The Development of an Algorithmic Model of Object Recognition Using Visual and Sound Information Based on Neuro-fuzzy Logic

Sabina Shahbazzade

Abstract The chapter considers the problem of recognition of visual and sound information by constructing a virtual environment, which allows a qualitatively simple system to carry out experiments and create an algorithmic model of pattern recognition comparable to human capabilities. The aim of the research is to obtain an algorithmic model that can extract from the surrounding world, “meaningful” (visual and sound) objects to link with the relevant lexical concepts which are the atomic building blocks of intelligence. The research is dedicated to the development of machine intelligence with the phased increase in the complexity of the behavioral model of artificial personality (AP), with the goal being experimental research in the problem of artificial intelligence.

1 Introduction

This chapter is devoted to the creation of an algorithmic model of a system potentially capable of eventually recognizing visual 2D and 3D objects, and sound information in a complex artificial virtual environment.

A virtual environment is a space in which objects of recognition are placed from the library of objects (3D models and corresponding textures) and sound fragments. In this environment an artificial person imitating human behavior is placed.

The time in a virtual environment is calculated independently of the actual elapsed time out of system which allows one to create an artificial personality (AP) unlimited computing power.

Numerous experiments carried out on standard visual and sound objects, showed that in the recognition of visual objects or natural speech, there is some

S. Shahbazzade (✉)
Baku, Azerbaijan
e-mail: sh.sabina86@gmail.com

L. A. Zadeh et al. (eds.), *Recent Developments and New Directions in Soft Computing*, Studies in Fuzziness and Soft Computing 317,
DOI: 10.1007/978-3-319-06323-2_12,

threshold detail comparable with the size of the object. Threshold detail is characteristic of the optimization, proper selection, which provides maximum information content with minimal interference, and, accordingly, the deviation from this value greatly reduces the reliability of recognition results.

In fact, the main thing is not the specific details of object, but the sufficient volume required for the recognition of information, contained in a given area, and the geometric dimensions of objects, which in each case is different. This geometric area which includes the necessary and sufficient volume of characteristic information is called a Unitary Square (US).

US for visual objects are measured in units of space, but sounds are measured in units of time. US for visual information represent some geometric area and for sound information it represents duration of the sound signal, including the whole or a large part of the visual object (or the sound object) is sufficient for its identification or classification.

2 Methods of Local Focus, the Visual and Sound Information

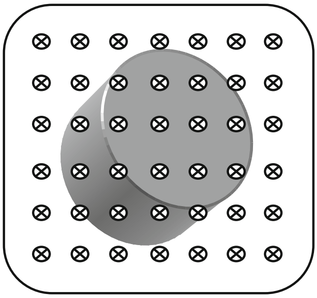
Research and experiments have shown that for that system the only available method for identification is the method of local focus, representing methods of imitation of the behavior of natural intelligence.

The parameter of the depth of the scene is very important in order to parse the scene for visual objects, due to the fact that relative to some of observers for each object corresponds to its own focal length. The essence of the method is a series of sets of measurements of focal lengths, the levels of which indicate an approximate map of surface of the visual object [4–6]. If the focal length increases sharply, then either the given area is out of the boundaries of the object or there is a hole in the object. In any case, within a finite amount of time, the system generates a map of the focal lengths and the actual image of this area (in the future image-contour).

The stage of recognition of visual information is a function of calculating the similarity of objects on which the system is already trained and information received by the image-contour of the target object. The algorithm of the similarity function is based on comparing, with a small level of allowable error, focus maps of compared objects [18, 23] (Fig. 1).

Parsing and recognition of sound objects is based on properties of sound information, which consist due to the fact that relative to some listeners of each sound, information in a certain time period corresponds to the proper frequency area [15].

