Pattern Recognition and Image Recognition

Y. I. Zhuravlev† **and I. B. Gurevich**††

†*USSR Academy of Sciences Computer Center, Moscow, USSR*

††*Scientific Council "Cybernetics" of the USSR Academy of Sciences, Moscow, USSR*

Abstract - The paper sums up the progress, and describes the present state, of recognition in the USSR as a data-processing technique (one of the first to be widely adopted and used in practice) that has a well-developed body of mathematics and excellent application capabilities. It sets forth mainly the problem of recognition as a whole and the present state and major results of the mathematical theory of recognition (largely for standard input data). The paper also analyzes the problem of image recognition, including the concept of the descriptive approach to the recognition and understanding of images.

Basic points of the review:

a) Pattern recognition has a well-developed and, in a sense, complete mathematical theory on the basis of the so-called "algebraic approach".

b) The problems of recognition using standard data and that of image recognition differ so much that the methods and devices of classical pattern recognition theory cannot be used for the latter directly. Although set and solved within the general methodology of recognition, image recognition problems require a special branch in recognition theory to be developed specifically for image processing.

c) Image recognition theory can be advanced on the basis of the descriptive approach as a descriptive theory of image recognition.

INTRODUCTION

A key issue in information science is the development, study and realization of synthesis methods by algorithmic procedures for transformation and analysis of data to solve information problems whose algorithms are unknown. For over fifty years now, these methods have been directly or indirectly the core of mathematical algorithm theory, cybernetics and later information science. The problems that require such methods arise in computer processing and transformation of structures which are formed from symbols, i.e. those which represent knowledge of the problem domain as a whole and knowledge of a specific problem in artificial intelligence programs. Although the problems are general, these methods, have been thoroughly studied, advanced and finally integrated into a complete mathematical theory for a single, albeit quite broad, class of problems of data transformation and analysis. These have long been known by the name (perhaps not quite accurate or apt) of "pattern recognition problems".

The review contains mainly a general description of the recognition problem as a whole (Chapter 1) and of the present state and major results of mathematical recognition theory largely for standard input data (Chapter 2). It also analyzes the problem of image recognition, including the concept of the descriptive approach to image analysis. (Chapter 3)[1].

Manuscript received May 20, 1991

[1] The review has used some other papers by the same authors, particularly [11, 20, 21]).

CHAPTER 1

THE RECOGNITION PROBLEM

1.1. Basic concepts

The term "pattern" is presently used in a very general sense in recognition, and especially artificial intelligence, to mean a structured approximate (partial) description (sketch) of an object or a phenomenon under study, the basic property of the pattern being partial definition of description. The pattern may have a recursive definition. A symbol is a pattern and a list of symbols is a pattern. Only expressions constructed in compliance with the two given conditions are patterns. Listing makes it possible to use the same representation to describe a pattern of an arbitrary type regardless of its "content". Its additional advantage is that the same algorithms can be used to handle patterns with different denotations. It is also natural to assume that a pattern consists of two groups of symbols which represent variable and constant characteristics of a described object, respectively.

The main function of the descriptions (patterns) is in establishing correspondence of objects, i.e. to prove their identity, analogy, similarity, resemblance, etc., by comparison. Two patterns are considered to be similar if their correspondence can be established. It can be assumed, for example, that two patterns correspond if their identity can be achieved by substituting some expressions for variables.

The comparison of patterns is the key problem in recognition and it plays an important part in information science in general. The problem arises, in particular, in various areas of artificial intelligence, including understanding of natural language, symbolic processing of algebraic expressions, expert systems, and transformation and synthesis of computer software. The comparison procedure has proved to be so important for artificial intelligence that many programming languages used in the latter include it as a primitive.

Notice that the concept of pattern has different meaning in different problems. Thus, for example, in recognition (in classical models), a pattern is usually described by a vector of features, each element of which represents a numerical value of one of the features that characterize an object. A pattern in the structural model of recognition is a statement generated by the grammar which characterizes the class the pattern belongs to. A pattern in word processing problems is a chain and thus the procedure to set up a correspondence is actually a search for occurrences of the chain (pattern) in a text.

The term "recognition" covers both the processes of perception and cognition peculiar to humans and living forms in general, and attempts to realize and use "mechanical" analogs (by function and result) of these processes, whose study and synthesis is the subject of recognition as a branch of information science. This Chapter will discuss only the latter.

Thus the purpose of creating computer systems for recognition is to automate the group of perception and cognition processes that involve the search for, and extraction, identification, classification, and description of, patterns on the basis of analysis of actual data obtained in some way. Patterns are usually searched for and extracted at the initial stage of analysis during input-data processing to produce intermediate results (i.e. to convert the input data into some other form) which "better" represent the patterns for the solution of a problem. The next stage - the development of a "classifier" - generally includes the analysis of sample (transformed) data, the synthesis of a model to take into account variation of patterns of a certain category, and the selection of a certain subset from a specified collection of characteristics, which adequately characterizes some categories of objects. Methods for selecting this subset are determined and the recognition algorithm (classification) is developed also at this stage.

1.2. Epistemological aspects of recognition

As more and more sophisticated methods of recognition were developed, a need arose for some regular basis and standardized techniques to compare individual heuristic recognition algorithms for their computational complexity, effectiveness, accuracy and speed, to select an optimum algorithm in a model, and finally to automate the selection and synthesis of algorithmic procedures for solving a specific recognition problem.

Not supported by a reliable foundation of a mathematical theory of recognition, these attempts produced only partial spontaneous achievements. They helped to realize that the capabilities of even the most sophisticated heuristic models of recognition algorithms are limited when they are used "independently" and that it is necessary to establish connections between individual models, i.e. to develop a general theory of recognition. This increased, of course, the interest in the epistemological status of basic concepts of recognition such as pattern, category, and recognition algorithm. These problems were studied by philosophers who were interested in the epistemological aspects of cybernetics and information science (see, for example, Ref. [31] by V.S. Tiukhtin) as well as mathematicians who tried to develop the theory of recognition. Since Chapter 2 describes in sufficient detail the most important principles of the mathematical theory of recognition based on a so-called "algebraic approach" to the problems of recognition and classification [20], we will dwell here only on the idea of combinatorial regularity put forward by U. Grenander [7] and the paradigmatic symbol by S. Watanabe [36], which made a strong impact on the establishment and development of the theory of recognition.

In his theory, U. Grenander proceeded from the hypothesis that the search for regularity was the predominant theme in human attempts to understand the surrounding world and any such attempt was based on an explicit or implicit assumption that natural phenomena and events of the man-made world were governed by certain laws which determined their order and structure. He also relied on the following thesis of D. Hume, which underlies the reasoning by incomplete induction: "If we were governed by reason, we would follow the principle that events which we have had no experience of should be similar to events which we have experienced and that the laws of nature never change" [33].

Grenander's theory is based on the idea that the world is structured, i.e. that there exists a sufficiently strong regularity manifested by permanent connections and laws. The point of departure for the theory is an object - a pattern as such - and the problems of the nature of the pattern, prototype and category, which it gives rise to. The patterns are regarded within the framework of strict formalism, which is used as a basis for the synthesis and analysis of patterns and thus helps to understand the way the patterns are set and processed. As a result, the procedures employed to describe, approximate, restore and recognize patterns take the form of natural consequences of the procedures used to form and transform objects.

Principles of describing regularities are known to be found or "invented", logically analyzed and their consequences obtained by so-called "formal systems" understood as a number of basic assumptions, procedures and rules indicating how they should be applied to explain a particular phenomenon. A formal system that describes a regularity should have some constancy as regards time and space. If this system is applicable only to a certain time and a certain place, we do not have a natural law but data, the results of isolated observations, since it is assumed as a rule that a law, an order or a pattern is something more than just individual facts. Laws deal with several alternatives and interesting laws with a great number of alternatives. So, according to Grenander, a pattern should be correlated with an ensemble of possible cases and the order in this ensemble is regarded as unified execution of certain properties.

The principal object of Grenander's theory of patterns is combinatorial regular structures - regular configurations - which are logical constructions that allow the determination of various types of regularity. Formally, it is construction of new objects by combining specified objects according to certain rules.

It was postulated that patterns were formed from simple standard elements called "generators". These are indivisible parts (atoms) which are selected in accordance with the "physical" nature of the objects or phenomena to be studied. The elements may be symbols, sets, relations or functions. The generators may seem to be entirely different but their role in generating regular structures is identical.

S. Watanabe [36] discussed basic concepts of recognition using the conception of "paradigmatic symbol" (from the Greek "paradeigma" - an example or model) and relying on a comment by L. Wittgenstein: "To see something_1 as something_2". He understands a pattern as the opposite of chaos, as an entity which is not clearly defined but which can be named, i.e. something. According to Watanabe, the pattern corresponds to "something_2" and the latter does not exist as such in the language at the same level as "something_1" is present in an image. "Something_2" may or may not have a name, or the structure of a nameless construction must be described by listing its components and indicating a well-defined method for establishing their relationship. Notice, by the way, that, as shown by numerous data of perception psychology and psychiatry (visual illusions, L. Wittgenstein's images of double meaning of the "duck/rabbit" type, Gestalt data on switching from a figure to the background, or Rohrschach inkblots), the important thing is that "something_2" is generated by interaction of an external stimulus and thinking processes which establish the connection between an object under consideration and another object similar or at least related to the former.

Substitution of the verb "to recognize" for "to see" and "pattern" for "something_2" in L. Wittgenstein's formula leads to the idea of a category combining a number of individual objects. The next natural step is to see recognition as identification of some object which is an element of a certain known set, i.e. as a process of mapping whereby the same element of a set is placed in correspondence with different elements of another set.

Recognition is usually associated with two functions, namely assignment of an object to a category of objects unknown to the classifier, and identification of an object as an element of a category known to the classifier. The first function is the process of isolating new categories - so-called "clustering" - and the second function is recognition proper (recognition in the narrow sense or recognition by paradigm according to Watanabe). The latter interpretation of recognition is based on the idea of the pattern as some object used as a sample of the objects which comprise a category. And it is assumed that some group of objects is described ("represented" or characterized) by a typical example, model, or prototype. On the other hand, the etymology of the word "pattern" allows also an interpretation which does not require the existence of any "main object" wherefrom all other versions of the "pattern" originated. This duality is reflected in the epicentral and aggregate approach to the definition of a pattern and it greatly affected the methodology of recognition in general.

1.3. The purpose of recognition. Characterization and types of recognition problems

The central task of pattern recognition is to produce efficient computer software on the basis of systematic theoretical and experimental studies to be used to assign formalized descriptions of situations and objects to appropriate categories. This assignment (recognition or classification) is based on some aggregated estimate of a situation obtained from its description. When a correspondence is set up between equivalence classes specified for a set of decisions and a set of recognition objects (situation), the automation of recognition procedures becomes an element of automation of the decision making process.

Recognition problems are in effect discrete analogs of problems of searching for optimum solutions. These include a broad class of problems which must establish from some, generally quite heterogeneous, perhaps incomplete, unclear, distorted or indirect information, whether some complex situations (objects or phenomena) under study have the fixed finite collection of properties required for them to be assigned to a certain category. Those are problems of recognition and classification. Or it is necessary to find out on the basis of the same type of information about a finite set of sufficiently similar processes what region out of a finite number of regions these processes will be in after a certain period of time. These are prediction problems that comprise problems of technical and medical diagnostics, geological forecasting (in particular, the restoration of geological fields), or prediction of properties of chemicals, alloys and new materials. Other areas where they are used include recognition and characterization of properties of dynamic and static objects in a complex noise environment with a great deal of passive and active interference with images produced by various technical devices, prediction of the progress of large construction projects, processing of data from remote exploration of natural resources, crop forecasting, detection of forest fires, control of production processes (estimation of the possibility of parameter values of high-speed processes entering critical regions), etc.

1.3.1. General description of recognition problems

Practical problems suitable for pattern recognition methods have a number of specific features, including:

1. Information problems to be solved by using a system of transformations of accessible input data. In the general case, these comprise two basic stages: a) reduction of input data to some standard form convenient for recognition; this is the synthesis of a formalized description of a situation (object) from available heterogeneous information (empirical data, results of measurements, knowledge of logical aspects of the phenomena (processes) being studied, information about the design, function and performance (probably assumed) of the object, expert data, and available a-priori semantic information), and b) recognition proper - the transformation of the formalized description into a standardized answer matrix that corresponds to the choice of one possibility out of some finite fixed collection of possibilities for an answer (classification decision) which indicates that the situation (object) belongs to a certain class.

2. It is possible to introduce into problems the concept of a certain similarity between objects (situations) or, to be exact, between their descriptions. A generalized concept of closeness can be formulated as a basis for assigning situations (objects) to the same class or to different classes.

3. In some problems one can operate with a certain collection of precedents or examples whose classification (in the sense of the problem to be solved) is known and which can be produced to a recognition algorithm for adjustment in the process of learning.

4. Problems that are hard to formulate a theory about, and to use classical mathematical methods since one of the following two cases takes place in the situations where they arise: a) the level of formalization of the appropriate object domain and/or accessible information is such that it cannot constitute a basis for the synthesis of a mathematical model which could satisfy classical mathematical or mathematical-physical canons and allow study by classical analytical or numerical methods, or b) a mathematical model can be built in principle but its synthesis or study involve costs so high (collection of the required data, computational resources and time) that they greatly exceed the gain obtained from the solution or are beyond technical capabilities, or make the solution of the problem simply senseless.

5. The input information that the problems contain is "bad" by definition. It characterizes a situation (an object in some environment) which is complex semantically and structurally. The information is limited, incomplete (with gaps), heterogeneous, indirect (it provides characteristics of outward manifestations of a process in operation, not always basic for the mechanism that drives it), unclear, ambiguous, and probabilistic. On the whole, these are problems about which too little is known to use classical solution methods (models) but still enough for a solution to be possible.

1.3.2. Types of recognition problems

1. Assignment of an object (situation) to one specified category on the basis of its formalized description is a recognition problem (supervised learning).

2. Partitioning of a set of situations (objects) into a system of disjoint subsets (categories) on the basis of their formalized descriptions is a problem of automatic classification (taxonomy, cluster analysis, or unsupervised learning).

3. Determination of an informative sample of features to compile a formalized description of the object of recognition, and estimation of the information content of individual features and their combinations is a problem of selection of an informative sample of features in recognition.

4. The compilation of a formalized description of the object of recognition is a problem of reducing the input data to a form suitable for recognition.

5. Problem 1 allowing for the dynamics of an object (situation).

6. Problem 2 allowing for the dynamics of an object (situation).

7. Problems 5 and 6 where a solution should be referred to a certain moment and time in the future. This is a prediction problem.

1.3.3. Types of input information in the problems of recognition and prediction

All the problems listed in 1.3.2 can be solved when their input data are set in one of the following forms or combinations thereof:

1. Images obtained in various parts of the radiation spectrum (optical, infrared, ultrasonic, etc.) in various ways (television, photographic, laser, radar, radiological, etc.) and converted into digital form.

2. Signals (long numeric sequences).

3. Expert data, numerical and other types of symbol information in the general form.

4. Series of images ("motion pictures") of any type listed in 1.

CHAPTER 2

PATTERN RECOGNITION. MATHEMATICAL THEORY

2.1. Evolution of the pattern recognition problem and approaches to its solution

The problem of recognition has long attracted experts in applied mathematics and then in information science. Particularly noteworthy are the work done by R. Fisher in the 1920s, which led to formation of discriminant analysis, a problem of separation of a mixture of two distributions formulated by A.N. Kolmogorov and A.Ya. Khinchin in the early 1940s, and the theory of statistical solutions. The 1950s and 1960s saw a great number of papers devoted to the search for and application of algorithms that could assign a new object to one of specified categories or divide a set of objects into several disjoint categories. The image of recognition as an independent scientific branch began to change somewhat by the mid-1970s as a normal mathematical theory of recognition became possible.

One of the prerequisites for the possibility was a number of models of recognition algorithms selected and developed in the course of solving application problems of data processing. This was a family of algorithms to solve classification problems. The following models have been studied and put into practice since then.

1. Models based on separation principles (R models). These differ mainly in specification of the category of surfaces wherein a surface (or a collection of surfaces) is selected to separate elements of different categories in the best possible way (e.g. see [30 (Ch. 2)], [26 (Ch. 4)]).

2. Statistical models. These types of recognition algorithm models are based on mathematical statistics. The models are used where probabilistic characteristics of categories, e.g. appropriate distribution functions, are known or can be found quite easily (e.g. see [5, 6 (Ch. 2, 3)], [30 (Ch. 4, 6, 7)]).

3. Models formed by the so-called "method of potential functions" (P models). The model is based on the idea of potential, borrowed from physics. The potential can be determined for any point in space and depends on the position of its source. The potential function - a distance function which is positive everywhere and decreases monotonically - is used as a function to define whether an object belongs to a category [1].

4. Estimation (voting) models (V models). These are based on the principle of partial precedence. The "closeness" between parts of descriptions of objects classified earlier and an object to be recognized is analyzed. The presence of closeness is a partial precedent and is evaluated by a certain preset rule (by numerical estimate). Using a collection of estimates of closeness, a general estimate of the object is worked out for a category. It is this estimate that is used as the value of the function to define whether an object belongs to a category [6 (Ch. 3)], [12, 16, 23].

Models based on computation of statements using, in particular, logic algebra (L models). Categories and attributes of objects are regarded in these models as logical variables and the description of categories in the language of attributes is represented in the form of Boolean relations (e.g. see [6 (Ch. 4)], [24]).

The merits, achievements and prospects of recognition are perceived outwardly in essentially "classification" perspective. But another side of the matter is no less important. The point is that the advancement of recognition is an excellent model of development of a mathematical theory of data processing and transformation. In the process, heuristic (at least in essence) methods were rigorously substantiated and started to be used within quite formalized regular procedures. It is interesting to note that recognition itself is a sufficiently well-developed version of such theory since its main task - to synthesize and select algorithmic means for extracting useful information from data of the above-described nature - can be carried out today.

It is well-known that a recognition problem is set when it is difficult to construct formal theories and apply classical mathematical methods. That usually happens for two reasons: a) the level of formalization of the appropriate object domain and/or available information are such that they cannot make the basis for the synthesis of a mathematical model to satisfy classical mathematical or mathematical-physical canons and to allow study by classical analytical or numerical methods, or b) a mathematical model can be built in principle but its synthesis or study involve costs so high that they greatly exceed the gain obtained from the solution or are beyond existing technical capabilities, or make the solution of the problem simply senseless.

Thus, the "duality" of recognition was manifested in that the solution of such problems introduced into practice a great number of inaccurate (heuristic) algorithms. The bulk of the recognition theory applications has long been concerned with poorly formalized fields such as medicine, geology, sociology, or chemistry. It is still difficult to make formal theories and use standard mathematical methods there. The best that could be done is to provide a mathematical form to some intuitive principles and then to use the resultant "empirical formalisms" to solve partial problems. This was the reason why so many different methods and algorithms appeared at the early stage of the theory and practice of recognition, to be used for solving practical problems without any serious validation. When a problem or a class of problems was studied on the basis of so-called "likelihood" reasoning, a non-rigorous but fairly reasonable method of solution and its algorithm were suggested. And the validation was found directly during experiment with the problems. The algorithms which passed this test by experiments, i.e. those which proved to be successful in solving certain practical problems, continued to be used despite the absence of mathematical validation.

It became obvious that the appearance of each heuristic algorithm of this type could be regarded as an experiment and the total set of experiments and their results could be handled as a set of objects new for mathematics. In other words, the set of inaccurate procedures to solve poorly formalized problems could be studied using rigorous mathematical methods.

