E. (2019). Ekspertnaya sistema produkcionnogo tipa dlya sozdaniya bazy znaniy i konstrukciyah letatel'nyh apparatov [Rule-based expert system for creating a knowledge base on aircraft structures]. *Aviakosmicheskoe priborostroenie [Aerospace Instrumentation]*, 6, 38-52. (in Russian).
5. Singh, M., Khan, H., & Gupta, R. (2018). Digital image processing a formal grammar approach. *International Journal for Research in Applied Sciences and Biotechnology (IJRASB)*, 5(2), 3-5. Retrieved from <https://www.ijrasb.com/ojs/index.php/ojs-ijrasb/article/view/44>.

6. Hodzhabeqyan, M. S., Avramenko, A. A., & Zejtunyan, V. M. (2017). Porozhdayushchaya grammatika N. Homskogo [N. Chomsky's generating grammar]. In *Razvitie nauki i tekhniki: mekhanizm vybora i realizacii prioritetov. Chast' 5. [Development of science and technology: a mechanism for selecting and implementing priorities. Part 5]* (pp. 92-94). Ufa, Russia : Aeterna. (in Russian).
7. Glushkov, V. M. (1961). The abstract theory of automata. *Russian Mathematical Surveys*, 16(5), 1-53. doi: 10.1070/RM1961v016n05ABEH004112. (in Russian).
8. Vizil'ter, Yu. V. (2008). Obobshchennaya proektivnaya morfologiya [Generalized projective morphology]. *Komp'yuternaya optika [Computer Optics]*, 32(4), 384-399. (in Russian).
9. Shashev, D. V., & Shidlovskij, S. V. (2015). Morphological processing of binary images using reconfigurable computing environments. *Optoelectronics, Instrumentation and Data Processing*, 51(3), 227-233. doi: 10.3103/S8756699015030036. (in Russian).
10. Sidnyaev, N. I. (2016). *Nejroseti i nejromatematika [Neural networks and neuromathematics]*. Moscow : Bauman Moscow State Technical University, 83 pp. (in Russian).
11. Finogeev, A. G., & Chetvergova, M. V. (2012). Razrabotka i issledovanie metodiki raspoznavaniya izobrazhenij dlya sistem rasshirenoj real'nosti [Development and research of image recognition techniques for augmented reality systems]. *Izvestija volgogradskogo gosudarstvennogo tehničeskogo universiteta [Izvestia VSTU]*, (15), 130-136. (in Russian).
12. Favorskaya, M. N., & Goroshkin, A. N. (2008). Model' raspoznavaniya izobrazhenij rukopisnogo teksta [The invariant model for image recognition of hand-written text]. *Vestnik of SibGAU*, (2), 52-57. (in Russian). Retrieved from https://vestnik.sibsau.ru/en_US/vestnik/929/.