Figure 2a shows the frequency density of a sound fragment with the word "rock" (circled in red oval), obtained from the appropriate unitary area (Fig. 2b) spoken by female voice with background noise. If the word was spoken by a male voice, then the fundamental frequency density would be in the area circled in the blue square. If several clusters of densities are recognized, the procedure is carried out on each of them regardless of their form [2, 7]. The longer the sound



∞	∞	∞	∞	∞	∞	∞
∞	∞	2	2	2	∞	∞
∞	1.3	2	2	2	2	∞
∞	1.3	2	2	2	2	∞
∞	1	1.3	1.6	1.8	∞	∞
∞	∞	∞	∞	∞	∞	∞

Fig. 1 Illustration of the application of a method of local focuses

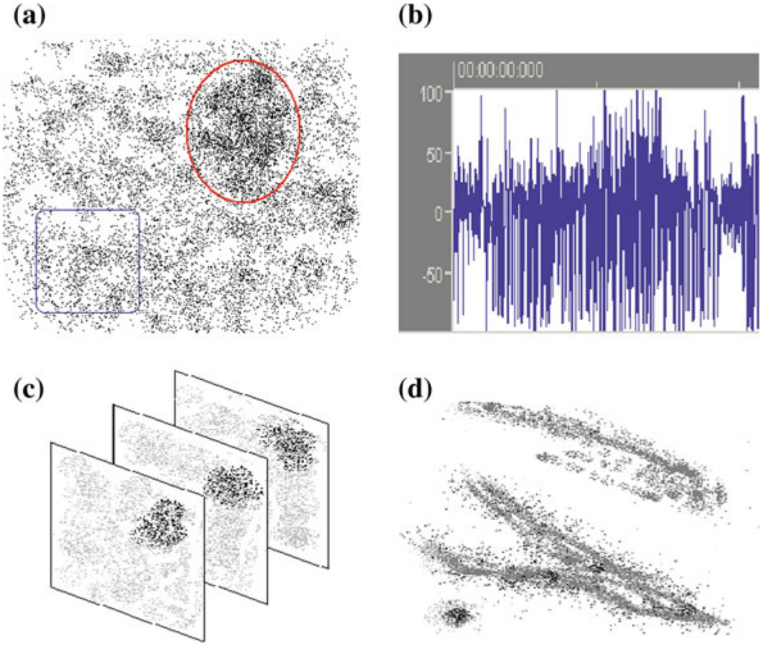


Fig. 2 **a** Frequency density of the word “rock” with a lot of background noise; **b** The unitary square (US) of sound object “rock”; **c** An example of multiple frequency densities; **d** Frequency density of non-verbal information

information, the greater the density of data that is negatively for the quality of analysis. Therefore, to achieve 25–30 % coverage is necessary to add a new bitmap (Fig. 2c). Analysis of the frequency density of artificial sounds has characteristic appearances in the form of dense points, lines or waves. Figure 2d shows analysis of symphonic musical instruments, some forms of violin strokes are represented in the upper part of the figure, brass bifurcated is in the center of the figure, and drums are represented by the point at the bottom left corner.

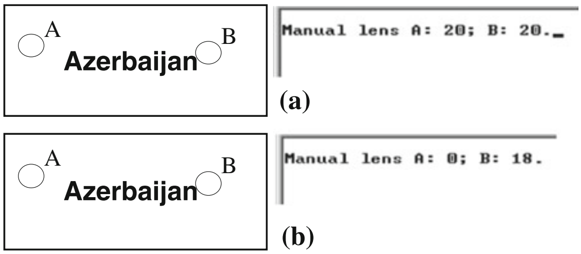


Fig. 3 Imitation of volume of flat geometric shapes: **a** usual 2d model, **b** 2d model with emulation third dimension

3 Recognition of Flat Objects (2D)

Classically presented sheet and painted on it geometric shapes are the flat objects (Fig. 3a) and, accordingly, the focal lengths of them are same, that does not allow creating a map of focal lengths in the case of 3-dimensional objects. To preserve the universality of the recognition subsystem of visual information, the problem was solved by an imitation of the third dimension for flat objects.