Therefore, the second stage in the development of recognition theory was distinguished, for one thing, by attempts to set and solve the problem of selecting the best (in a sense) algorithm in a specific situation and, for another, by attempts to transfer from description of individual inaccurate algorithms to description of the principles for their formation. These were the attempts to provide unified descriptions for sets of procedures which were heuristic but successful in solving actual problems. Such a set is specified by indicating variables, objects, functions and parameters and by defining exactly the regions of their variation. The specification of these variables, objects, functions and parameters makes it possible to isolate some particular algorithm out of the appropriate set, i.e. a model. The class of estimation algorithms was the first to be represented as a model but later appeared descriptions of other models.

The need for the synthesis of models of recognition algorithms was dictated in the first place by the necessity to specify somehow the class of algorithms where an optimum, or at least acceptable, procedure is selected to solve a specific problem. Attempts to construct such models aroused, in turn, an interest in "mathematical" properties of recognition algorithms as such and particularly in their rigorous validation. It turned out that the problem of describing a class of algorithms is similar to that of the classical definition of an algorithm. So the necessary condition for formulating the theory of recognition is to conduct classical algorithmic investigations for the concept of recognition algorithm.

As inaccurate recognition algorithms were accumulated, an analysis of the whole lot made it possible to isolate and describe, in addition to individual partial algorithms, also the principles of their formation. The principles already valid for subsets of algorithms and formulated at first in a poorly formalized form, could be realized then as accurate mathematical descriptions. The choice of a principle was heuristic at this stage while algorithms generated on the basis of the principle could be compiled as usual. It was in this sense that formalization of various principles of compiling recognition algorithms resulted in the appearance of models of recognition algorithms.

The use of models of recognition algorithms as such did not lead to a universal model nor to formalization of selection of a particular model to solve a specific recognition problem. But the models allowed one to set and solve, within a certain model, the problem of selecting an algorithm extreme by the quality functional of classification or prediction. The construction of such optimum algorithms boils down usually to the study, realization and development of computational schemes for non-standard extreme problems.

Parametrization of some recognition algorithms (models) and the ability to determine values of the parameters from available information about categories truly makes it possible to select correct algorithms for some subsets of problems. In most practical cases, however, this subset proves to be rather narrow since otherwise a very large volume of a-priori information would be required to synthesize the model of the recognition algorithm, to describe categories and to select features of the objects of recognition. But this information can be obtained only if we have a sufficiently accurate model of the objects and phenomena under study. In addition, the construction of an optimum algorithm in a multiparametric model involves solutions of difficult extreme problems (often NP-complete ones). More often than not a global extreme cannot be found while the use of algorithms that correspond to a local extreme reduces considerably the quality of recognition and does not allow one to realize the potential of the model. It turns out sometimes that low-parametric models with which a global extreme can be found produce a greater effect than a local extreme algorithm in a multiparametric model. Moreover, there is no guarantee that the algorithm, optimum in the model, will remain the same in handling objects which do not participate in learning [4, 18, 20].

The validation is done at the second stage using one of the following three methods:

1. By experiment. That it is possible to obtain a "solution" of the problem, acceptable to the user, by means of the appropriate recognition algorithm is regarded as proof of its validity for the given problem.

2. By solving an optimization problem and using a recognition algorithm which is optimum within the framework of the selected model. The validation consists in the use of the best possible recognition algorithm for this model.

3. The validation is the same as in 2 but it is proved, in addition, that if "a number" of "natural" hypotheses (conditions) valid for this class of problems is satisfied, the algorithms which are optimum in the model truly ensure high quality of recognition, i.e. both the choice of the algorithms and the choice of the model are validated.

The next stage in the development of recognition involved the study of the structure of all ill-defined algorithms as a whole. Since it was found that a more complex model often did not produce comparatively better results and, moreover, there was a natural limit to the complexity of any model, it occurred to some researchers that algorithms could be selected from existing families and an optimum algorithm could be compiled from the original ones using appropriate corrective operations.

The so-called "result corrector" was one of the first variants of the idea. With it, a solution of a recognition problem was formed from the results of the input-data processing by individual algorithms [29]. It proved, however, that there were no "good" simple operations in the natural sense of the word, which could provide the necessary correction even where the answers "YES", "NO" and "DO NOT KNOW" were considered as acceptable algorithms. The trouble was that the space of input informations and the set of possible answers depended on the content of a problem. Therefore the former consisted of sufficiently complex elements (usually vectors of very large dimensions) and the latter was very poor - ($\{0.1\}$).

As a solution, a method was proposed for defining a recognition algorithm, which covered all the existing types of algorithms, combined with the so-called "algebraic approach" to the problems of recognition and classification that ensured efficient study and a structural description of classes of recognition algorithms.

As we have mentioned above, a description of the classes of recognition algorithms is a problem similar to the general definition of an algorithm (as it was provided in the classical papers by Church, Turing, A.A. Markov and others). In these papers, the intuitive perception of an algorithm as a regular mass procedure was transformed into rigorous definitions which set the mathematical model of the concept of algorithm. The correspondence of the model to intuitive perceptions is impossible to prove mathematically since it requires the comparison of a formally specified object with objects that have no rigorous formal descriptions. The mathematical model of an algorithm is valuable in that it agrees quite well also with the intuitive perceptions of the algorithm, the agreement having been established by questioning experts. In fact, the validity of the definition of an algorithm is proved by the fact that an overwhelming majority of experts consider it to be correct. Thus, even for the most formalized mathematical discipline - mathematical logic and the theory of algorithms - one of the basic concepts is validated by a sort of processing of observation results.

To formulate a rigorous recognition theory required similar developments for the concept of recognition algorithm. In other words, as in the case for the concept of algorithm, it was necessary to translate intuitive (heuristic, ill-defined) perceptions into the language of definitions. The latter is equivalent to the need to formulate a mathematical model of the recognition algorithm no less convincing (by the expert criterion) than the one for the concept of algorithm.

Thus the necessary condition for the rigorous theory of recognition was the classical algorithmic study of the concept of recognition algorithm. The definition was to satisfy also some other conditions, the key condition being that this definition could be used at the next stages of the study of the recognition algorithm. This requirement was quite strong in actual problems [20].

The study of models of recognition algorithms produced interesting theoretical results and helped to solve various application problems. At the same time, the method used to solve recognition problems had some serious drawbacks which could not be remedied, it seems, in handling only individual models. To overcome the difficulties, a general theory of recognition algorithm was proposed. It is based on an algebraic approach to problems of recognition and classification, which ensures effective study and a constructive description of the class of recognition algorithms, and provides a definition of the recognition algorithm, which covers all the existing models of algorithms [18-20].

The algebraic approach consists in enriching original heuristic families of algorithms by means of algebraic operations and constructing a family which guarantees a correct algorithm that can solve the given class of problems. It is based on the idea of inductive generation of mathematical objects through a generalized inductive definition. Basic algorithms and recognition models are isolated and operations with them are introduced to generate successively new algorithms and models. Conditions are found wherein the given family of algorithms is basic with respect to the operations, as are the properties which a model must have to contain an algorithm that can classify correctly all the objects of an arbitrary final sample. Methods for building such algorithms are formed. The idea of the approach is that the family of such algorithms is regarded as an algebra whose operations can be used to construct an expansion of a family of algorithms on that family's basis such that it contains a correct algorithm which can classify the final sample in all the categories.

The algebraic approach makes extensive use of the peculiarities of the structure typical of any recognition procedure. It introduces the so-called "space of estimates" which is intermediate with respect to initial descriptions and acceptable answers. And the recognition algorithm is regarded as the superposition of two operators. The first of them - the recognition operator - forms elements called estimates as answers, while the second operator (the decision rule) finds final answers from the estimates. So the necessity to deal with "inconvenient" spaces of initial descriptions and acceptable answers is replaced by the possibility of performing corrections in the space of estimates (more often than not a set of real numbers).

An important aspect of the algebraic approach is the concept of completeness that connects individual problems and models of algorithms; the completeness of a problem relative to a model means that with an arbitrary collection of a-priori classifications an algorithm can be constructed for the objects under consideration within the model to give a correct answer at all times. The completeness of a problem with respect to a model directly implies that the model contains an algorithm that ensures absolute accuracy on the learning data. It is important that the formulation of an extreme algorithm proves in most cases to be a problem which is rather easily solvable by standard mathematical methods.

A number of investigations has been conducted with the algebraic approach to study and validate advanced methods (some of those studies were described in issues 1 and 2 of the yearbook). It was found that the problem of the boundary of a set of corrective operations, beyond which the extension does not produce a real effect, is related to specification of an acceptable method for the use of information by the algorithms. Formalization and later studies of a meaningful idea of an acceptable method for application of information by recognition algorithms produced some final estimates for models of algorithms and sets of corrective operations. Thus, in particular, a universal upper bound was obtained for the power of the sets of operation of the polynomial type, and lower bounds were set for the complexity of models of recognition operators for computation of estimates and of models of recognition operators based on the partitioning principle.

It was demonstrated that the families of algorithms which are formed with the algebraic approach have a limited capacity and so the use of such families is correct only when sufficiently general hypotheses of statistical nature are true [27, 28]. Extreme algorithms formed under the algebraic approach proved to have in many cases a nonzero radius of stability. That means that when input information varies but little in a sense, the classification generated by an extreme algorithm is preserved. In other words, when sufficiently general assumptions about compactness are satisfied, the classifications generated by extreme algorithms converge almost everywhere to a true classification. Research was also conducted to study the possibility of the simplest representation of extreme algorithms

Along with the transfer in recognition from individual algorithms to models, another branch of research developed, which was concerned with application of algebraic methods to extend the types of input informations acceptable in recognition problems. Noteworthy in this respect is the pattern theory of Grenander and descriptive theory of the analysis of images developed within the framework of the algebraic approach (see Ch. 3 and [7-11, 13, 18]).

To sum up the above we would like to emphasize that the methodology of recognition is used in two capacities in information science:

- first, for its proper purpose to solve problems of recognition in the classical sense; and

- second, as a tool for accurate study of ill-defined problems.

The methodology is realized in the latter case roughly as follows. Let there be, for example, some data obtained by a physical or simulation experiment. The data characterize an object or a situation under study in some quite limited respect. One must try to bring the data together to find what laws are reflected in the available information. For this purpose, a simple hypothesis is put forward and it is given a mathematical appearance. An attempt is made to use the hypothesis "to explain" the available data. A successive use of a number of heuristics (realizations of the hypothesis) may help to guess the model. Otherwise there is a search within the framework of the model generated by a heuristic and then a search for an optimum (adequate) heuristic principle - a model. If it turns out that the principle is non-existent or it cannot be used in practice, a certain conglomeration of principles should be formed to produce a "federative" principle. It is this upper level that is adequate to the capabilities and function of the algebraic approach.

2.2. The mathematical formulation of the recognition problem

Given a set M of objects ω, there is a partitioning into a finite number of subsets (categories) Ω_i, $i = 1, ..., m$, $M = \bigcup_{i=1}^{m} \Omega_i$ on this set. The partitioning M

Table 2.1. Standard form of the training table $T_{N,\, m}$

Objects	Features and their values							Classes
	x_1	x_2		x_j		x_N		
ω_1	$a_{1,1}$	$a_{1,2}$	...	$a_{1,j}$	...	$a_{1,N}$		Ω_1
ω_2	$a_{2,1}$	$a_{2,2}$	...	$a_{2,j}$	...	$a_{2,N}$		
....	....	....	...	....	...	....		
ω_{r_1}	$a_{r_1,1}$	$a_{r_1,2}$	...	$a_{r_1,j}$	...	$a_{r_1,N}$		
$\omega_{r_{i-1}+1}$	$a_{r_{i-1}+1,1}$	$a_{r_{i-1}+1,2}$	...	$a_{r_{i-1}+1,j}$	...	$a_{r_{i-1}+1,N}$		Ω_i
$\omega_{r_{i-1}+2}$	$a_{r_{i-1}+2,1}$	$a_{r_{i-1}+2,2}$	...	$a_{r_{i-1}+2,j}$	...	$a_{r_{i-1}+2,N}$		
....	....	....	...	....	...	....		
ω_{r_i}	$a_{r_i,1}$	$a_{r_i,2}$	...	$a_{r_i,j}$	...	$a_{r_i,N}$		
$\omega_{r_{m-1}+1}$	$a_{r_{m-1}+1,1}$	$a_{r_{m-1}+1,2}$	...	$a_{r_{m-1}+1,j}$	...	$a_{r_{m-1}+1,N}$		Ω_m
$\omega_{r_{m-1}+2}$	$a_{r_{m-1}+2,1}$	$a_{r_{m-1}+2,2}$	...	$a_{r_{m-1}+2,j}$	...	$a_{r_{m-1}+2,N}$		
....	....	....	...	....	...	....		
ω_{r_m}	$a_{r_m,1}$	$a_{r_m,2}$	...	$a_{r_m,j}$	...	$a_{r_m,N}$		
ω'	b_1	b_2	...	b_j	...	b_N		$\Omega_?$

is not defined completely. Only some information I_0 about the classes Ω_i is specified. The objects are defined by values of some features x_j, $j = 1, ..., N$ (this collection is always the same for all objects considered in the solution of a certain problem). The collection of values of the features x_j determines the description $I(\omega)$ of the object ω. Each feature can assume values from different sets of acceptable values of features. For example, from the following: $\{0, 1\}$ - the feature is not satisfied or satisfied, respectively; $\{0, 1, \Delta\}$, Δ - no information about the feature; $\{0, 1, ..., d - 1\}$ - the degree of distinctness of the features has different grades, $d > 2$; $\{a_1, ..., a_d\}$ - the feature has a finite number of values, or $d > 2$; $[a, b]$, $(a, b]$, $[a, b)$, a, b are arbitrary numbers or symbols, $-\infty$, $+\infty$. The values of the feature x_j are functions of a category, and those of x_j are functions of distribution of a random variable. The description of the object $I(\omega) = (x_1(\omega), ..., x_N(\omega))$ is called standard if $x_j(\omega)$ assumes a value from the set of acceptable values.

The recognition problem with standard information is to compute values of predicates $P_i(\omega)$ – "$\omega \in \Omega_i$", $i = 1, ..., m$ from learning information I_0 Ω_1, ..., Ω_m about categories and the description $I(\omega)$ for a given object ω and a collection of classes $\Omega_1, ..., \Omega_m$. The information whether ω is contained in Ω_i is coded by symbols "1"($\omega \in \Omega_i$), "0"($\omega \bar{\in} \Omega_i$), while Δ means that it is unknown whether ω belongs to the category Ω_i or not. It is written in the form of a so-called "information vector" as follows

$$\tilde{\alpha}(\omega) = (\alpha_1(\omega), ..., \alpha_m(\omega)) \quad \alpha_i \in \{0, 1, \Delta\} \tag{2.1}$$

The standard information

$$I_0(\Omega_1, ..., \Omega_m) \tag{2.2}$$

is a collection of sets $(I(\omega_1), ..., I(\omega_{r_m}))$ and $(\tilde{\alpha}(\omega_1), ..., \tilde{\alpha}(\omega_{r_m}))$ (it is assumed that there are no vectors of the form $(\Delta, ..., \Delta)$ among information vectors). The a-priori information in a recognition problem with disjoint classes is often specified in the form of a so-called "training table" $T_{N,m}$ (see Table 2.1). Obviously, the objects $\omega_1, \ldots, \omega_{r_1}$ belong to the class Ω_1, $\omega_{r_{i-1}}, \ldots, \omega_n$ to Ω_1, and $\omega_{r_{m-1}+1}, \ldots, \omega_{r_m}$ to Ω_m.

2.3. The synthesis of a model of a heuristic recognition algorithm

As mentioned above, the analysis of a collection of inaccurate algorithms can reveal their formation principles while the formalization of the latter can produce mathematical models of heuristic recognition algorithms. We illustrate the construction of such a model taking as an example the formalization of the partitioning principle which consists in that in many problems where descriptions of objects are specified by collections of values of numerical attributes (the objects are points in n-dimensional space), such descriptions that belong to different categories can be divided by surfaces of a quite simple form.

Consider one possible formalization. Let us use the simplest class of partitioning surfaces called hyperplanes:

$$\sum_{i=1}^{n} a_i x_i + a_{n+1} = 0 \qquad 2.3$$

Let a set of acceptable objects be divided into two categories: K_1, K_2, $K_1 \cap K_2 = \varnothing$. Let it be also known that the objects $S_1, \ldots, S_m$ belong to K_1 and $S_{m+1}, \ldots, S_q$ to K_2. Generally speaking, these objects are not equivalent. So we introduce their numerical characteristics $\gamma(S_i) = \gamma_i$, and the weight of the object S_i, $i = 1, 2, \ldots, m, m+1, \ldots, q$. Thus the set of algorithms is characterized by specifying the parameters $a_1, \ldots, a_{n+1}$, the coefficients in the hyperplane equation, and $\gamma_1, \ldots, \gamma_q$, weights of the objects whose classification was done earlier. The recognition process for $I(S) = (\alpha_1, \ldots, \alpha_n)$ is carried out as follows.

Let

$$f(x_1, \ldots, x_n) = \sum_{i=1}^{n} a_i x_i + a_{n+1}. \qquad 2.4$$

Divide the objects $S_1, \ldots, S_m$ into sets K_1^+, K_1: $S_i \in K_1^+$, if $f(I(S_i)) \geq 0$ and $S_i \in K_1^-$ if $f(I(S_i)) < 0$. Similarly, divide the objects $S_{m+1}, \ldots, S_q$ into sets K_2^+, K_2. Consider the quantities

$$\gamma(K_I^+) = \sum_{S_i \in K_1^-} \gamma(S_i), \; \gamma(K_1^-) = \sum_{S_i \in K_1^-} \gamma(S_j) \qquad 2.5$$

and $\gamma(K_2^+)$, $\gamma(K_2^-)$ similar to them.

Compute $f(I(S))$. Compare two numbers with S: $\Gamma_1(S)$ and $\Gamma_2(S)$ that correspond to values of the function belonging to classes K_1 and K_2, respectively.

If $f(I(S)) \geq 0$, then

$$\Gamma_1(S) = \frac{\gamma(K_1^+) + \gamma(K_2^-)}{\gamma(K_1^-) + \gamma(K_2^+)}, \qquad \Gamma_2(S) = \frac{\gamma(K_2^+) + \gamma(K_1^-)}{\gamma(K_1^+) + \gamma(K_2^-)}. \qquad 2.6$$

For $f(I(S)) < 0$:

$$\Gamma_1(S) = \frac{\gamma(K_1^-) + \gamma(K_2^+)}{\gamma(K_1^+) + \gamma(K_2^-)} \qquad 2.7$$

and $\Gamma_2(S)$ can be computed in the same way.

A decision is taken on the basis of values of $\Gamma_1(S)$ and $\Gamma_2(S)$ whether S can be assigned to K_1 or K_2. This procedure is specified by the decision rule. Consider a class of decision rules determined by the parameter:

if $\Gamma_1(S) - \Gamma_2(S) > \delta$, then $S \in K_1$

if $\Gamma_2(S) - \Gamma_1(S) > \delta$, then $S \in K_2$ 2.8

if $|\Gamma_1(S) - \Gamma_2(S)| \leq \delta$, then the decision cannot be taken and the algorithm refuses to classify S.

Thus, one of the possible models based on the partitioning principle was constructed. It is based on two hypotheses: a) elements of the categories K_1 and K_2 are partitioned by a hyperplane (at least a considerable number of the elements whose classification we are interested in) and b) the elements of the categories are not equivalent in importance and the measure of the importance can be expressed by a number.