If one does not take into account the particular solution that provides almost 100 % results, the only method of teaching which gives satisfactory results is the method of converting the flat form of geometric shapes into 3D. The sheet on which the figure is drawn is a distant background (In Fig. 3b, focal distance of the point A is 0, i.e. as far as possible).

This approach makes it possible to produce a clear analysis of visual objects, but a steady recognition method has not yet been found. For example, about 65 % accuracy is attained in recognizing letters and about 75 % accuracy is attained in word recognition.

4 Learning and Self-Learning

An education system and quality control of procedure for pattern recognition is performed using the base previously prepared of training models. On the whole education system is based on modules:

- Decision-making process under the experimental conditions;
- The mechanism of formation of a database object recognition;
- Subprogram control achieved during the learning outcomes.

The four fundamental learning stages are illustrated below:

1. Object preparation—the creation of two-dimensional models for photography and three-dimensional models for visual objects;
2. Selection of lexical concepts—an important intellectual stage. The researcher should make a list (possibly with one element) of lexical concepts that in his opinion is consistent and is closely associated with a new object [17];

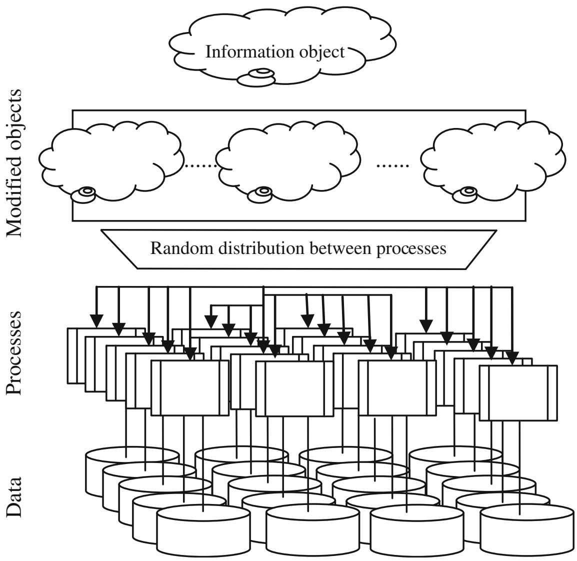


Fig. 4 Carrying out of learning—a stage of storing of object

3. Learning process—learning is a special process in which the system is proposed to recognize the new object of information. This is usually performed in a room in which only one new object is located at any given time. During this phase of the experiment, after selecting an object from the environment and at the end of stage recognition, the system receives a list of lexical concepts for corresponding new object;
4. Saving of object—In this critically important stage, quality implementation is crucial for sustainable functioning of the system as a whole. Saving of the object is a function that creates about 20 copies of the object, each of which has some minor distortion of the geometric dimensions of objects and their orientation, and color distortion, including black and white image. After receiving the modified copies and the original, the system randomly distributes between 1 % of the available system of modular processes and saves them in the relevant database [11].

The basis of self-learning is the idea of self research by AP, using available visual and sound information, to parse and save object recognition as nameless patterns (Fig. 4) [24].

Procedurally, self-learning is carried out in the following stages:

1. An unknown object (system has not been previously trained to this object) is placed in the virtual room and loading the procedure of recognition;
2. If during the recognition of a target object, recognition accuracy of less than 50 % is achieved, or an uncertain controversial situation with 70 % reliability occurs, then the object is stored under the nameless marker;
3. The same object is placed in the virtual room again and re-start the procedure of recognition;

4. If the result of recognition of the target object gives reliability of recognition similar to the corresponding nameless marker, comparable to 70 % and above, and it does not create controversy between other objects, then the object is considered as identified;
5. Otherwise, the process must revert to the second stage.

Upon the successful identification of the target object in 8–12 cases for the same nameless marker, the system is allowed to ask the question “What is it?” and operator introduces the lexical concepts, and it becomes a full marker.