The hypotheses were realized in building the models

$$M(a_1, \ldots, a_{n+1}, \gamma_1, \ldots, \gamma_q, \delta), \; -\infty < \gamma_i, \; a_i < +\infty, \; \delta \geq 0. \qquad 2.9$$

All the parameters of a model are specified by its element - a particular recognition algorithm.

2.4. The synthesis of the extreme in the model of a recognition algorithm

While the main difficulty at the level of creating an individual fixed algorithm is how to build effective computational schemes and conduct experiments, the level of models involves many new mathematical problems. Particularly noteworthy among them are problems of the synthesis of algorithms which are extreme by the quality of recognition within the framework of the given model. The quality functional of an algorithm can be determined by different methods. It is usually determined on the basis of the following principle. A method for construction of objects in each category is specified. An estimate is made for a fixed algorithm from the given model of what part of the objects it can classify correctly, i.e. assign to a given category. The value thus produced is averaged by categories and is called the quality functional of the algorithm. The objective is to find an algorithm with a maximum value of the quality functional. For example, the following law of generating the classes K_1 and K_2 can be specified.

Let the description $I(S)$ of the objects S comprise the collections $(a_1(S), ..., a_n(S))$ of numerical features $-\infty < a_i(S) < +\infty$, $i = 1, 2, ..., n$.

Two normal distributions are set in n-dimensional space with the mathematical expectations m_1, m_2 and dispersions σ_1, σ_2, respectively. Points (descriptions of objects) are selected at random and the category where they are assigned is played out according to the specified laws. After that, the object S assigned, for example, to K_1 with the probability p is included in a learning set and that with probability $1 - p$ in a check sample. The same is done with K_2. Let the learning set and the check sample be formed this way. The former contains objects $S_{11}, ..., S_{1m}$ in K_1 and $S_{21}, ..., S_{2t}$ in K_2, the latter $S_{31}, ..., S_{3v}$ in K_1 and $S_{41}, ..., S_{4u}$ in K_2. An algorithm A is compiled in the model. It gives the maximum value of the quality functional $\varphi(A) = q'/q''$ from the descriptions $I(S_{11}), ..., I(S_{1m})$ and $I(S_{21}), ..., I(S_{2t})$, where q' is the number of objects from the check sample of correctly classified algorithms A and $q'' = v + u$ is the number of objects in the check sample.

$\varphi(A)$ is a random value and its characteristics (moments) give an idea of the accuracy of the model in a certain type of recognition problem. The computation of the characteristics is not at all a trivial matter. The results in such problems can be obtained only for relatively simple models and laws of formation of categories.

More standard is the approach where, for fixed input information I_0 and a model, we must find an algorithm in the model, which can classify the given union S_i, $i = 1, 2, ..., m$ of test objects with a maximum accuracy when it is known whether they belong to the categories $K_1, ..., K_l$.

Naturally, the information of the type $S_i \bar{\in} K_j$, $S_i \in K_j$ is not introduced into the algorithm. The construction of extreme algorithms in a model on a specified check sample leads to a solution and study of new types of extreme problems. A great number of papers is devoted to such studies, particularly in R models and V models. Suppose we have object descriptions $I(S_1), ..., I(S_m)$ in K_1, $I(S_{m+1}), ..., I(S_q)$ in K_2, and $I(S_i) = \alpha_{i1}, ..., \alpha_{il})$. No input information. An R model is constructed and the separation is done with a hyperplane $f(\dot{x}) = \sum_{i=1}^{n} a_i x_i + a_{n+1}$. The parameters of the model are coefficients of the hyperplane $a_1, ..., a_{n+1}$.

The decision rule. If $f(I(S_i)) \geq 0$, then $S_i \in K_1$, for $f(I(S_i)) < 0$ the object S_i is included into K_2, $i = 1, 2, ..., q$.

It is easy to write the condition of correct classification for each S_i. Having written the conditions successively, $S_1, ..., S_m$ and $S_{m+1}, ..., S_q$, we obtain a system of linear inequalities with the unknown quantities $a_1, ..., a_{n+1}$:

$$
\begin{aligned}
&a_1\alpha_{11} + \ldots + a_n\alpha_{1n} + a_{n+1} \geq 0 \\
&\ldots \\
&a_1\alpha_{m,1} + \ldots + a_n\alpha_{m,n} + a_{n+1} \geq 0 \\
&\ldots \\
&a_1\alpha_{m+1,1} + \ldots + a_n\alpha_{m+1,n} + a_{n+1} < 0 \\
&\ldots \\
&a_1\alpha_{q1} + \ldots + a_n\alpha_{qn} + a_{n+1} < 0.
\end{aligned}
\qquad 2.10
$$

The system (2.10) is incompatible, generally speaking. To construct the desired algorithm we must find a maximum joint subsystem in (2.10). By solving it we get the values $a_1, ..., a_{n+1}$ and so an algorithm which is extreme on the sample $S_1, ..., S_q$.

The isolation of a maximum joint subsystem is a difficult problem even for linear systems and special methods are required to solve it. The construction of extreme algorithms on a given sample in more complex models can be also reduced to searching for maximum joint subsystems but the inequalities in the analogs of (2.10) are not linear.

2.5. The AVO model: recognition algorithms based on computation of estimates

First recognition algorithms of the estimate computation class (abbreviated in Russian as AVO) were proposed in the 1960s [14] and were studied in detail and developed during the 1970s [6, 12, 16, 18, 20, 23]. A detailed bibliography on the AVO can be found in Refs. [6, 18, 20]. The algorithms of this class are based on a very natural heuristic principle which is used frequently and readily. It is the principle of precedence or partial precedence, i.e. taking decisions by analogy, namely that one should act the same in the same (or at least similar) situations.

AVO computes priorities (similarity estimates) that characterize the "closeness" of the object(s) of recognition to the prototype(s) by the system of ensembles of features, which are a system of subsets of a specified set of features.

The importance of the AVO class is due to a number of circumstances:

1. This class of algorithms was a proving ground for the mathematical theory of recognition known today as the "algebraic approach to the solution of problems of recognition and classification" and concerned with validation of algorithms for the solution of ill-formalized problems involved in data processing and analysis [20].

2. As applied to this class, the concept of the "model of a recognition algorithm" was first formulated and appropriate models were actually built and used to solve practical problems.

3. The construction of the model of the recognition algorithm made possible systematic selection of an algorithm by formalizing to solve a specific problem. This selection is equivalent to isolation of an algorithm

A out of a certain known set of recognition algorithms $\{A\}$, such that a functional φ_A characterizing the quality of A becomes a maximum. It is well-known that the classical optimization problem consists in finding an extreme of a functional when there are restrictions to the extent of variables. And both the functional and the variables can be static and dynamic. The result is an extreme problem for a special functional in a certain space selected in some special way. It is in this sense that the synthesis of a recognition algorithm is described as a solution of a discrete extreme problem.

The AVO class uses the recognition error that arises in application of a specific algorithm as a functional set in the space of recognition algorithms. Thus there is a need to define more accurately the concept of the recognition error. The problems of recognition and classification rather often have an important feature: the objects whose categories are known make up only a very small part of the number of objects which are hypothetically possible ("acceptable"). Obviously, what remains in these situations is to compute the error or some "generalized" error for all objects which are known to belong to certain categories and to minimize exactly that functional. On the other hand, it is a well-established fact that when limited "experimental data" are used, the transfer to other experimental data may result in a radical change in the error functional. The alternative for practical applications of recognition is that either the recognition problem with limited data to be based on cannot be solved in principle, or an error functional must be used for the data. The difficulty can be overcome if we find criteria of stability of the statistical sample by the error functional (some aspects of the issue are discussed in a monograph by V.N. Vapnik [4]).

True, there are also problems that are solved by formulating a priori a hypothesis about different objects belonging to appropriate categories, whereupon the hypothesis is checked using known statistical data. Unfortunately, in the most common recognition applications, such as medical diagnostics, geological prediction or sociological studies, researchers usually fail to formulate the desired reasonable hypotheses due to low formalization of the object domain (and sometimes hypotheses require so much input information and data of such kind that their availability make the solution of the given recognition problem unnecessary.)

4. Since optimization procedures are realized in a specially selected space, for a problem of the synthesis of an optimum algorithm to be solved we must determine the class of algorithms to be optimized so that the optimization can truly be carried out. That means in effect that we need, first, a model which could represent a sufficiently broad class of recognition algorithms and, second, this model should be specified by a number of parametric objects characterizing the given class of recognition algorithms. This model can set up a mutually unambiguous correspondence between the algorithms and the collection of numerical parameters (the class of algorithms being a region in a multidimensional space). Thus, the specification of an actual algorithm A that belongs to the class of recognition algorithms $\{A\}$ under consideration makes it possible to associate it with the value of the generalized-error functional φ_A, and so to determine the functional which estimates the quality of the recognition algorithm at the points of the parametric space.

Third, the values of this functional should be computed efficiently. The algorithm's parametric model being specified, it becomes possible to select and vary the parameters so that the value of the functional changes in the right direction. If there exists an efficient way of computing the functional φ_A, it is possible in principle to construct the algorithm A^* where the extremum φ_A is achieved. Although it is well-known that the latter problem cannot be solved completely in all cases (we have to confine ourselves to local extrema in some situations), when an absolutely extreme algorithm can be found, it is guaranteed that there is no better recognition algorithm than A^* for the given input data in the given category and with the given verification data [16, 20]. Experience with solving recognition problems has shown that frequently main "discriminating" information is not contained in individual features but in various combinations thereof. Attempts were made as early as the mid-1960s to build recognition algorithms that could take into account the information contained in combinations of features. The most famous among algorithms of this type were so-called "test algorithms" (first described in [14]), the Geometry program (Kora unit) [2] and Kora algorithm [3]). Computational complexity (the size of search) confined the two latter algorithms to conjunctions of complexity 3. Test algorithms encountered difficulties involved in the synthesis of the set of all deadlock tests (see below) for the given problem. The AVO class brings the idea of using feature aggregates to its logical end; since it is not always known exactly what combinations of features are the most informative, the degree of similarity is computed in the AVO while comparing all possible (or certain - where the combinations of features of the greatest discriminating power are known) combinations of features in the description of an object. As mentioned above, to estimate the closeness of objects the AVO provides simple analytical formulas which eliminate the need for exhaustive search during realization of the recognition procedure (to be exact, when the parameters of an algorithm are adjusted to the problem in the process of learning.) In addition, the AVO allows one to take into account differences in the information content (discriminating power) of individual features and their combinations, and differences in representativeness of individual objects included in the training table.

5. An important difference between the AVO and other classes of recognition algorithms is considerably less stringent demands upon input information since it does not require knowledge of moments and other statistical characteristics. The input data can be presented both in numerical form or specified by descriptions in a natural language.

6. The AVO class enables the solution of not only static but also dynamic problems which are often prediction problems.

A recognition algorithm based on the principle of precedence or partial precedence compares descriptions of an object of recognition - $I(\omega')$ with $T_{N,m}$ - and decides which category (class) the object should be assigned to. The decision is made by computing the degree of resemblance of the object (row) with the rows whose categories are known.

Let a standard description of objects $\{\omega^{\Omega_i}\}$, $\omega^{\Omega_i} \in \Omega_i$ and $\{\omega^{\bar{\Omega}_i}\}$, $\omega^{\bar{\Omega}_i} \in \bar{\Omega}_i$ be specified. It is necessary to determine whether the object ω' presented for recognition belongs to the category Ω_i, $i = 1, ..., m$. If there is a method for determining the closeness for some parts of the description $J(\omega')$ and corresponding parts of the descriptions $\{J(\omega^{\Omega_i})\}$ and $\{J(\omega^{\bar{\Omega}_i})\}$, then a "generalized closeness" can be formulated between ω' and the set of the objects $\{\omega^{\Omega_i}\}$ and $\{\omega^{\bar{\Omega}_i}\}$, respectively. In the simplest case, the generalized closeness is equated to the sum of closeness between parts of the descriptions. As a result, the characteristic of the form $\Gamma_i(\omega') = \Gamma^{\Omega_i}_i - \Gamma^{\bar{\Omega}_i}_i$, where $\Gamma^{\Omega_i}_i$ and $\Gamma^{\bar{\Omega}_i}_i$ are values of respective generalized closeness, can be naturally assumed to be the value of the function of ω' belonging to the category Ω_i. $\Gamma_i(\omega')$ is called an estimate of the object ω' by the category Ω_i (sometimes we will denote it by $\Gamma(\omega', \Omega_i)$.

The descriptions of the objects $\{\omega'\}$ presented for recognition are translated into a numerical matrix $\{\Gamma_i\}_{\{\omega'\}\times m}$ - the estimate matrix - by the recognition algorithm. The procedure includes two stages. First, the estimate ω' is calculated for each row in $T_{N,m}$ and then the estimates thus obtained are used to produce total estimates for each of the categories Ω_i. The application of the decision rule to the estimate matrix produces the matrix $\{\alpha_i\}_{\{\omega'\}\times m}$ of the information vectors of the objects $\{\omega'\}$.

Consider the procedure of estimating $\Gamma_i(\omega')$ used in test algorithms and AVO.

Test algorithms are based on the concept of tests [32]. The test of a table $T_{N,m}$ is a set of columns $x_{t_1}, ..., x_{t_q}$, such that after all the columns except those numbered $t_1, ..., t_q$ have been discarded from $T_{N,m}$, all the pairs of rows belonging to different categories in the resultant table $T_{N-q,m}$ will be different. The test $\{x_{t_1}, ..., x_{t_q}\}$ is called a deadlock if none of its parts is a test.

Let $\{T\}$ be the set of all deadlock tests of $T_{N,m}$ and $T = \{x_{t_1}, ..., x_{t_q}\} \in \{T\}$. Isolate the part $(b_{t_1}, ..., b_{t_q})$, in the description of the object of recognition $I(\omega')$, which corresponds to the features $x_{t_1}, ..., x_{t_q}$ and compare it with all partial descriptions $(a_{r_{t_1}}, ..., a_{r_{t_q}})$ of the objects $I(\omega_r)$ in the table $T_{N,m}$, $r = (r_{i-1} + 1), ..., r_i$, $i = 1, ..., m$. Calculate the number of coincidences $\Gamma_T(\omega', \Omega_i)$ of the partial descriptions $(b_{t_1}, ..., b_{t_q})$, with all partial descriptions $(a_{r_{t_1}}, ..., a_{r_{t_q}})$ of the objects of the i-th category. $\Gamma_T(\omega', \Omega_i)$ is the number of rows of this category, which are close to the row ω' being recognized by the test T, i.e. it is an estimate of ω' for Ω_i by T. The estimate for ω' by the other tests is computed in a similar fashion for all categories. The quantity

$$\Gamma(\omega', \Omega_i) = \frac{1}{r_i - r_{i-1}} \sum_{T \in \{T\}} \Gamma_T(\omega', \Omega_i) \qquad 2.11$$

is an estimate of ω' by the category Ω_i.

There are versions of test algorithms wherein the formation of the estimates $\Gamma_T(\omega', \Omega_i)$ takes into account the differences in representativeness ("importance") of individual rows in $T_{N,m}$ and of the attributes included into standard descriptions of objects. Numerical coefficients - weights of features and weights of objects - are used for this purpose. There exist quite a few methods for introduction of such weights. Most often they are specified for heuristic considerations, because of some special features of a problem to be solved, by means of an expert estimate, etc. A quite natural measure of the importance of a feature - the information weight - has been proposed [14] in the form

$$p(x_j) = \frac{r_{x_j}(N, m)}{r(N, m)}, \qquad 2.12$$

where $r(N, m)$ is the number of deadlock tests of the table $T_{N,m}$ and $r_{x_j}(N, m)$ is the number of deadlock tests of $T_{N,m}$ containing the feature x_j. The more deadlock tests the feature x_j contains the greater information weight $p(x_j)$ it has and the greater its importance in describing the objects in $T_{N,m}$.

If the weights of the attributes $p(x_1), ..., (x_N)$ and the objects in the table $T_{N,m} \gamma(\omega_1), ..., \gamma(\omega_{n_m})$ are taken into consideration, each coincidence of the partial description of the object $(b_{t_1}, ..., b_{t_q})$ of recognition with that of an object in $T_{N,m}(a_{r_{t_1}}, ..., a_{r_{t_q}})$, corresponding to a test T is estimated by

$$\Gamma_T(\omega', \omega_r) = \gamma(\omega_r)(p(x_{t_1}) + ... + p(x_{t_q})) \qquad 2.13$$
$$r = (r_{i-1} + 1), ..., r_i;\ \omega_r \in \Omega_i$$

So the estimate of ω' by the category Ω_i takes the form

$$\Gamma(\omega', \Omega_i) = \frac{1}{r_i - r_{i-1}} \sum_{T \in \{T\}} \sum_{r = r_{i-1}+1}^{r_i} \Gamma(\omega', \omega_r) \quad 2.14$$

The changeover from test algorithms to the AVO was caused by increasing the types of subsets of the set of features used to compare an object of recognition with objects in $T_{N,m}$ and by deriving efficient formulas to compute estimates $\Gamma(\omega', \Omega_i)$ for different cases of the specification of subsets of features (called support sets of the recognition algorithm in the AVO). The AVO considers two cases: the presence [16, 23] and the absence of restrictions on the system of an algorithm's support sets [12]. In the former, the most common are the systems of support sets composed of all subsets of a set of features with a fixed length q, $q = 2, \ldots, N - 1$, or of all non-empty subsets of the set of features.

Consider a full collection of the features $\langle x_1, \ldots, x_N \rangle$ and isolate the system of subsets of the set of attributes (the system of support sets of an algorithm) $S_1, \ldots, S_l$. Discard an arbitrary subcollection of features from the rows $\omega_1, \omega_2, \ldots, \omega_{r_m}, \omega'$ and denote the resultant rows by $S\tilde{\omega}_1, S\tilde{\omega}_2, \ldots, S\tilde{\omega}_{r_m}, S\tilde{\omega}'$.. The rule of closeness that can be used to estimate the similarity of $S\tilde{\omega}_r$ and $S\tilde{\omega}'$ consists in the following. Let the "truncated" rows contain q first features, i.e. $S\tilde{\omega}_r = (a_1, \ldots, a_q)$ and $S\omega' = (b_1, \ldots, b_q)$, and let the thresholds $\varepsilon_i, \ldots, \varepsilon_q, \delta$ be specified. The rows $S\tilde{\omega}_r$ and $S\omega'$ are considered to be similar if at least δ inequalities of the form $|a_j - b_j| \le \varepsilon_j, j = 1, \ldots, q$ are satisfied. The quantities $\varepsilon_1, \ldots, \varepsilon_q, \delta$ are included as parameters in the model of the class of algorithm of the AVO type.

Consider the procedure used to compute estimates by the subset S_1. It is exactly the same for other subsets. Columns are found in Table 2.1. $T_{N,m}$ to correspond to the features included in S_1. All other columns are crossed out. A check is made of the closeness of the row $S_1\tilde{\omega}'$ to the rows $S_1\tilde{\omega}_1, \ldots, S_1\tilde{\omega}_{r_1}$ that belong to the category Ω_1. The number of the rows of that category, which are close to the row $S_1\tilde{\omega}'$ being classified by the selected criterion, is denoted by $\Gamma_{S_1}(\omega', \Omega_1)$. That is an estimate of ω' for Ω_1 by the support subset S_1. Estimates for the other categories are computed in the same way; $\Gamma_{S_1}(\omega', \Omega_2), \ldots, \Gamma_{S_1}(\omega', \Omega_m)$. The application of this procedure to all other support sets of the algorithm can produce the system of estimates

$$\Gamma_{S_1}(\omega', \Omega_1), \Gamma_{S_2}(\omega', \Omega_m), \ldots, \Gamma_{S_l}(\omega', \Omega_1), \ldots, \Gamma_{S_l}(\omega', \Omega_m).$$

The quantities

$$\Gamma(\omega', \Omega_1) = \Gamma_{S_1}(\omega', \Omega_1) + \Gamma_{S_2}(\omega', \Omega_1) + \ldots + \Gamma_{S_l}(\omega', \Omega_1) = \sum_{S_A} \Gamma(\omega', \Omega_1);$$
$$\ldots \quad 2.15$$
$$\Gamma(\omega', \Omega_m) = \Gamma_{S_1}(\omega', \Omega_m) + \Gamma_{S_2}(\omega', \Omega_m) + \ldots + \Gamma_{S_l}(\omega', \Omega_m) = \sum_{S_A} \Gamma(\omega', \Omega_m)$$

are estimates of ω' for appropriate categories by the system of support sets of S_A. From the analysis of these a decision is taken whether the object ω' can be assigned to one of the categories Ω_i, $i = 1, \ldots, m$ or whether it cannot be recognized. The decision rule may take different forms. In particular, the row being recognized can be assigned to the category whose estimate is maximal, or to the one whose estimate exceeds those of all other categories by at least a certain threshold value η_1, or the ratio of an estimate to the sum of estimates for all other categories should not be less than a certain threshold η_2, etc. Such parameters as η_1 and η_2 are also included in the AVO model.

Example. A training table and an object ω' in Table 2.2 to be recognized are specified.

Table 2.2.

Objects	Features and their values						Categories
	x_1	x_2	x_3	x_4	x_5	x_6	
ω_1	0	0	0	0	0	0	
ω_2	0	1	0	0	1	1	Ω_1
ω_3	1	1	0	1	1	1	
ω_4	0	1	0	1	0	1	
ω_5	1	1	1	1	1	1	Ω_2
ω_6	1	1	0	0	0	1	
ω'	1	1	0	0	0	0	$\Omega_?$

Let $S_1 = \langle x_1, x_2 \rangle$, $S_2 = \langle x_3, x_4 \rangle$, $S_3 = \langle x_5, x_6 \rangle$. The rows are assumed to be close if they fully coincide. The above estimation procedure allows one the following:

$$S_1: \Gamma_{S_l}(\omega', \Omega_1) = 1; \; \Gamma_{S_l}(\omega', \Omega_2) = 2$$

$$S_2: \Gamma_{S_2}(\omega', \Omega_1) = 2; \; \Gamma_{S_2}(\omega', \Omega_2) = 1$$

$$S_3: \Gamma_{S_3}(\omega', \Omega_1) = 1; \; \Gamma_{S_3}(\omega', \Omega_2) = 0$$

$$\Gamma_{S_A}(\omega', \Omega_1) = \Gamma_{S_1}(\omega', \Omega_1) + \Gamma_{S_2}(\omega', \Omega_1) + \Gamma_{S_3}(\omega', \Omega_1) = 1 + 2 + 1 = 4;$$

$$\Gamma_{S_A}(\omega', \Omega_2) = \Gamma_{S_1}(\omega', \Omega_2) + \Gamma_{S_2}(\omega', \Omega_2) + \Gamma_{S_3}(\omega', \Omega_2) = 2 + 1 + 0 = 3.$$

According to the decision rule which realizes the principle of a simple majority of votes, the row ω' is assigned to the category Ω_1 since $\Gamma_{S_A}(\omega', \Omega_1) > \Gamma_{S_A}(\omega', \Omega_2)$.

The determination of the recognition part in the AVO class involves formalization of the following stages that correspond to the sequence of realization of the recognition procedure: 1) the algorithm's system of support sets is isolated and the sets are used to analyze the objects of recognition; 2) the concept of closeness on the set of partial descriptions of objects is introduced; 3) the following rules are specified: a) the rule for computing estimates of pairs of objects by the degree of similarity between the prototype and the object of recognition; b) the rule for formulating estimates for each of the prototype categories by the fixed support sets on the basis of estimates of pairs of objects; c) the rule for formulating a total estimate for each of the prototype categories by all support sets; and d) the rule of decision whether to assign the object of recognition to one of the categories or to refuse to classify this object on the basis of estimates of the categories.

The specification of structural parameters, i.e. of the method for selection of the system of support sets, of the type of closeness function, of the rules for computing estimates, and of the decision rule determines the choice of the subclass of algorithms of the AVO type while the specification of appropriate parameters determines the concrete algorithm of the AVO type. The AVO-type model is parametric, i.e. there is a mutually unambiguous correspondence between specific algorithms and collections of numerical parameters. In this case, the specification of a concrete algorithm that belongs to the class under consideration allows one to associate it with a value of a functional of the quality of recognition (e.g. the number of errors and refusals to recognize in the training table) and thus define the latter at points of the parametric space of the algorithm.

If the computational procedure is formulated from a given description of an algorithm, the computational complexity becomes great when the power of the system of support sets is large. Thus, when all the subsets of the set of attributes of power q are chosen as a system of support sets of an algorithm, the number of the support sets is C_N^q and that of summands in the formula which determines $\Gamma_{S_A}(\omega', \Omega_i)$, is $(r_i - r_{i-1})C_N^q$.

As has been mentioned above, an important advantage of the AVO is that simple analytical formulas are provided to compute the estimates that determine which of the specified categories the object of recognition belongs to. The formulas replace complex sampling procedures (that are involved in estimating closeness by the system of support sets). Since the efficiency (in computational terms) of computing the quality functional in the AVO depends entirely on the efficiency of the estimation procedure, it is possible in principle to build an optimum algorithm. Where an absolute extreme algorithm can be found, there is a guarantee that the given class does not contain a better recognition algorithm for the given input data $T_{N,m}$.

Two methods are known to find, when combined, sufficiently simple formulas for interesting AVO models in practice, provided threshold closeness functions (assuming values 0 or 1) and $p(S) = p_{x_{t_1}} + \dots + p_{x_{t_q}}$ (the weight of a support set equals the sum of weights of the attributes it contains) are used.

1. The first method [20] uses the property of estimates for the category $\Gamma(\omega', \Omega_i)$ where an estimate of the form (2.13) is used to evaluate the category by the support set S_u, $u = 1, \dots, l$ as follows:

$$\Gamma_{S_u}(\omega', \Omega_i) = \frac{1}{r_i - r_{i-1}} \sum_{\omega_r \in \Omega_i} \gamma(\omega_r)(p_{t_1} + \dots + p_{t_q}) B_{S_u}(\omega', \omega_r), \qquad 2.16$$

where $t_1, \dots, t_q$ is a collection of unit coordinates of the characteristic vector which determines the support set S_u and B_{S_u} is a function of closeness of partial descriptions of the objects $S\omega'$ and $S\omega_r$ assuming values "1" or "0" according to the number of satisfied inequalities of the form $|a_j - b_j| \le \varepsilon_j$, $j = t_1, \dots, t_q$ (see above).

Let $v_j(\omega', \omega_r)$ be the number of different values that can be assumed by the number of the support sets $S_u \in S_A$ containing a set attribute x_j, such that $B(S\omega', S\omega_r) = 1$. It can be shown that the estimate for the category Ω_i takes the form

$$\Gamma(\omega', \Omega_i) = \frac{1}{r_i - r_{i-1}} \sum_{\omega_r \in \Omega_i} \gamma(\omega_r) \times \sum_{j=1}^{N} p_j v_j(\omega', \omega_r). \qquad 2.17$$

If the number of different values $v_j(\omega', \omega_r)$ is small, the inner sum is contracted to a small number of summands (summing by the system of support sets is practically impossible) and the complexity of computing $\Gamma(\omega', \Omega_i)$ becomes proportional to the length of the learning set. It was proved, in particular, that when a threshold function is used, its value depending only on the number of satisfied and unsatisfied inequalities of the form $|a_j - b_j| \le \varepsilon_j$, the quantities $v_j(\omega', \omega_r)$, $j = 1, \dots, N$ do not assume more than two values [20].

Here are analytical formulas which are quite effective in estimating $\Gamma(\omega', \Omega_i)$ for two methods of restricting the system of support sets of an algorithm [23]:

a) S_A coincides with the system of all subsets of power q of the set of attributes $\{1, \dots, N\}$.

1° the closeness function assumes the value "1" if at least δ (threshold) inequalities of the form $|a_{t_u} - b_{t_u}| \le \varepsilon_{t_u}$:

$$\Gamma(\omega', \Omega_i) = \frac{1}{r_i - r_{i-1}} \sum_{\omega_r \in \Omega_i} \gamma(\omega_r) \times \left(\sum_{j=0}^{\delta} C_{z(\omega', \omega_r)}^{q-j} C_{N-z(\omega', \omega_r)}^{j} \right), \quad 2.18$$

are satisfied.

Here $z(\omega', \omega_r)$ is the number of satisfied inequalities of the form $|a_j - b_j| \le \varepsilon_j$ for the pair (ω', ω_r) $j = 1, ..., N$.

2° the closeness function assumes the value "1" if all inequalities of the form $|a_{t_u} - b_{t_u}| \le \varepsilon_{t_u}$ $(\delta = 0)$ are satisfied, i.e. the support set includes only the features which coincide by the threshold ε_{t_u}:

$$\Gamma(\omega', \Omega_i) = \frac{1}{r_i - r_{i-1}} \sum_{\omega_r \in \Omega_i} \gamma(\omega_r) C_{z(\omega', \omega_r)}^{q} \quad 2.19$$

b) S_A coincides with the system of all non-empty subsets of the set of attributes $\{1, ..., N\}$:

1° the closeness function is the same as in a) 1°:

$$\Gamma(\omega', \Omega_i) = \frac{1}{r_i - r_{i-1}} \sum_{\omega_r \in \Omega_i} \gamma(\omega_r) \times \left[(2^{z(\omega', \omega_r)} - 1) \sum_{j=0}^{\delta} C_{N-z(\omega', \omega_r)}^{j} \right]; \quad 2.20$$

2° the closeness function is the same as in a) 2°:

$$\Gamma(\omega', \Omega_i) = \frac{1}{r_i - r_{i-1}} \sum_{\omega' \in \Omega_i} \gamma(\omega_r) \times (2^{z(\omega', \omega_r)} - 1). \quad 2.21$$

2. The second method that can be used to produce efficient estimation formulas [12] is based on the following two statements:

a) if a system of support sets S_A consists of disjoint subsets, then the estimate $\Gamma(\omega', \Omega_i)$ for this algorithm is equal to the sum of estimates for the algorithms whose support sets are the subsets which form S_A (a separate algorithm corresponds to each subset):

$$S_A = \bigcup_u S_u; \; \Gamma_A(\omega', \Omega_i) = \sum_u \Gamma^{A_u}(\omega', \Omega_i); \quad 2.22$$

b) if the characteristic function of an AVO algorithm is an elementary conjunction, then an efficient estimation formula can be derived by the above method 1.

The practice of recognition has shown that subcollections of features are known a priori in some cases and they should be taken into account in comparing the object of recognition with objects in the training table. These subsets of features do not always coincide with partial cases; the subsets may be of different lengths, prohibited or "excessive" combinations may be specified, etc. Ref. [12] has obtained analytical estimation formulas for the case of arbitrary support sets.

The formulas were derived by introducing the characteristic Boolean function of the system of support sets of the algorithm f_{S_A} and by establishing a mutually unambiguous correspondence between subsets of the set of features and Boolean length vectors (vertices of an n-dimensional unit cube).

Example. We consider the same training table and the object of recognition as in the previous example but confine ourselves to only the first four features. Having coded the presence of an feature in a support set by "1" and the absence by "0", we can associate a binary vector or, which is the same, a vertex of a four-dimensional unit cube with each subset of the set of features $\langle x_1, x_2, x_3, x_4 \rangle$. The characteristic Boolean function can be determined on the set of these vectors, the units of the function determining the subsets of features included into the system of support sets of the algorithm S_A.

Let $S_A = \{S_1, S_2\}$, $S_1 = \langle x_2, x_3 \rangle$ (vertex), $S_2 = \langle x_1, x_2, x_3 \rangle$. In this case,

$$f_{S_A} = \bar{x}_1 x_2 x_3 \bar{x}_4 \mathrm{V} x_1 x_2 x_3 \bar{x}_4 = x_2 x_3 \bar{x}_4.$$

It was shown in [12] that where the set of the units f_{S_A} forms an interval or the sum of disjoint intervals in an N-dimensional unit cube, efficient estimation formulas can be derived. Remember that the subset of the vertices of an N-dimensional unit cube are called intervals if it corresponds to an elementary conjunction. Obviously, all the faces, edges and vertices of the N-dimensional unit cube are intervals.

The system of support sets is organized as follows (the appropriate interval is represented by an edge that connects the vertices). It includes all features contained in the disjunctive normal form (DNF) of the characteristic function without negation (x_2 and x_3) and does not include the features contained in the DNF with negation (x_4). There is a full variation for other features (x_1), i.e. consideration is given to all subsets both including and not including these features.

When an interval corresponds to the characteristic function of the system of support sets and the closeness function assumes that $\delta = 0$, the efficient estimation formula has the form

$$\Gamma(\omega', \Omega_i) = \frac{1}{r_i - r_{i-1}} \sum_{\omega_r^* \in \Omega_i} \gamma(\omega_r^*) 2^{z^*(\omega', \omega_r^*)}, \quad 2.23$$

where ω_r^* is an object of the training table, which is "efficient" for ω'.

Eq. (2.23) considers the contribution of only those objects of $T_{N,m}$ ("efficient" ones) whose constant part is close (in the sense of the closeness function) to that of ω'; $z^*(\omega', \omega_r^*)$ is the number of satisfied inequalities of the form $|a_j - b_j| \le \varepsilon_j$ on the variable part.

Thus, on the condition that $\varepsilon_1, ..., \varepsilon_4 = 0$ and $\gamma_1, ..., \gamma_6 = 1$ and considering that the objects ω_2 and ω_3 are efficient for ω' in Ω_1, and ω_4 and ω_6 are efficient in

Ω_2, $z^*(\omega', \omega_2) = 0$, $z^*(\omega', \omega_3) = 1$, $z^*(\omega', \omega_4) = 0$, $z^*(\omega', \omega_6) = 1$, we obtain $\Gamma(\omega', \Omega_1) = (1/3)(1\cdot 2^0 + 1\cdot 2^1) = 1$ and $\Gamma(\omega', \Omega_2) = (1/3)(1\cdot 2^0 + 1\cdot 2^1) = 1$. The result means that ω' cannot be classified for this choice of system of support sets.

If the sum of disjoint intervals (represented by the orthogonal DNF) corresponds to the characteristic function, as, for example, in the cases $S_A = \{S_1, S_2, S_3, S_4, S_5\}$, $S_1 = <x_2, x_3>$, $S_2 = <x_1, x_2, x_3>$, $S_3 = <x_1, x_3, x_4>$, $S_4 = <x_1, x_3>$, $S_5 = <x_1, x_2, x_4>$, $f_{S_A} = x_2x_3\bar{x}_4 \vee x_1\bar{x}_2x_3 \vee x_1\bar{x}_3x_4$ and when the formula (2.23) is used to compute the estimates, it is applied to each interval separately whereupon the results are added together.

Ref. [12] demonstrated that the complexity of the AVO estimation formulas for arbitrary S_A is proportional to that of the DNF representing the characteristic function of the system of support sets of the algorithm. That means that the derivation of the simple estimation formula $\Gamma(\omega', \Omega_i)$ involves minimization of Boolean functions in the DNF class or, to be exact, construction of the shortest orthogonal DNF, or the DNF where each interval has the greatest number of intersections with neighbors. In the general case, the problem of this synthesis in insolvable and so use should be made of approximate algorithms that can produce "sufficiently simple" orthogonal DNF or DNF with a small number of intersections of intervals [12, 17].

Thus, if there exists an efficient algorithm to compute distances $\rho_j(a_j, b_j)$, $j = 1, \ldots, N$ and the number of operations for one such computation does not exceed a certain value Q, then the number of operations to compute all the values of $\Gamma(\omega', \Omega_i)$, $i = 1, \ldots, m$ does not exceed $2QNm$. The number of operations for recognition of one object in a fixed AVO algorithm is proportional to the "area" of the table $T_{N,m}$, the factor of proportionality not exceeding $2Q$ (see Table 2.1). The justification for reducing the problem of constructing extreme AVO algorithms to finding extrema of the function of many variables was substantiated in [16]. The optimization can be done by methods of exhaustion (for a small number of parameters), those of gradient type or random search.

Note that there are other methods to specify the system of support sets. Thus, studies have been made lately of an AVO model with support sets specified in $T_{N,m}$ by local neighborhoods of low orders, so-called two-index support sets, and a DAVO model (AVO with two-dimensional support sets) has been proposed [10, 11].

The AVO class is used to advantage to solve problems in medical diagnostics, geological prediction, sociological data processing, identification and control of production processes, optimum selection of an algorithm, automatic processing of experimental data, etc. The algorithms of this class can solve recognition problems of all basic types, including assignment of an object to a specified category, automatic classification, and the selection of a system of features to describe objects of recognition and to estimate their information content. In addition, the AVO model served as a basis for determining the first parametric class of algorithms of image recognition [10, 11].

2.6. Recognition where input data are represented as long sequences[2]

Long sequences are one of the most universal forms of information representation in recognition problems. Since the vector of features of any object can be described by a binary sequence, this form is common in a sense. The specific feature of long sequences is that their great volume of information makes them practically impossible to process as a whole. So descriptions more compact but sufficient for recognition have to be found taking into account additional difficulties caused in practical problems by incomplete data and the influence of noise. Problems involved in construction of such sets of features have been widely discussed in the literature for the last few years.

The most natural way of constructing simple and effective sets of features to describe and recognize long sequences consists in computing frequency characteristics to be compared with prototypes. With this approach, a certain collection of vectors only small dimensions must be specified, the frequencies of occurrences of vectors from this collection in the sequence as a subsequence must be computed, and the resultant frequencies assumed to be features describing the object. A great advantage of these features is that they are very easy to compute and require a small memory space for storage and processing. In addition, the features eliminate difficulties faced as a rule in practical problems. The low noise that distorts the form of the sequence little affects the frequencies, the displacement of the starting point for discretization and for feature count is not very important either, and the choice of the order of scanning of multidimensional objects, which changes substantially the form of the sequence, can be neutralized to some extent by a proper selection of characteristic configurations.

The following example can be given to illustrate the latter statement. Let there be a binary image of $n \times n$ in size, which is scanned from left to right and from top to bottom. For a point with coordinates (i, j) numbered $k = ni + j$ in the sequence, the horizontal neighbors are numbered $k - 1$ and $k + 1$ while vertical ones $k - n$ and $k + n$. Disregarding boundary effects, we can say that the number of subwords of the form "11" is the number of occurrences of two unities in succession horizontally. To determine a similar characteristic vertically we must calculate the number of the vectors "11" in the collection of pairs (a_1, a_{n+1}), (a_2, a_{n+2}), ..., $(a_{n\cdot n-n}, a_{n\cdot n})$, i.e. in the collection of subsequences of length 2, where the number of the second element differs from that of the first by n. By calculating the number of vectors "11" in

[2] Section 2.6 was written jointly with Yu.G. Smetanin.

the collection of pairs $(a_1, a_2), (a_2, a_3), ..., (a_{n\cdot n-1}, a_{n\cdot n}), (a_1, a_{n+1}), (a_2, a_{n+2}), ..., (a_{n\cdot n-n}, a_{n\cdot n})$ we get the number of neighborhoods of unities horizontally and vertically, i.e. a characteristic which is invariant with respect to a turn of the image through 90°.

The difficulties involved in the use of such frequency features include above all the absence of clear-cut criteria for relevancy and of methods for the selection of attributes.

It is sometimes possible in practical problems to find heuristic procedures for the selection of a small number of features to recognize long sequences. These features generally have the following form. Let $a = (a_1, ..., a_N)$ be a sequence and $r_1, r_2, ..., r_{k-1}$ be nonnegative integers. The subsequence $a(i; r_1, ..., r_{k-1})$ with delays $r_1, ..., r_{k-1}$ is referred to as a subsequence of the form $(a_i, a_{i+r_1}, ..., a_{i+r_1+...+r_{k-1}})$. If some $r_d = 0$, the element $a_{i+r_1+...+r_{d-1}+r_d}$ occurs in the subsequence only once, not twice. Obviously, $a(i; r_1, ..., r_{k-1}) \in E_{k-v}$, where v is the number of zeroes in the collection $(r_1, ..., r_{k-1})$ and E_l is a binary l-dimensional cube. The symbol $t_\gamma(a; r_1, r_2, ..., r_{k-1})$ will denote for each vector $\gamma \in E_{k-v}$ the number of occurrences of the vector γ in the sequence $\{a(i; r_1, ..., r_{k-1})\}$, $i = 1, 2, ..., N-(r_1+ ... + r_{k-1})$. The quantities $(r_1, ..., r_{k-1})$ are selected as features for recognition and $t_\gamma(a; r_1, ..., r_{k-1})$, $0 < k < N$, $\gamma \in E_{k-v}$ as values of these features.

The recognition can be carried out on the basis of a single feature defined as follows. The method of successive search (from shorter to longer delays) is used to find the value r_1 for which $t_{\alpha\beta}(a^i; r_1) - t_{\alpha\beta}(b^j; r_1)$, where a^i are elements of the category A and b^j are elements of the category B, assumes a value exceeding a specified threshold for some $\alpha, \beta \in \{0,1\}$. The value of r_2 for which $t_{\bar{\beta}\gamma}(b;r_2) - t_{\bar{\beta}\gamma}(a^r;r_2), \beta, \gamma \in \{0, 1\}$ is large enough can be found in the same way. The only feature for recognition is $t_{\alpha\beta\gamma}(a; r_1, r_2)$. Since $t_{\beta\gamma}(x; r_2)$ decreases as $t_{\bar{\beta}\gamma}(x;r_2)$ increases, this junction ensures better distinction of $t_{\alpha\beta\gamma}(a^i; r_1, r_2)$ from $t_{\alpha\beta\gamma}(b; r_1, r_2)$ than that of each of the "individual" features, and the "unified" feature works better.

The problem of reconstruction of sequences from frequencies of subsequences with fixed delays is closely connected with the problem of reconstruction of words from fragments, which arises in the theory of coding [25]. In particular, the sufficiency of the single feature for dichotomy follows from the fact that for any vector $a, b \in E_n$ there exists a delay r, $0 \le r \le [N/2]$ and the vector $\gamma \in E_2 \cup E_1$ such that $t_\gamma(a; r) \ne t_\gamma(b; r)$. This fact makes it possible to confine ourselves later on to the case of subsequences with one delay. In addition, this delay can be selected so that for any three vectors $a, b, c \in E_n$ there is a vector $\gamma \in E_2 \cup E_1$ such that

$$|t_\gamma(a; r) - t_\gamma(b; r)| \ne |t_\gamma(a; r) - t_\gamma(c; r)|. \quad 2.24$$

The distinguishability of arbitrary sequences by the frequencies of subsequences with one delay allows one to use for recognition the algebraic expansion of a set of estimation algorithms based on the use of these features.

The estimation algorithm is specified in this case as follows. Let $a_1, ..., a_m$ be a learning set of binary sequences with length N. An information vector is specified for each of these sequences to define the categories $K_1, ..., K_l$ of the sequence while $\tilde{K}_u \ne \tilde{K}_v$ for $u \ne v$, where $\tilde{K}_u = \{a_i | a_i \in K_u\}$, $\tilde{K}_v = \{a_i | a_i \in K_v\}$. Let $a^1, ..., a^q$ be the sequences to be recognized. An feature r^{ik} and a vector γ are selected for each pair a^i, a^k such that

$$t_\gamma(a^i, r^{ik}) \ne t_\gamma(a^k; r^{ik}). \quad 2.25$$

This procedure is performed for all pairs a^i, a^k. The selected attributes r^{ik} are numbered arbitrarily. We use in the work the descriptions of the learning set and of the sample to be recognized by collections of these features:

$$a_i \to S_i = (t_{i1}, ..., t_{iS}), a^p \to S^p = (t_1^p, ..., t_S^p). \quad 2.26$$

The condition (2.24) guarantees nonisomorphy of each pair of objects described in this way in the sample to be recognized in the sense of [19]. Along with the condition $\tilde{K}_u \ne \tilde{K}_v$, it ensures regularity of the recognition problem in the sense of [19]. Hence, the algebraic closure of the estimation algorithm class for the recognition problem

$$\langle S_1, ..., S_m, \tilde{K}_1, ..., \tilde{K}_l, S^1, ..., S^q \rangle$$

is correct with the vectors S being specified by the features of the form $t_\delta(\alpha; r)$, all the elements a^u being different, $a^u \ne a^v$ for $u \ne v$ and the conditions $\tilde{K}_u \ne \tilde{K}_v$ for $u \ne v$ being satisfied for the categories $K_1, ..., K_l$. This method can be used also for integer-valued sequences $a_i \in \{0, 1, ..., p\}$ without recoding them into binary form since the condition (2.24) is valid for p-value sequences. The method can be generalized directly for the case of two-dimensional arrays. The proof of a statement similar to (2.24) is practically the same. It can be shown that the number of features in the above estimation algorithm is $q - 1$ for the worst case.

2.7. Basic provisions of the algebraic approach

The algebraic approach makes substantial use of the peculiarities of the structure typical of any recognition procedure. A solution of each specific recognition problem considers the information the recognition is based on, a list of categories, and the objects $\omega', ..., \omega^q$ for which it is necessary to decide what categories on the list they belong to. Any recognition algorithm must define the vector of this object (2.1) from standard information of the form (2.2) and from the description $I(\omega)$, i.e. to compute the values of the predicates $P_i(\omega)$ – „$\omega \in \Omega_i$", $i = 1, ..., m$.

Thus, any recognition algorithm translates the recognition problem z with q objects of recognition and m

categories into a matrix of answers - an information matrix whose rows are information vectors for each of the recognition objects ω^j, $j = 1, ..., q$

$$\begin{Vmatrix} \alpha_{11}\ldots\alpha_{1i}\ldots\alpha_{1m} \\ \ldots \\ \alpha_{j1}\ldots\alpha_{ji}\ldots\alpha_{jm} \\ \ldots \\ \alpha_{q1}\ldots\alpha_{qi}\ldots\alpha_{qm} \end{Vmatrix} \qquad 2.27$$

where the element α_{ji} of the matrix assumes values 1, 0 or Δ and indicates exactly which value was computed by the recognition algorithm for the property $P_i(\omega^j)$, $i = 1, 2, ..., m, j = 1, 2, ..., q$.

Obviously, the application of different recognition algorithms to the same problem z may result in different information matrices. Naturally, therefore, a problem arises of formulating corrective methods which could compare information matrices of different algorithms to produce a single answer matrix with the least possible number of errors. It is a difficult problem since there are no easy operations with the elements "YES", "NO" and "DO NOT KNOW" ("1", "0" and "Δ") having natural properties such as associativity, commutativity, etc. (the proof can be found in [20]. So corrective operations with information matrices are simply impossible in a sense. A more thorough study of different models of recognition algorithms, however, reveals specific features of translating input learning information into an information matrix of answers.

Consider some examples.

In statistical recognition algorithms, the process starts with formation of a matrix of probabilities $\|P_{ji}\|_{q\times m}$, where P_{ji} is the probability that the j-th object belongs to the i-th category. Then the matrix of final answers is constructed from this probability matrix.

In the AVO, the process starts with construction of a matrix of estimates (votes) $\|\Gamma_i(\omega^j)\|_{q\times m}$, where $\Gamma_i(\omega^j)$ is a numerical estimate of the occurrence of the j-th object in the i-th category. Then a final decision is made on the basis of the matrix whether the object of recognition belongs to a certain category.

A similar analysis can be made for other families of recognition algorithms. It is important here that the processing of input information by a recognition algorithm into an information matrix of answers can be divided into two successive stages. The first involves translating input information into a numerical matrix of standard size with the number of rows equal to that of the objects to be recognized in the problem z, and the number of columns equal to that of the categories considered for the solution of z. At the second stage, the recognition algorithm processes this numerical matrix into that of final answers with the same number of rows and columns.

The existing division of data processing in the recognition algorithm into two stages demonstrated by the above examples is meaningful and is rigorously validated by a theorem that any recognition algorithm can be represented as two successive algorithms[20]. The first of them - algorithm B - translates the learning information and descriptions of objects of recognition into a numerical matrix of dimensions $q \times m$ and the second - algorithm C - translates the latter into a matrix of answers composed of the symbols 1, 0 and Δ, and of the same dimensions.

In theory, algorithm C can be made the same for all recognition algorithms. It is a so-called "threshold decision rule" with positive thresholds:

$$1)\ C(\|a_{ji}\|_{q\times m}) = \|C(a_{ji})\|_{q\times m} \qquad 2.28$$

It means that the decision rule is applied to the numerical matrix $\|a_{ji}\|_{q\times m}$ element by element;

$$2)\ C(a_{ji}) = \begin{cases} 1, & a_{ji} > d_2 \\ 0, & a_{ji} < d_1\ 0 < d_1 < d_2 \\ \Delta, & d_1 \le a_{ji} \le d_2 \end{cases} \qquad 2.29$$

Here d_1 and d_2 are arbitrary fixed positive numbers.

Hence the second part of the recognition algorithm is very simple and, moreover, it is standard for recognition algorithms. It implies that the main job in the information processing is done in the first part of the recognition algorithm, i.e. algorithm B.

Generally called a recognition operator, it translates input information for the given problem into a numerical matrix of standard size, i.e.

$$B(I, \Omega^m, \omega^q) = \|a_{ji}\|_{q\times m}. \qquad 2.30$$

If there are several or a whole family of recognition algorithms, we can get a collection or a family of respective recognition operators by discarding component C of the recognition algorithm. Operations of addition, multiplication and multiplication by a number can be defined for them as mappings on matrices of standard size.

Let B_1 and B_2 be recognition operators:

$$B_1(I, \Omega^m, \omega^q) = \|a^1_{ji}\|_{q\times m}$$
$$B_2(I, \Omega^m, \omega^q) = \|a^2_{ji}\|_{q\times m} \qquad 2.31$$

The recognition operator $B^+ = B_1 + B_2$ is the sum of the recognition operators B_1 and B_2 and is given by

$$B^+(I, \Omega^m, \omega^q) = \|a^1_{ji} + a^2_{ji}\|_{q\times m}. \qquad 2.32$$

The recognition operator B^+ is applied to z by applying the recognition operators B_1 and B_2 and by adding the resultant matrices element-wise.

The product $B(\cdot) = B_1B_2$ of B_1 and B_2 can be found in a similar way:

$$B(\cdot)\,(I, \Omega^m, \omega^q) = \|a^1_{ji} \cdot a^2_{ji}\|_{q \times m}. \qquad 2.33$$

Notice that the element-wise multiplication of respective matrices rather than addition of recognition operators is used to determine the product of recognition operators.

The multiplication of the recognition operator B by a number is defined as

$$(c \cdot B)\,(I, \Omega^m, \omega^q) = \|c \cdot a_{ji}\|_{q \times m}. \qquad 2.34$$

The operations of addition, multiplication and multiplication by a number thus introduced have all the properties of addition and multiplication of numbers, namely commutativity, associativity, distributivity, etc. If a family of recognition algorithms $\{A\}$ is specified, then according to the theorem of representation of the recognition algorithm, families of recognition operators $\{B\}$ and decision rules $\{C\}$ can be specified. The closure $L\{B\}$ of the operators in the family by the operations (2.32) and (2.24) is called the linear closure $\{B\}$ and the set

$$L\{A\} = L\{B\} \cdot \{C\} \qquad 2.35$$

the linear closure of the family of algorithms $\{A\}$. We represent the operators of linear closure through the operators B_g that belong to $\{B\}$, i.e.

$$a_1 B_1 + \ldots + a_g B_g + \ldots + a_k B_k, \qquad 2.36$$

where a_g are real numbers, $g = 1, \ldots, k$.

The closure $\mathfrak{A}\{B\}$ of the operators of the family $\{B\}$ with respect to the operations (2.32)-(2.34) is called algebraic closure $\{B\}$, and the set

$$\mathfrak{A}\{A\} = \mathfrak{A}\{B\} \cdot \{C\} \qquad 2.37$$

the algebraic closure of the property of algorithms $\{A\}$. In this case we can use the notation of operator polynomials, i.e. expressions of the form

$$\sum C_{g_1} \ldots C_{g_k} \cdot B_{g_1} \ldots B_{g_k}, \qquad 2.38$$

where $C_{g_1}, \ldots, C_{g_k}$ are constant values and $B_{g_1}, \ldots, B_{g_k}$ are recognition operators of the initial recognition algorithms.

The collection of operators in $\mathfrak{A}\{B\}$ that can be represented by such polynomials of a power no greater than k is called the algebraic closure $\mathfrak{A}_k\{B\}$ of the set $\{B\}$, $k = 2, 3, \ldots, v$, and the set

$$\mathfrak{A}_k\{A\} = \mathfrak{A}_k\{B\} \cdot \{C\} \qquad 2.39$$

the algebraic closure of the power k in the family $\{A\}$ of recognition algorithms.

The introduction of operations with recognition operators makes it possible to extend the original collection of recognition operators and thus of recognition algorithms. Indeed, if $L(B_1, \ldots, B_k)$ is an operator polynomial, then the expression $L(B_1, \ldots, B_k)C(d_1, d_2)$ defines already a recognition algorithm. The product symbol in this case means a successive performance of the respective parts of the algorithm and $C(d_1, d_2)$ is the threshold decision rule with the thresholds d_1 and d_2.

Thus extended, the recognition algorithms have very powerful corrective properties. It was proved [20] that even when the initial family of recognition algorithms lacks an algorithm that can solve correctly the given recognition problem z, if easily verifiable assumptions about the input learning information are satisfied and the objects of recognition are described, the desired algorithm for z exists in the extension, and can be written explicitly.

To illustrate consider an AVO collection as an initial set of inaccurate recognition procedures (heuristics). As has been mentioned above, the algorithm of this class is specified by parameters $\varepsilon_1, \ldots, \varepsilon_N$ (the threshold of accuracy for features), $p_1, \ldots, p_n$ (weights of the features), $\gamma_1, \ldots, \gamma_{rm}$ (weights of the objects in the table of learning information), and l (the number of attributes in the algorithm's support sets). The latter parameter is used to specify only some versions of the AVO.

Consider the recognition problem z with learning objects

$\omega_1, \ldots, \omega_{r_m}$, $\omega_r = (a_{r_1}, \ldots, a_{r_m})$, $r = 1, \ldots, r_m$ and categories $\Omega_1, \ldots, \Omega_m$ which may intersect. The solution z presupposes recognition of the objects $\omega^1, \ldots, \omega^q \cdot \omega^j = (b^{1N}, \ldots, b^{qN})$, $j = 1, \ldots, q$. In addition, let the method for measuring the distance x_t, $t = 1, \ldots, N$ be introduced in the set of mutual values of the feature $\rho_t(y, z)$.

It was found that when sufficiently natural conditions are satisfied, a recognition algorithm can be formulated explicitly in the algebraic extension to solve correctly the given z. The recognition operator of the algorithm has the form:

$$R_A = (d_1 + d_2) \sum_{j=1}^{q} \sum_{i=1}^{m} C_{ji} B_{ji}^{k_{ji}}. \qquad 2.40$$

The notation is the same in this expression for the recognition algorithm and k_{ji} is the exponent of the operator polynomial (the order of the algebraic closure). For fixed recognition operators in the AVO class acting as recognition operators B_{ji}, each of them is described completely, i.e. all the parameters $\varepsilon(j, i)$, $p(j, i)$, $\gamma(j, i)$ and $k(j, i)$ are specified. The values of the parameters are computed by special procedures from the learning information I and descriptions of the objects of recognition. So each recognition operator B_{ji} is specified by a collection of $2N + r_m + 1$ numbers. The value of k_{ji} is also computed from the learning information I and descriptions of the objects of recognition, but in all cases [22]

$$k_{ji} \le \{ (\ln q + \ln m |\ln (d_1 + d_2)| - |\ln d_1|) \, [|\ln (1 - (q + m - 2 - \eta)^{-1})|]^{-1} \} + 1 \qquad 2.41$$

where η is an arbitrarily small number.

The exponent k_{ji} can be reduced to q - 1 but in this case the recognition operators cannot be written simply.

The constant values C_{ji} can not always be defined solely from learning information. All the input data in this case is divided into two parts. The objects included in the first part are regarded as learning data and the rest form the array $\omega^1, ..., \omega^q$ to be recognized. But in fact it is known which of the specified categories these objects belong to. It is assumed in this case that $C_{ji} = 1$ if $\omega^j \in \Omega_i$ and $C_{ji} = 0$ if $\omega^j \bar{\in} \Omega_i$.

The algebraic theory of recognition algorithms proves that a recognition operator of the type R_A maintains classification if the objects $\omega', ..., \omega^q$ are replaced by objects from a certain neighborhood of theirs. That means that an algorithm with a recognition operator of the form R_A has a nonzero radius of stability of classification. Hence it is easy to derive the theorem that if the boundaries of the categories are smooth enough and the number q is sufficiently great, the algorithm with a recognition operator in the form R_A nearly always gives a correct answer, i.e. with a probability close to 1. This is a fact in principle also for more general assumptions about the structure of categories. It is valid if the so-called "hypothesis of compactness" is true.

When no information is available about the structure of the categories and reasonable considerations give no grounds to believe that the requirements of the hypothesis of compactness are met, it can be proved using the technique proposed by V.N. Vapnik and A.Ya. Chervonenkis [5] that a recognition algorithm with recognition operator of the R_A type gives a correct answer over the whole collection of acceptable recognition objects with a certain guaranteed probability. The closer this probability is to 1 the greater is q, i.e. the number of objects included in the second part of the partitioning of the learning array. This part of the learning array is often called a check part.

When a collection of objects of a known classification is divided into two parts, the check group should be made as large as possible. And only two conditions must be satisfied: a) the categories should be distinguished in the learning information, and b) for any pair of objects ω^u, ω^v in the check array there should be at least one object ω_r in the learning array and at least one feature x_t, $r = r(u, v)$, $t = t(u, v)$ such that

$$\rho_t(a_{rt}, b_{ut}) \neq \rho_t(a_{rt}, b_{vt}). \qquad 2.42$$

In other words, this means that for any objects in the check array there will be an object and an feature in the learning array such that a certain method of measuring the distance ρ_t in the set of values of the x_t-th feature produces different values for distances from the learning object thus found to the first and second check object.

Let us sum up the basic features and results of the algebraic approach to the general theory of recognition algorithms.

The study of different models of algorithms made it possible to formulate a general definition of the recognition or classification algorithm and to investigate the properties of the set of such algorithms. It turned out that the set of recognition algorithms is an algebra, the operations of the algebra having a collection of properties that could be used to study the set of recognition algorithms in detail. Appropriate algebraic methods solved problems of finding recognition algorithms in a collection, conditions were outlined under which a model contains an algorithm that can classify any final sample with an absolute accuracy, and so on. The algebraic methods can be also used to solve efficiently the problem of selection of an extreme algorithm in the total set of recognition algorithms rather than within the framework of one model.

The algebraic approach has demonstrated that if a collection of non-rigorous algorithms designed for the solution of problems of certain type has a system of relatively easily verifiable properties, it can be extended by adding formal procedures to a system of algorithms that has the following property: this extended system based on "intuitive" algorithms has an algorithm which can provide an exact solution of each problem of the given type.

Thus, the "intuitive algorithms" whose families are usually constructed to solve poorly formalized problems are extended to the set of rigorous algorithms.

The introduction of additional restrictions on the classes of solvable problems and the sets of heuristic algorithms enables us not only to prove these theorems of existence but also to indicate methods for the search to extend such correct algorithms.

Consequently, a basically different path is possible instead of trying to build formal models in fields which are difficult to formalize. It is enough to construct a family of "intuitively reasonable" algorithms to solve appropriate problems and then to introduce algebra for the set of such problems and to derive an algebraic closure of the intuitive family of algorithms. This closure was found to provide in principle a solution for any problem from the set of problems involved in the study of poorly formalized situations. Naturally, the collection of problems itself should be described formally. A set of input information and a set of questions (predicates) to be answered as the problem is solved should be specified.

The main result of the algebraic theory of recognition algorithms is that to describe a class of algorithms that assign correctly the final sample to all categories it is enough to take any complete model, to consider the linear closure of the collection of its recognition operators and to join any correct decision rule to the model. This algorithm will be optimum for any quality functional.

A model of recognition algorithms is complete if its linear closure contains recognition operators such that the matrices obtained from initial information as the operators are applied form the basis in the space of numerical matrices of the dimension $q \times m$. A decision rule is correct if for any final collection of acceptable objects there exists a matrix of numerical estimates (produced by applying the recognition operator) such that the application of the decision rule makes it true information - the matrix of final answers. The completeness of widely adopted models of algorithms has been verified and the results are known.

It was demonstrated that the class of algorithms for standard initial information, which are defined by piecewise linear partitioning surfaces and weights of objects in the learning set, is complete [20]. The same paper has shown that the AVO class with threshold functions of closeness defined by specifying accuracy thresholds for features, weights of features and weights of objects in the learning set is correct for initial information if some restrictions are imposed on the latter and the class of recognition operators has an addition in the form of an operator which separates maximum elements of the numerical matrices.

These proofs are based on general criteria that allow one to establish the correctness of appropriate closures. The closures themselves, however, contain infinite sets of algorithms. Naturally, the theorems of existence cannot by themselves produce an algorithm which will be correct for the given problem if its information matrix is unknown. Thus it was demonstrated that a finite set of algorithms could be effectively isolated in a closure and that the algorithms contained an algorithm which is correct for the given problem (the information matrix of the problem being unknown). A similar construction is made when there is some knowledge about the information matrix [19].

In addition, it was proved that for any problem z set by the initial information I and a collection of objects of recognition ω^q there existed a neighborhood in the space of the problems centered at (I, ω^q) for which the correct algorithm does not change while the information matrix is maintained. A method was demonstrated for computing the radius of the neighborhood and so it was established that correct algorithms were stable in closures [19].

If the hypothesis of compactness is true for specified categories, a method for formulating correct algorithms can be indicated. But if there are problems with different information matrices in any neighborhood of a problem, one correct algorithm for a problem in the neighborhood cannot be formulated in principle. In this case it is possible to isolate finite sets of algorithms of minimum power, which will be correct for problems in the neighborhood. Ref. [19] describes a method for constructing such sets.

Most of the studies within the algebraic approach were made in extensions of the AVO model but they can be made in just the same way for any model whose extensions are complete and thus correct.

The basic scheme of the recognition algorithm class can be described as follows [18]:

1) Form models of recognition algorithms.

2) Select an extreme algorithm and a model.

3) Divide extreme algorithms into classes, formulate a linear closure in each class, and select an algorithm which is optimum in terms of coefficients of linear combination and of the parameters of the decision rule.

4) Work out an optimum corrective operation and apply it to correct the optimum algorithms obtained in stage 3.

A detailed description and the proof of the results produced by the algebraic approach can be found in Refs. [19, 20].

CHAPTER 3

IMAGE RECOGNITION

3.1. General description of the problem

In the last few decades, particularly since the 1970s, problems whose input information is visually represented have moved more and more to the foreground among application problems in data processing. This was brought about by the appearance of new devices for collection and reproduction of information. These display efficiently and graphically data recorded and accumulated in the form of images. This was accompanied by the growing popularity of recognition as a new information technology - a powerful, practical and, in a sense, universal methodology for mathematical processing and evaluation of information and detection of hidden patterns [11].

The problem of image recognition is still rather changeable. It may arise in the form of recognition problems proper, or as an analysis of scenes, as problems of image understanding, or as problems of so-called machine vision. The objects of recognition (of analysis or understanding) may be images obtained in various parts of the full radiation spectrum (optical, infrared, ultrasonic, etc.) by various methods (television, photographic, laser, radar, etc.), transformed into digital form and represented as a certain integer-valued matrix.

The role of an image as an object of information technology is due to the fact that the image is a special type of information that combines and mixes the input (represented) information and the form of the representation. It also combines the information model and the physical model of objects, phenomena and processes thus represented. Regretfully, the specific character of the image is not yet well understood, nor has it been studied in information science or in research into visual perception of humans and animals (psychology, psychophysics and neurophysiology).

Images have great information capacity, they are compact and graphic while vision is the most natural human mechanism for perception of information about and from the outside world. Moreover, this it is the oldest way of representation and perception of information in terms of evolution. When perceiving an image, as far as we can judge, the human mind does not try to describe it in words but handles it as a whole pattern or a system of patterns with a non-linguistic inner comprehension. In developing methods and systems of automatic image recognition one has to find ways of efficient formalization of images in order to be able to process perceptions (descriptions) reflecting the semantics of an image. In other words, one has to find information contained in the image's interior structure and the structure of external relations of the part of the real world (scene) reproduced by the image.

The specific character and complexity of image recognition problems, and the difficulties they imply result from the need to compromise between quite contradictory facts that reflect the requirements for analysis, the nature of vision, methods used to produce, to form and to reproduce images and the existing mathematical and technical capabilities. Obviously, the main contradiction lies between the nature of an image and the analysis based on the use of the formal apparatus (in fact, a model) of an object. The trouble is that to benefit from the advantages offered by data representation in the form of an image the data must be given a "non-pictorial" shape since appropriate algorithms are adapted for processing only certain symbolic descriptions. Images are by nature an excellent object for parallel processing methods. However, most of the existing recognition techniques are serial, particularly because the organization of the recognition procedure requires recording of results produced at intermediate stages. An overwhelming majority of image processing methods is purely heuristic and their merits, in fact, depend on how successfully they can overcome the "pictorial" character of an image by "non-pictorial" means. Thus they rely on procedures that do not need the organization of information as an image.

An analysis of the present state of image recognition leads to the following conclusions:

1. The mathematical theory of image recognition has not yet materialized and unfortunately the specific character of images does not let us apply directly the methods and devices of the classical ("unidimensional") theory of pattern recognition and digital signal processing.

2. The absence of a mathematical theory of image recognition is an obstacle to a well-grounded and systematic development, selection, comparison and use of image recognition algorithms. Neither does it help to obtain a verified and reliable estimate of their efficiency and adequacy.

3. The task of image recognition should be set, studied and carried out as a mathematical problem.

4. The development of an informational theory of image recognition involves:

a) mathematical formulation, characterization and systematization of image recognition problems;

b) development, study, characterization and systematization of methods and means for formulation of models of images oriented towards the recognition problem;

c) development, study, characterization and systematization of transformations that can reduce an image to a form suitable for recognition;

d) development of formal structures to describe models of image recognition algorithms using the latter as a knowledge base to define and characterize classes of image recognition algorithms;

e) standardization of models for individual categories of images as applied to the recognition problem;

f) development of systematic methods for automatic selection of an optimum model of a recognition algorithm for a specific image;

g) realization of the above methods and models in computer software designed to compile, study and estimate image recognition algorithms.

3.2. Types of image recognition problems

The following recognition problems arise in dealing with input information provided in the form of images:

1. Comparison of two images as a whole to establish whether they belong to the same category (to determine whether the images represent the same object or a scene).

2. Comparison of an image as a whole with a set or a series of successive (in time) images which represent a certain category of images (i.e. objects or scenes), the purpose being the same as in problem 1 above.

3. Problems 1 and 2 for the case of several categories.

4. Search for regularity/irregularity (of an object or situation) in an image presented for recognition to pay attention to although it was not specified in the a-priori list of prototypes (the associative search and boundedly determinate collection of categories - the problems of logical and semantic filtration combined with self-learning).

5. Search for a regularity/fragment of a specified form in an image presented for recognition.

6. Partitioning of a set of images into disjoint subsets (the problem of automatic classification).

7. Solution of the problem of automatic classification in one image (the division of an image into uniform domains, groups of objects, segmentation of a domain, and isolation of features of objects).

8. Combined solution of problems 6 and 7.

9. Automatic isolation of nonderivative elements, characteristic objects of an image, feature objects, and spatial and logical relations for the synthesis of formalized representations and descriptions of the image.

10. Reduction of an image to a form suitable for recognition, and an automatic synthesis of formalized representations and descriptions of images.

11. Solution of problems 9 and 10 in dialog mode.

12. Problems of reconstruction of:

- missed frames in a sequence of images;

- images as a whole from fragments;

- fragments of images (and objects) on the basis of nonderivative elements, features and productive procedures taking into account the context of the entire image;

- the path of a problem from its fragments and unknown fragments of the paths.

13. Selection and formation of the path of an image recognition problem (in the sense of a recognition problem with standard learning information).

14. Solution of problems 1-13 when the image has dynamic objects and a complex background situation (including dynamic and static noise) and taking into account the methods for production, formation and representation of images.

3.3. Mathematical formulation of the image recognition problem

When images are analyzed and recognized, the information thus processed is represented as a numerical matrix that reproduces the properties of the depicted object (scene) and distortions due to the method and process used to produce the images. To formalize image recognition we define three sets (models) of images, on which the existence of equivalence classes is postulated, and the sets of acceptable images specified on equivalence classes. The introduction of the equivalence classes on the set of models reflects the hypothesis that any image has a certain regularity or a mixture of regularities of different types. Under this assumption, a recognition problem can be reduced to the division of images maintaining its own regularity and those whose inherent regularity was upset (naturally, the problem can be set to find a regularity or distortions of regularity of certain types in an image).

Consider the following model. Let I be a true image of an object under study. The operations of production, formation, discretization, etc. (all the procedures required to process an image) can be regarded as transmission of a true image along a channel with interference. As a result, it is not the true image that is analyzed but a certain actual image I^* being observed. The analysis should classify the latter, i.e. define its prototype in the true equivalence class K_i. Or a regularity (regularities) of a specified type J^R should be found in the image I^*.

Hence we can define the sets $\{I\}$, $\{I^*\}$ and $\{I^R\}$, $\{I\} = \bigcup_i K_i$ and transformations of the formation $\{T^F\}$ and recognition $\{T^R\}$ of images:

$$T^F : I \to J^* \qquad 3.1$$

$$T^R : I^* \to J^R \qquad 3.2$$

Thus, image recognition is reduced to defining the set $\{I\}$ of algebraic systems of transformations $\{T^F\}$ and $\{T^R\}$ on the equivalence class and to applying them to the images I^* being observed in order to: a) analyze "backward" - to divide images according to the nature of regularity (to reconstruct true images, i.e. to indicate the equivalence class they belong to), and b) analyze "forward" - to find regularities of a certain type I^R and to localize them in the image I^*.

This formulation of the recognition problem makes it possible to define the class of image processing procedures. This class is characterized by a fixed structure of the process, and the interpretation (specific realization) of the structure depending on the purpose and type of the analysis. The following main stages are distinguished in the process of recognition:

1^0. Synthesis of a model of an observed image $\mathfrak{K}\{I^*\}$. This is a so-called "stage of reduction to a form suitable for recognition", i.e. obtaining a certain formalized description of the image suitable for processing with appropriate transformations - algorithmic recognition procedures.

2^0. Logical filtration of images. This stage involves preprocessing of the observed image to provide its preliminary classification which is necessary to select the set of transformations $\{T^F\}^{-1}$. A correspondence is assumed to exist between the type and/or nature of the model $\mathfrak{K}\{I^*\}$ and the equivalence class defined on the set $\{T^F\}$. It is assumed, in addition, that there is a weak equivalence on $\{T^F\}$ to place in correspondence the subsets $\{T^F\}$ and the equivalence class of models of true images $K^i[\mathfrak{K}\{I\}]$ in the sense of that weak equivalence.

3^0. The establishment of an equivalence class of the true image $K_i[\mathfrak{K}\{I^*\}]$ that produces the given image I^* being observed. For this purpose, inverse transformations of the formation $\{T^F\}^{-1}$ are applied to the model $\mathfrak{K}\{I^*\}$. In addition, the analysis at stage 2^0 serves as a basis for a hypothesis about the equivalence class true for $\mathfrak{K}\{I^*\}$. This makes it possible to apply the transformations T^F to the model of the true image prototype to verify the permissibility of the generation of I^* under consideration in the appropriate equivalence class and to compare the results of application of $\{T^F\}^{-1}$ to $\mathfrak{K}\{I^*\}$ and T^F to $K_i[\mathfrak{K}\{I^*\}]$. In keeping with the methodology of the algebraic approach, both transformations can be used in the form of linear and algebraic closures of appropriate transformations:

$$L\{T^F\}^{-1} : \mathfrak{K}(I^*) \Rightarrow K_i[\mathfrak{K}(I^*)]' \qquad 3.3$$

$$L\{T^F\} : K_i[\mathfrak{K}(I^*)] \Rightarrow \mathfrak{K}(I^*)'. \qquad 3.4$$

The process stops when the equivalence of $K_i[\mathfrak{K}(I^*)]'$ and $K_i[\mathfrak{K}(I^*)]$, $\mathfrak{K}(I^*)$ and $\mathfrak{K}(I^*)'$ or the equivalence of intermediate results of transformations in the analysis "forward" and "backward" is achieved. This mechanism of reconstruction of the equivalence class which is true for $\mathfrak{K}(I^*)$ is called the procedure of reversible algebraic closure [8].

4^0. Transformations of recognition T^R are selected according to $K_i[\mathfrak{K}(I^*)]$ since it is also assumed that the subsets of transformations $\{T^R\}$ and $K_i[\mathfrak{K}(I)]$ correspond:

$$K_i : \{T^R\} \to T^R(K_i) \qquad 3.5$$

5^0. "Recognition", i.e. detection of the desired regularities I^R on I^* by applying the analysis "forward" to $\mathfrak{K}(I^*)$ and simultaneously the transformations $\{T^R\}^{-1}$ to $\mathfrak{K}(I^R)$, i.e. analysis backward. This involves the procedure of reversible algebraic closure, just as in stage 3^0, but this time to establish whether the desired regularity $\mathfrak{K}(I^R)$ can be generated by the model of the observed image $\mathfrak{K}(I^*)$:

$$L[T^R(K_i)] : \mathfrak{k}(I^*) \Rightarrow \mathfrak{k}(I^R), \quad 3.6$$

$$L[T^R(K_i)]^{-1} : \mathfrak{k}(I^R) \Rightarrow \mathfrak{k}(I^*)'. \quad 3.7$$

The process stops when the equivalence of $\mathfrak{k}(I^R)$ and $\mathfrak{k}(I^R)'$, $\mathfrak{k}(I^*)$ and $\mathfrak{k}(I^*)'$ or the equivalence of intermediate results of transformations of the "forward" and "backward" analysis is achieved. The equivalence missing, new iterations are performed for stages 1^0 - 5^0 with other $\mathfrak{k}(I^*)$ and hypotheses about $K_i[\mathfrak{k}(I^*)]$.

This formulation reduces the recognition to defining the set $\{I\}$ of algebraic systems of transformations $\{T^F\}$ and $\{T^R\}$ on the equivalence class and applying them to the images $\{I^*\}$ being observed in keeping with the method of algebraic reversible closure. It is done: a) to analyze "backward", that is to divide images according to the nature of their regularity ("to reconstruct" true images, i.e. to indicate the equivalence class they belong to), and b) to analyze "forward", that is to find regularities of a certain type I^R and to localize them in the images I^*.

Thus, the mathematical formulation of the image recognition problem has the following form:

a) We are given

1^0 $\{I\}$: a set of ideal images;

2^0 $\{I^*\}$: a set of observed images;

3^0 $\{I^R\}$: a set of images as the results of realization of the recognition process (the set of decisions);

4^0 $\{T^F\}$: a set of acceptable transformations of the formation of images;

5^0 $\{T^R\}$: a set of acceptable transformations of image recognition.

b) Let

$$1^0 \ \{I\} = \bigcup_i K_i,\ K_i,\ i = 1, \ldots, l;$$

$$2^0 \ \exists \mathfrak{k}(I) \in K_i(\mathfrak{k}(I)) \mid \forall I \in K_i;$$

$$3^0 \ K_i(\mathfrak{k}(I)) = K_i(\mathfrak{k}(I^*)) \mid t^F : I \to I^*,\quad t^F \in \{T_i^F\};$$

$$4^0 \ \langle K_i(\mathfrak{k}(I))\rangle \underset{R}{\Leftrightarrow} (\{T_i^F\} \in \{T^F\},\quad \{T_i^F\}^{-1} \in \{T^F\}^{-1});$$

$$5^0 \ \langle K_i(\mathfrak{k}(I))\rangle \underset{R}{\Leftrightarrow} (\{T_i^R\} \in \{T^R\},\quad \{T_i^R\}^{-1} \in \{T^R\}^{-1});$$

$$6^0 \ \langle K_i(\mathfrak{k}(I))\rangle \underset{R_M}{\Leftrightarrow} K_i(\mathfrak{k}(I^R)),\quad t^R : I^* \Rightarrow I^R,\ t^R \in \{T_i^R\},$$

where K_i is the equivalence class of the images; $\mathfrak{k}(I)$ is a formal description of an image, R_M is a morphism, and R is the correspondence relation.

c) It is required to compute the value of the predicate

$$P(\mathfrak{k}'(I^R) \in K_i(\mathfrak{k}(I^R)) \mid \{t^R\} : \mathfrak{k}(I^*)$$

$$\{t^R\} \in \{T_i^R\}\ i = 1, \ldots, l; \quad 3.8$$

d) The structure of the solution of the recognition problem:

$$1^0 \ \mathfrak{k}(I^*);$$

$$2^0 \ P((\mathfrak{k}(I^*)) \in K_i(\mathfrak{k}(I^*))) = ?$$

$$3^0 \ L(t_i^{F-1}) : \mathfrak{k}(I^*) \underset{K_i}{\Rightarrow} \mathfrak{k}'(I^*);\quad L(t_i^F) : \mathfrak{k}(I) \Rightarrow \mathfrak{k}'(I);$$

$$4^0 \ P_{k_i}(\mathfrak{k}'(I^*) \underset{R_{K_i}}{\Leftrightarrow} \mathfrak{k}'(I)) = ?$$

$$(P_{k_i} = 1) \Rightarrow 5^0,\ (P_{K_i} = 0) \Rightarrow (2^0 \div 4^0),$$

$$5^0 \ L(t_i^{R-1}) : \mathfrak{k}(I^R) \underset{K_i}{\Rightarrow} \mathfrak{k}'(I^R),\quad L(T_i^R) : \mathfrak{k}(I^*) \underset{K_i}{\Rightarrow} \mathfrak{k}''(I^*);$$

$$6^0 \ P_R(\mathfrak{k}'(I^R) \underset{R_{K_i}}{\Leftrightarrow} \mathfrak{k}''(I^*)) = ?$$

$$(P_R = 1) \Rightarrow \text{stop}$$

$$(P_R = 0) \Rightarrow (2^0 \div 5^0)$$

where L is a linear (algebraic) closure and R_{k_i} is an equivalence relation specified on the wheel K_i.

3.4. The descriptive theory of image recognition

The basic features of this theory are determined by its objectives - to create regular methods for the selection and synthesis of algebraic data processing procedures in image recognition problems - and by the specific character of the formulation and solution of the recognition problem where the input information is in the form of an image and a structural model of the image recognition algorithm. We describe below the fundamentals of the descriptive theory.

1. The principle of generation. The recognition process includes information which reflects the formation of a visual pattern. The structure of the pattern is found by finding what subpatterns can be isolated in the whole pattern, to what degree they can or must be elementary, and how the elements relate to each other. The main task is to study and to use the structures of the relationships of the elements that comprise the pattern. As a result, the description of a complex object on the image is constructed as a hierarchical structure formed by simpler objects. In other words, it becomes possible to use and represent explicitly the hierarchical (tree-like) information contained in the image.

As applied to recognition, this method of the specification of images can be realized as detection and processing of regular structures. The representation of the image by the hierarchical structure naturally leads to combinatory regular structures. The latter make it possible to obtain a practically unlimited variety of descriptions with a quite limited number of atomic (primitive) elements and a limited number of combination rules. This involves the unlimited (e.g. recurrent) application of the rules to the initial elements and to the results of the realization of appropriate recurrent procedures. When used as a mechanism for describing the structure of images, the combinatory regularity saves description resources considerably.

This method of specification ("inductive generation") of a category of objects is called the generalized inductive definition in mathematics. It can be outlined as follows: (1) some (initial) objects are specified in the category to be defined, (2) some rules are specified to obtain other objects of this category out of the objects already defined, and (3) the objects in this category are only those which were constructed as specified in 1 and 2 above.

So the idea of the principle of generation is that the formalized description of an image is specified in recognition as a certain system of objects bound by structural relations. The objects are isolated in the image using a system of transformations and are specified by transformations which indicate the acceptable method for their construction. Notice that these transformations ("generative procedures") have functional completeness with respect to an appropriate equivalency class induced on the set of ideal images.

2. Formalized description. The input information for image recognition is a formalized description of an image - a model of the image. All the transformations applied to the formalized description of the image are introduced for one of the following three purposes: a) to produce a new formalized description of the image, b) to reduce the image to a form suitable for recognition, or c) to obtain an aggregate estimate of the formalized description, i.e. to transfer from the space of input informations to that of estimates. The latter usually serves to realize the decision making process for classification in recognition.

3. The specific character of the formalized representation of an image. The formalized representation is understood as a formal scheme designed to obtain explicitly objects of recognition and generative procedures, i.e. a standardized formal description of forms of the surfaces that generate the image. The formalized description is understood as specific realization of the formalized representation. The input data for the synthesis of formalized image representations may be values of brightness which depend on geometric properties. The latter include spatial organization and the reflective power of visible surfaces, illumination of the scene and the position of an observer. The system of representations used in image recognition should include characteristic objects of the image, which could be associated with the attributes that indicate values of variables such as brightness, orientation, dimensions and location. The characteristic objects should correspond to the actual physical features of the surfaces. This means that the structure and properties of the real world - the object of the image - reproduced by physical restrictions play a key role in obtaining information about a surface.

The following four aspects determine the specific character of formalized image representation.

a. Multitude of levels in terms of scale and morphology. Image recognition uses a system of representations that includes several layers of formalized descriptions of images, each of which corresponds to its scale level or morphological level. The division into morphological levels depends on the types and complexity of the characteristic objects and of non-derivative elements specified and generated at each level. The division into the scale levels is determined by the scale of the characteristic objects and of non-derivative elements specified and generated at each level, and by appropriate scale transformations. The multilevelness of the representation regularizes the selection of the system of generative transformations and creates an information redundancy to compensate for the "partial" character and incompleteness of formalized descriptions corresponding to individual levels. It should be noted that Pavlidis classes are one of the versions of the specification of morphological levels [11].

b. Syntactical and relational information. Due to the way it is generated, the information which characterizes the syntactical structure of an image is used extensively in the synthesis of representations. The relational information is important for specifying a representation as a formal construction since it defines connections between objects of the representation. Note that the predominance of syntactical and structural information in implicit image representations was reflected in the wide use of structural methods for the analysis of images, particularly at the early stages of its evolution.

c. Features. These are used for image recognition in two capacities: as primitives whose assembly by generative transformations produces characteristic objects of recognition, and as characteristics which are associated with characteristic objects to reflect and fix the types of changes in the variables taken from the images. They are useful, for example, in defining variation of the visible surface relative to an observer and distances from the latter. The features make it possible to use local and global information contained in the image and they determine the relationship of the global and local information used in the formation of a description.

New types of features arise in image processing (with respect to classical recognition models with standard information). They allow one to reproduce the two-dimensional character of the object of recognition. The notion of the feature of an image is based on the concept of local algorithms for computation of information [15]. It involves, in particular, considerable use of the concept of local neighborhood.

An feature is regarded as a manifestation of a property in an image, interesting in the context of the problem, which may either carry some semantic load, or reproduce physical or geometric properties of a scene, or have a quantitative measure (the numerical value of the functions or characteristics associated with a fragment of the image). The feature of the image is specified through a predicate that determines the distinctness or manifestation of an appropriate property in a local neighborhood of the image - the support set of the predicate (support set of the feature). This support set is defined as a minimal local neighborhood which maintains a stable computability (or just computability in a weaker case) of the feature's predicate. The stable computability is characterized by the local character of the computation (by the complexity and required resources of the algorithm) and the acceptable range of variation of pixels in the local neighborhood. (Note that one of the variants of the reduction of an image to a form suitable for recognition is a partitioning of the image into support sets of features or construction of an optimum packing of the image with the support sets of the features).

The estimate of a property may be, in particular, some numerical value that reflects the properties of a local segment of the image (the distribution of pixel values over this segment, the presence or absence of a characteristic object in this section, the type of form of the object isolated in the section, and the power of the local neighborhood.) In addition, useful local features, i.e. those computed with respect to the local neighborhood of an image, may comprise: (1) characteristics based on the Shannon measure computed for the distribution of the type of neighborhoods of individual image elements, (2) characteristics based on the distribution of types of Boolean functions, which are specified on the neighborhoods of image elements, (3) characteristics based on the distribution of types of partially defined Boolean functions, which are specified on the neighborhoods of image elements, and (4) characteristics based on numerical estimates of the properties of the connectivity graphs and of partially defined graphs of connectivity of uniform parts of the image elements' neighborhoods.

The features used for the synthesis of the model of an image are classified: (1) by function (generative (descriptive) features as non-derivative elements, features as characteristic objects, feature generating procedures, and parametric features), (2) by the nature of the represented information (global and local features), (3) by the production method (measured or isolated in an image, and computable), (4) by the mathematical techniques used to form or compute the features (statistical, algebraic, topological, spectral, geometric, or matrix features), (5) by the type of the images they comprise (brightness, binary or textured features), and (6) by the types of objects represented by the attributes (skeletal, contour, or segmented features).

Note, in particular, that the representation of the "raw sketch" type to specify information characterizing the variation of brightness and distribution, and the geometric characteristics of two-dimensional images uses primitives of the following types: intersections of the zero level, spots, breaks, discontinuities of edges, boundaries, and lines. But the representation of the type "two-and-a-half-dimensional sketch" to specify information characterizing geometric properties of visible surfaces (orientation, depth, and contours of discontinuities) uses primitives of the following types: local orientation, distance from the observer, and discontinuities [11].

d. Models. As implied by the principle of generation and by the types of attributes used for the synthesis of formalized image representations, images are represented in recognition by models of two basically different types. These are descriptive (generative) models and attributive (parametric) models.

The former models reflect the structural organization of the image and represent the information to be processed for transformation of representations. The latter reflect actual physical properties of the image objects and the properties of the numerical matrix whereby the observed image is specified. The attributive models are generally more convenient to use for aggregate estimates of a formalized description. On the whole, the descriptive models are more suitable for characterizing the processes of transition between representations of different levels while the attributive models are better associated with the results of transformations of representations of specific levels.

Note that the establishment of correspondences of these two types of models as applied to a level or step of the recognition process can serve as a criterion for control of the progress of the process, or as the stopping rule.

4. Duality. A basic feature of the image recognition process is that it has dualities of two types, including that between a formalized description of an image and the recognition procedure, and the duality between a formalized description of an image and its formalized representation. The duality in this case is understood as the property of inner symmetry typical of a number of axiomatic theories and expressed in compatibility of some basic concepts. As mentioned above, image recognition changes the concepts of initial and final information. The multilevelness of the recognition process reduces it to a sequence of transformations of formalized image descriptions related to various morphological or scale levels. And the formalized description of an image acts both as initial and final information since the desired result of the analysis of images may require a classification decision as well as a formalized description corresponding to certain conditions defined by the context of the problem.

The duality of the second type is associated with the fact that the type of chosen representation determines to a great extent the properties used to produce an appropriate formalized description (characteristic objects and non-derivative elements) while the latter, in turn, affects the types of formalized representations included in the path of transformations, which realizes the recognition process (connected, among others, with the problem of admissibility of generative transformations).

5. Characterization of the recognition process. The organization of the image recognition process is governed by several basic principles which determine the methods of formation of recognition transformations, the structure of the transformations, the mechanism of the organization of the process as a whole and the model of the image recognition algorithm.

a. When recognition transformations are selected and formed, the basis is the methodology of the algebraic approach to the problems of recognition and classification. The methodology is known to include three main stages - the selection and use of basic heuristic procedures and optimization in a model, and the synthesis of the problem solution procedure by correcting heuristic procedures on a set of basic models (using the mechanisms of linear and algebraic closure).

b. In contrast to the classical model of the recognition algorithm, that of the image recognition algorithm consists of three rather than two elements, including:

- an operator for the reduction of the image to a form suitable for recognition:

$$R_f^n(I_n) = P_n(I, \Omega^m, \omega'),\ n = 1, \ldots, t, \qquad 3.9$$

where I_n is an image that corresponds to a morphological (scale) level of the formalized description and $P_n(I, \Omega^m, \omega')$ is a model of the image at the n-th level of the formalized description, the model having been produced by the operator for the reduction of the image to a form suitable for recognition;

- a recognition operator that transforms the formalized description into a numerical matrix $\|a_{jt}\|$ of standard size with the number of rows equal to that of the objects to be recognized in the problem, and the number of columns equal to that of the categories considered in solving the problem:

$$B(I, \Omega^m, \omega') = \|a_{jt}\|_{q \times m} \qquad 3.10$$

- the decision rule to process the numerical matrix into a matrix of final answers $\|a^{jt}\|_{q \times m}$ with the same number of rows and columns:

$$c(\|a_{jt}\|_{q \times m}) = \|a_{0jt}\|_{q \times m}. \qquad 3.11$$

Obviously, the main function of the reduction operator is to obtain a formalized description of the image such that it will be suitable for the recognition operator (3.10), i.e. to produce the formalized description which will include numerical estimates of the information contained in the image to take into account the two-dimensional nature of the information. Note that the reduction operator in the DAVO class (algorithms to compute estimates from two-dimensional information) [10] is a discretization grid with cells of an arbitrary (regular) shape, each of which covers a collection of adjacent pixels of the initial image. The specific method of estimation for a local neighborhood is defined by a set of structural parameters which indicate the cell's size, shape, power, and orientation, the method of computation (the averaging rule, the majority rule, or both, but taking into account pixels vertically, diagonally, horizontally, etc.), the weights of pixels, and values of discretization thresholds. This means that the reduction operator is specified by a parametric model which can be handled in just the same way as other models of the heuristic recognition procedures.

The three conditions above (3.9-3.11) comprise a standard element of the image recognition process.

c. On the whole, the structure of the recognition process is hierarchical due to the duality between the use of recognition procedures and formalized descriptions of images. Curiously, the transformation of representations proceeds not only "vertically" - between different morphological and scale levels - but also "horizontally" - in optimizing the selection of a formalized representation and description within the same level. Thus the path of the recognition process includes vertical segments (the process being controlled by horizontal transformations) and horizontal segments (the process being controlled by vertical transformations). The basic control mechanism is the establishment of a correspondence between generative and parametric models at individual levels in the image description hierarchy.

d. The correspondence of formalized descriptions synthesized during recognition and the satisfaction of the conditions for the process stopping rules are checked using the mechanism of reversible algebraic closure (see Section 3.3 above). The meaning of the transformations (3.3-3.7) is that the recognition process is iterative and reversible. Hence the correspondence is achieved at an intermediate stage (in the sense of formalized representations and descriptions) and the transformations are applied in parallel to the initial models and to final models if their form is specified, or to appropriate hypothesis models. Naturally, this mechanism is used both for vertical and horizontal transformations of the hierarchy of image formalisms. The basic structural element is the trio of operators (3.9-3.11) while the resolution power (and complexity) of the transformations is increased from a basic heuristic to the model of the heuristic procedure to correction on the set of basic models.

6. Logical filtration. Obviously enough, the image recognition process makes extensive use of knowledge about the data domain, the nature of a problem, physical realities of the image object (scene), the universal physical, logical and mathematical laws that naturally govern also the image object (e.g. geometric laws), methods used to reproduce and form the image, and the circumstances accompanying these processes. The knowledge is used in partitioning into equivalence classes, for hypotheses about the type and character of

final information put forward during recognition, and in selecting and comparing basic heuristics and models. It is also applied to formation of the rules of stopping and controlling the recognition process, to the selection and assignment of non-derivative elements and characteristic objects, and types and levels of formalized representations of an image. The basic forms taken by these a-priori values are semantic information and systems of physical restrictions. The main mechanism for realization of the knowledge is logical analysis, particularly logic inference.

7. The set of transformations to solve a specific image recognition problem forms the path of the solution. The path is formed, in a sense, similarly to the stage of learning or adjustment of an algorithm to the problem in the classical version of recognition. Obviously, for recognition problems of the same class (the analysis of ideal images in the same equivalence class), the paths should be close in the transformation space. In addition, there must be a closeness of appropriate formalized representations and descriptions. In this case, the synthesis of the path of a solution of a new problem is actually a problem of selecting an optimum path from a set of paths associated with an equivalence class (at least after the equivalence class has been restored for the observed image). Another version of the problem is regeneration of the path from its fragments associated with certain types of descriptions, representations and their transformations. This problem may arise in connection with that of efficient use of the knowledge about the problem and, in particular, at the stage of logical filtration. Obviously, the bundles of problem paths can form the basis for the synthesis of appropriate parametric models where, among others, the choice optimization problem can be set and solved under certain conditions. Naturally, it leads the solution of the path synthesis problem to the algebraic approach of the "heuristic-to-model-to-correction" type.

A vital open issue in the descriptive theory of image recognition today is the systematization of image description means in recognition problems (largely features, primitives and characteristic objects) and of methods for the synthesis of formalized representations and descriptions (mainly generative structures).

CONCLUSION

RECOGNITION AS THE INFORMATION TECHNOLOGY FOR DEVELOPMENT OF ALGORITHMIC KNOWLEDGE BASES

This review summed up the advancement and present state of recognition as an information technology (one of the first to be actually and widely adopted in practice), which has a well-developed body of mathematics and excellent application capabilities. Having analyzed the progress made by pattern recognition and reviewed the image recognition problem we came to the following main conclusions:

a) Pattern recognition has a well-developed and, in a sense, complete mathematical theory formulated on the basis of the so-called "algebraic approach".

b) The problem of recognition using standard data and that of image recognition differ so much that the methods and devices of classical pattern recognition theory cannot be used for the latter directly. Although set and solved within the general methodology of recognition, image recognition problems require a special branch in the theory of recognition to be developed specifically for image processing.

c) The theory of image recognition can be advanced on the basis of the descriptive approach as a descriptive theory of image recognition.

The purpose of the algebraic approach to recognition is to produce an algorithm which can isolate all useful information out of input data and obtain a solution which corresponds completely and accurately to the "information content" of the useful data thus extracted. This solution is characterized by a minimum (relative) computational complexity, stability to noise and distortion of the input data and statistical reliability. The solution process makes extensive use of the principle of precedence, the formalized concept of generalized closeness, automatic adjustment of algorithms to the problem, including automatic selection of an algorithm class optimum for the particular class of problems, and the principle of correction of the final solution by extending the basic set of algorithm models used to produce the solution.

The solution proceeds at many levels. The first stage involves the formulation of an heuristic model of the algorithm to reflect the specific character of the problem. The next stage deals with models of families of algorithms generated on the basis of a principle selected heuristically by standard means. The recognition algorithm is optimized at this stage within the framework of individual models. At the third stage, the desired algorithm is synthesized from algorithms that belong to different models.

Thus, the algebraic approach to data processing in problems of recognition, prediction and artificial intelligence realizes a philosophy that helps to synthesize an algorithm which can solve a specific problem under certain non-rigorous and easily verifiable conditions. It is actually a methodology for automatic synthesis of recognition and prediction algorithms to provide preliminary analysis of a problem and take into account its peculiarities whereupon a solution method can be selected and an appropriate algorithm proposed on its basis. The main difference between the algebraic approach and other recognition methods is that the latter lack stage 3 and so cannot truly produce an exact solution. The idea of the stage is that it eliminates the difficulties that arise at stage 2 (the inaccuracy of the models of heuristic recognition algorithms and the difficulties involved in actual optimization and multiparametric space), and can ensure an absolutely exact solution in the above sense unlike the locally extreme solutions provided as a rule at stage 2.

The descriptive theory of image recognition was formulated and is advanced as a development of the philosophy of the algebraic approach in recognition problems to process images. It provides solutions of problems that arise in generation of formalized representations and descriptions of images as objects of recognition, and in the synthesis of recognition procedures. The solutions are produced by studying the inner structure and the content of an image as a result of the generative operations which can be used to construct the image from non-derivative elements and objects isolated in the image at various stages of its analysis. Since this method of characterizing an image is operational, the whole process of image processing and recognition, including the construction of a formalized description - a model of the image (the path of the problem) - is regarded as the realization of a system of transformations of images. The transformations are defined on equivalence classes which are ensembles of acceptable images (an ensemble is also specified descriptively by a system of prototypes and a collection of generative transformations, which is functionally complete relative to the equivalence class). The recognition process involves a hierarchy of formalized descriptions and representations of images, particularly models related to different morphological and scale levels of representation, multilevel models that allow one to select and vary the required extent of details in the description of an object of recognition.

The descriptive approach was brought about due to some inherent properties of an image as a way of representing information, and peculiarities of organization and realization of image recognition processes. Obviously, these include above all the following:

a) The synthesis of a formalized description of an image as an object of recognition becomes an independent problem which is set and solved within, but not without, the recognition process, as is usually the case in the classical recognition methods with standard information.

b) The formalized description of an image should maintain the advantages due, in fact, to the use of the image as a means of representation of input data. Therefore, when the description is formulated, the image should not be "destroyed". In other words, the description should preserve the "pictorial nature" of the image. It should be specified by a certain formal construction whose synthesis is based consistently on the principles of generation and the hierarchy in the structure of the object of recognition. The synthesis is also underlaid by the relations that exist between the descriptions included in the hierarchy and their non-derivative elements and characteristic objects both within individual morphological levels and scale levels, and between them.

c) The connection (duality) between procedures for generation of formalized descriptions and those of recognition determined by both the role of the models proper in the image recognition process and the possibility of representing the result of the recognition as a new formalized description compared with the image being analyzed.

d) The realization of the image recognition process as a path of transformations of formalized descriptions presupposes control of its progress using information obtained by setting up correspondences between formalized descriptions of the image within and between hierarchical levels.

The concept that underlays the descriptive theory of image recognition involves standard organization of procedures of recognition and representation of information for the solution of recognition problems. The structure of the process is multilevel. Features, primitives, characteristic objects and generative procedures are selected, found or calculated at each level to be used to construct a model of the image there. When the means to form a description are defined, extensive use is made of knowledge about the data domain, about the problem, and the universal physical, logical and mathematical laws and restrictions that reflect the realities of the scene which is an object of the image. The structure of the recognition process reflects specific features of the model of the image recognition algorithm (it includes the operator of the reduction of the image to a form suitable for recognition, the recognition operator and the decision rule), the possibility of transition from the space of formalized descriptions to the space of aggregate estimates, and the realization of transformations by the mechanism of reversible algebraic closure.

The most important points of the descriptive theory of image recognition are:

1. The principle of generation.

2. Formalization of the description and representation of an image by taking into account the multitude of levels in the structure of descriptions, by using syntactical and relational information, by generalizing the concept of the feature of the object of recognition as applied to the problem of image recognition, and by using generative and parametric models in parallel.

3. The duality of procedures for the synthesis of formalized descriptions and recognition procedures, and the duality of formalized descriptions and representations.

4. Organization of the recognition process including the methodology of the algebraic approach (from a heuristic to a model to correction on the set of models), generalization of the concept of recognition algorithm as applied to the problem of image recognition, the hierarchical character of the transformation procedures, and their realization by the mechanism of reversible algebraic closure.

5. Logical filtration based on the use of knowledge about the image object, the data domain, universal physical, logical and mathematical laws, and the objectives, methods and means of analysis.

6. The concept of the path of a problem, which reduces the selection of the synthesis of transformations

in the image recognition problem to the solution of a recognition problem with standard information.

The entire theory of recognition at present as well as the philosophy of the algebraic approach and of the descriptive approach to images this theory is based on the desire to regularize the selection and synthesis by learning algorithmic procedures of data transformation and analysis to solve information problems whose algorithms are unknown. For all their functional and object diversity, these problems have some basic features which make them a new class of data transformation problems, independent and exceptionally interesting in theoretical terms. The process of solution in these problems involves two stages: reduction of input data to a standard form (e.g. the synthesis of a formalized description) and transition from this standardized description to the space of generalized estimates which form the basis for a judgement about an object or phenomenon "on the whole". They may include the concept of similarity between standardized descriptions to be manipulated in decision making. Precedents can be specified to be used to adjust an algorithm to the problem in the process of learning. Problems arise where it is impossible (due to limited theoretical conceptions) or impractical (due to high costs or limited resources) to formulate normal mathematical models. The input information in these problems is bad due to the nature of the problem and its improvement eliminates the problem in the given formulation anyway. Note once again that these are problems about which too little is known to make it possible or advisable to use classical methods of solution (models) but still enough for a solution to be probable.

Historically, these problems were associated with attempts to use the experience of a skilled expert to solve so-called "problems with a poor structure". The latter generally comprised diagnostic, classification or prediction problems. In present terms, such problems are referred to as those of development of expert systems but one should not forget the great amount of work done in this connection in the area of heuristic programming during the 1950s and 1960s. Some papers of that period devoted to expert systems of zero generation led to some interesting conclusions. It was found that, generally speaking, problems with "bad" and even contradictory information could be solved in practice, and heuristic algorithms (empirical axiomatics) could be applied if it was proved by experiment that they were acceptable. It was also proved that all attempts to create universal methods for the solution of these problems within the framework of traditional mathematical logic led to algorithmic insolubility, i.e. to exhaustive solution methods. In fact, the advantages of heuristics are precisely that they allow one in particular cases to make efficient use of the available information to reduce or order the exhaustive search.

Later on it was proposed as a solution to turn to extensions of classical mathematical logic and mathematical statistics by introducing techniques which could allow for unreliability and contradictoriness of information and the heuristic character of the inference rule. A classical method of mathematical cybernetics - the estimation method - was adopted for the purpose. Any empirical axiom and heuristic inference rule included in the expert system was associated with a measure (estimate) that characterized its validity. The validity estimates were associated with inference chains in just the same way.

Thus, the first method of extension involves structuring of input information, the selection of systems of empirical axioms and heuristic inference rules, juxtaposition of nongeneral mathematical axioms and validity estimation rules, estimation of heuristic inferences and their results as a whole, and taking decisions from the estimates.

From those years till the mid-1970s, other extension methods made great progress. They were discussed in detail in this Chapter. It suffices to remember that the methodology of the second type of extension includes the selection of main heuristic structures, construction of an appropriate model on the basis of a heuristic, parametrization of the latter and transition to formulation and solution of optimization problems using the heuristic model.

As a result of all those activities, a great number of all kinds of heuristic algorithms was accumulated. Each of them was formulated as a rule in some partial language, and was not accompanied by instructions or recommendations for conditions of the use of or information about the relationship with other heuristic algorithms. This set of algorithms can be regarded as a basis for algorithmic knowledge bases of modern expert systems. While the expert systems of generation zero have fixed algorithmic bases, this set of heuristic algorithms can be regarded as a collection of experimental points each of which is a result of application of one heuristic algorithm. This approach makes possible a new method for structuring heuristic procedures to transfer to a unified language for the description of these procedures. This transfer is realized through parametrization of the description of algorithms, which in turn permits one to place an algorithmic envelope on the heuristics - experimental points - (to specify classes of algorithms differently - by varying values of parameters of descriptions of algorithms) and solve optimization problems on such parametric models.

When the introduction of a single description method enables any algorithm to be represented as a procedure of transformation of structured (i.e. reduced to a standard form) information into a matrix of estimates of a standard type, a new structuring (extension) by specifying operations with matrices will become possible. In this case we will be able to do without optimization in the parametric model and to synthesize new algorithms in the form of polynomials over basic algorithms associated with addition, multiplication, and multiplication by a scalar value.

Thus the new area of research is a systematic study of algorithmic knowledge bases of expert systems, which involves automation of the synthesis of new algorithms using expert quality functionals. Basic elements of this area are contraction of the space of original algorithms, development of techniques for a unified description of algorithms, structuring on the basis of isolation of major algorithms and indication of the radius of stability and order in a local neighborhood (as done, for example, in the theory of local algorithms [15]), and structuring on the basis of the algebraic approach, which is a continual extension of a finite number of basic algorithms. Another branch comprises methods and means of formalization in data transformation problems in extending the space of acceptable information, mainly for the case of specification of input information in the form of images.

REFERENCES

1. Aizerman, M.A., Braverman, E.M., Rozonoer, L.I., *Metod Potentsialnykh Funktsii v Teorii Obucheniya Mashin* (The Method of Potential Functions in the Theory of Machine Training), Nauka, Moscow, 1970.

2. Bongard, M.M., *Problema Uznavaniya* (The Problem of Recognition), Nauka, Moscow, 1967.

3. Vaintsvaig, M.V., *Kora Algorithms of Pattern Recognition Learning in Algoritmy Obucheniya Raspoznavaniyu Obrazov* (Algorithms of Pattern Recognition Learning), a collection of articles, ed. by Vapnik, V.N., Sovetskoe Radio, Moscow, 1973, pp. 110-116.

4. Vapnik, V.N., *Vosstanovlenie Zavisimostei po Empiricheskim Dannym* (Reconstruction of Relationships from Empirical Data), Nauka, Moscow, 1979.

5. Vapnik, V.N., Chervonenkis, A. Ya., *Teoriya Raspoznavaniya Obrazov* (Pattern Recognition Theory), Nauka, Moscow, 1974.

6. Gorelik, A.L., Gurevich, I.B., Skripkin, V.A., *Sovremennoe Sostoyanie Problemy Raspoznavaniya. Nekotorye Aspekty* (The Present State of the Recognition Problem. Some Aspects), Radio i Svyaz, Moscow, 1985 (Cybernetic).

7. Grenander, U., *Lektsii po Teorii Obrazov* (Lectures on Pattern Theory) in 3 vols., translated from English and ed. by Zhuravlev, Yu. I., Mir, Moscow, 1979 - 1983.

8. Gurevich, I.B., *Image Analysis Using the Method of Reversible Algebraic Closure in Problemy Iskusstvennogo Intellekta i Raspoznavaniya Obrazov* (Problems of Artificial Intelligence and Pattern Recognition), abstracts of reports and presentations at the Scientific Conference with the Participation of Researchers from Socialist Countries, Kiev, May 13-18, 1984, Section 2: Pattern Recognition, V.M. Glushkov Institute of Cybernetics, Kiev, 1984, pp. 41- 43.

9. Gurevich, I.B., *The Algebraic Approach to Image Analysis and Recognition in Matematicheskie Metody Raspoznavaniya Obrazov* (Mathematical Methods of Pattern Recognition), abstracts of presentations at the All-Union Conference in Dilizhan, May 16-21, 1985, published by the Academy of Sciences of the Armenian SSR, Erevan, 1985, pp. 55-57.

10. Gurevich, I.B., *Determination of the Class of Estimation Algorithms from Two-Dimensional Information for Image Recognition Problems in Metody i Sredstva Obrabotki Graficheskoi Informatsii* (Methods and Means of Graphic Data Processing), an interinstitute collection of papers, ed. by Vasin, Yu.G., Gorky University named after N.I. Lobachevsky, Gorky, 1986, pp. 47-66.

11. Gurevich, I.B., *Problems of Image Recognition, Raspoznavanie, Klassifikatsiya, Prognoz. Matematicheskie Metody i ikh Primenenie* (Recognition, Classification, Prediction. Mathematical Methods and their Application), yearbook, ed. by Zhuravlev, Yu. I., Nauka, Moscow, 1988, issue 1, pp. 280-329.

12. Gurevich, I.B., Zhuravlev, Yu.I., Minimization of Boolean Functions and Efficient Recognition Algorithms, *Kibernetika*, 1974, no. 3, pp. 16-20.

13. Gurevich, I.B., Zhuravlev, Yu.I., *On Formalization of the Principles of Selection of Image Analysis Algorithms in Metody i Sredstva Obrabotki Slozhnostrukturirovannoi Semanticheski Nasyshchennoi Informatsii* (Methods and Means of Processing Semantically Rich Graphic Data with a Complex Structure), abstracts of presentations at the 1st All-Union Conference in Gorky, September 1983, Gorky University, Gorky, 1986, pp. 20-21.

14. Dmitriev, A.N., Zhuravlev, Yu.I., Krendelev, S.P., *On Mathematical Principles of Classification of Objects and Phenomena in Diskretnyi Analiz* (Discrete Analysis), collection of articles, Institute of Mathematics SO of the USSR Academy of Sciences, 1966, issue 7, pp. 3-15.

15. Zhuravlev, Yu.I., Local Algorithms of Computation of Data, I-II, *Kibernetika*, 1965, no. 1, pp. 12-19; 1966, no. 2, pp. 1-11.

16. Zhuravlev, Yu.I., *Extreme Problems that Arise in Validation of Heuristic Procedures in Problemy Prikladnoi Matematiki i Mekhaniki* (Problems of Applied Mathematics and Mechanics), Nauka, Moscow, 1971, pp. 67-75.

17. Zhuravlev, Yu.I., *Algorithms of Construction of Minimum Disjunctive Forms for Functions of Logic Algebra in Diskretnaya Matematika i Matematicheskie Voprosy Kibernetiki* (Discrete Mathematics and Mathematical Questions of Cybernetics), vol. 1, ed. by Yablonsky S.V. and Lupanov, O.B., Nauka, Moscow, 1974, pp. 67-98.

18. Zhuravlev, Yu.I., Non-Parametric Problems of Pattern Recognition, *Kibernetika*, 1976, no. 6, pp. 93-123.

19. Zhuravlev, Yu.I., Correct Algebras Over a Set of Inaccurate (Heuristic) Algorithms, I-III, *Kibernetika*, 1977, no. 4, pp. 14-21; 1977, no. 6, pp. 21-27; 1978, no. 2, pp. 35-43.

20. Zhuravlev, Yu.I., *On the Algebraic Approach to the Solution of Problems of Recognition and Classification in Problemy Kibernetiki* (Problems of Cybernetic), a collection of articles, Nauka, Moscow, 1978, issue 33, pp. 5-68.

21. Zhuravlev, Yu.I., *On Algebraic Methods in Problems of Recognition and Classification, Raspoznavanie, Klassifikatsiya, Prognoz. Matematicheskie Metody i ikh Primenenie* (Recognition, Classification, Prediction. Mathematical Methods and their Application), yearbook, ed. by Zhuravlev, Yu.I., Nauka, Moscow, 1988, issue 1, pp. 9-16.

22. Zhuravlev, Yu.I., Zenkin, A.A., Isaev, I.V., Koltsov, P.P., Kochetkov, D.V., Ryazanov, V.V., Problems of Recognition and Classification with Standard Learning Information, *Zhurnal Vychislitelnoi Matematiki i Matematicheskoi Fiziki*, 1980, vol. 20, no. 5, pp. 1294-1309.

23. Zhuravlev, Yu.I., Nikiforov, V.V., Recognition Algorithms Based on Computation of Estimates, *Kibernetika*, 1971, no. 3, pp. 1-11.

24. Ladley, R.S., *Programmirovanie s Ispolzovaniem Tsifrovykh Vychislitelnykh Mashin* (Programming Using Digital Computers), translated from English, ed. by Kitov, A.I., Mir, Moscow, 1966.

25. Leontiev, V.K., Smetanin, Yu.G., On Reconstruction of Vectors from a Collection of Their Fragments, *Doklady Akad. Nauk SSSR*, 1988, vol. 302, no. 6, pp. 1319-1322.

26. Mazurov, V.D., Kazantsev, V.S., Beletskii, N.G., Krivogonov, A.I., Smirnov, A.I., *Questions of Validation and Application of Committee Recognition Algorithms, Raspoznavanie, Klassifikatsiya, Prognoz. Matematicheskie Metody i ikh Primenenie*, yearbook, ed. by Zhuravlev, Yu. I., Nauka, Moscow, 1988, issue 1, pp. 114-148.

27. Matrosov, V.L., Correct Algebras of Limited Capacity Over Sets of Inaccurate Algorithms, *Doklady Akad. Nauk SSSR*, 1980, vol. 253, no. 1, pp. 25-30.

28. Matrosov, V.L., Correct Algebras of Limited Capacity Over Sets of Estimation Algorithms, *Zhurnal Vychislitelnoi Matematiki i Matematicheskoi Fiziki*, 1981, vol. 21, no. 5, pp. 1276-1291.

29. Rastrigin, L.A., Erenshtein, R.Kh., *Metod Kollektivnogo Raspoznavaniya* (The Method of Collective Recognition), Energoizdat, Moscow, 1981.

30. Tou, J., Gonzalez, R., *Printsipy Raspoznavaniya Obrazov* (Principles of Pattern Recognition), translated from English, ed. by Zhuravlev, Yu.I., Mir, Moscow, 1978.

31. Tiukhtin, V.S. *Teoriya Avtomaticheskogo Opoznavaniya i Gnoseologiya* (The Theory of Automatic Recognition and Epistemology), Nauka, Moscow, 1976.

32. Chegis, I.A., Yablonskii, S.V., *Logical Methods of Control of Electric Circuits in a collection of articles on mathematical logic and its applications to some questions of cybernetics*, ed. by S.V. Yablonskii, Transactions of V.A. Steklov Institute of Mathematics under the USSR Academy of Sciences, published by the USSR Academy of Sciences, Moscow, 1958, vol. 51, pp. 270-36

33. Hume, D., *The Treatise on Human Nature, Collection of Works in 2 vols.*, vol. 1, Mysl, Moscow, 1966.

34. Becker, W., *Recognition of Patterns Using the Frequencies of Occurrence of Binary Words*, Springer Verlag, Wien-New York, 1978.

35. Pappentin, F., Kruger, M., Binary Sequences. 3. Complexity versus Homogeneity and Symmetry, *Information Sciences*, 1983, no. 31, pp. 33-39.

36. Watanabe, S., *Pattern Recognition: Human and Mechanical*, John Wiley and Sons., Inc., New York, 1985.

Yuri I. Zhuravlev, Dr. Sc. (Phys-Math), professor, corresponding member of the USSR Academy of Sciences, was born Jan. 14, 1935 in Voronezh, USSR. In 1957, he graduated from the Department of Mathematics and Mechanics, Moscow University, in information science, was a post-graduate student at Moscow University from 1957 to 1959 and was awarded the degree of Cand. Sc. (Phys-Math) in 1959. The subject of his thesis was "Normal Forms of Logical Functions". He worked as a section head and then as Deputy Director at the Institute of Mathematics, Siberian Branch of the USSR Academy of Sciences, in Novosibirsk from 1959 till 1969, and taught as a senior lecturer and then professor at Novosibirsk University from 1962-1969. He was awarded the degree of Dr. Sc. (Phys-Math) for the thesis "Local Algorithms for Computation of Information" in 1965. Starting 1969, he worked as a laboratory chief, then became a section head and is now Deputy Director of the USSR Academy of Sciences Computer Center in Moscow.

He is also a professor at the Moscow Physico-Technical Institute.

His work is principally concerned with mathematical logic, algebra, discrete optimization, pattern recognition, and the use of mathematical methods and computer technology for the solution of application problems for data processing and estimation in research automation, industry, medicine, geology, and sociology.

He is the author of over 150 publications in information science and applied mathematics, editor-in-chief of the *Pattern Recognition and Image Analysis* journal and the yearbook "Raspoznavanie, Klassifikatsiya, Prognoz. Matematicheskie Metody i ikh Primenenie" (Recognition, Classification, Prediction. Mathematical Methods and their Application) (Nauka Publishers, Moscow), and a member of editorial boards of international and Soviet journals on information science and applied mathematics.

Yu.I. Zhuravlev was awarded the USSR Lenin Prize for studies in discrete mathematics (1966) and the

USSR Council of Ministers Prize for a solution of application problems using pattern recognition methods (1986).

Igor B. Gurevich, Cand. Sc. (Phys-Math), Leading Researcher of the Scientific Council of the USSR Academy of Sciences on Cybernetics, was born in 1938 in Moscow. He graduated from the Moscow Power Engineering Institute and was awarded the degree of Cand. Sc. (Phys-Math) in 1975 in mathematical cybernetics. He is principally concerned with pattern recognition, and recognition, analysis and understanding of images. He has written two monographs (with co-authors), two textbooks on recognition (with co-authors) and over 80 papers. He is Scientific Secretary of the Information Science section of the Scientific Council of the USSR Academy of Sciences on Cybernetics, Deputy Editor-in-Chief of *Pattern Recognition and Image Analysis*, Managing Editor of the yearbook "Raspoznavanie, Klassifikatsiya, Prognoz. Matematicheskie Metody i ikh Primenenie" (Recognition, Classification, Prediction. Mathematical Methods and their Application) (Nauka Publishers, Moscow), and a member of the editorial board of *GVGIP: Image Understanding* (Academy Press, Inc.).