The system asks the operator a question about the meaning of the lexical concept recognized objects, to obtain the largest possible number of synonyms. Few there are objects which may have more than 12 synonyms, other objects may be only 2–3 synonym. To solve this problem, was picked up by an algorithm which, when receiving a large number of identical synonyms limited to eight questions that would have on the one hand reduce the burden on operators, on the other hand, eliminate or reduce the possibility of false labeling.

If the marker already occurs in the system, then the set of visual or sound examples of the object joins an existing set [22].

With sound objects there is an additional uncertainty, which requires the intervention of a researcher in the learning process associated with the uncertainty of the main group of frequencies carrying the information. For recognizing sound information, indication of the frequency range mostly required for 5–8 first objects of information that define a library of groups of sound frequencies used in the sequel to defined of voice information.

Basically, for system important the only verbal information, which has a fairly clear system of formation and it is well identified.

As nameless lexical concepts using simple numeric codes (0001, 0002, 0003,…). It is quite feasible, since the lexical concepts play a minor role in the recognition process, which allows the researcher to monitor the correct functioning of the system [9].

5 Method of Classification of Information Objects

As described above, visual and sound objects are stored in bundles with the lexical concepts exercising the function of the associative marker. The marker is the cornerstone of ensuring the viability of the method of classification.

In fact, the lexical concepts (a marker of information objects) are the links of a higher order of intelligence, which are based on the atoms of information—words of their native language, in which every person thinks [1].

As the number of information objects associated with the markers, it is definitely a system of classification of self-organization, with this classification as learning becomes more complex. In this case, regardless of the number of related markers, the quality of self-organization does not get worse, since the formation of bonds is a natural character. [6].

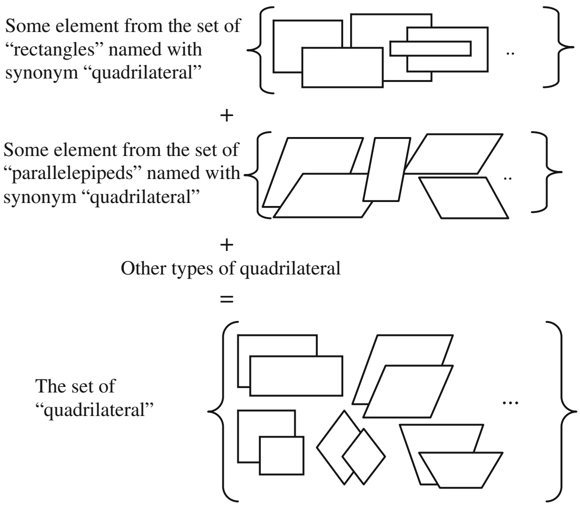


Fig. 5 Creating a virtual set of "quads" as an example of self organization of objects

All visual and sound information enters into the system as equal objects and as learning becomes associative, groups (bundles/ligament: the lexical concepts ↔ object), the growth of the number of objects of information will lead to the fact that some markers (A and B) will refer to each other like objects (A↔B), but at the set of marker (A) is more extensive than (B) (for diversity rather than on quantity, A↔C,…), therefore, the marker (B) is a subclass to the marker (A). In such a way, a hierarchy of classes and subclasses is formed (Fig. 5).

For example, it is assumed that the system is able to freely distinguish three types of quadrilaterals: squares, rectangles and parallelepipeds. Then into the further learning in the system was introduced the concept of a quadrilateral in combination with a square, a rectangle and a parallelepiped. A consequence of learning is to create a virtual (as nowhere is anything is created), parent class of quadrangles covering the child classes (markers) of squares, rectangles and parallelepipeds that is perfectly valid in terms of geometry. In addition, one can see that the class (marker) square could already be a child of the rectangle, if in the set of rectangles there is an object recognizable as the square [8].

6 Resolving a Dispute Situation: Hybrid Method of Decision-Making

One of the main tasks of the hybrid subsystem of decision-making is the task of improving the quality of recognizable images. [3] Decision made by the additional features of the image-analysis test circuits and reference objects with a fuzzy inference according to Fig. 6 [14].

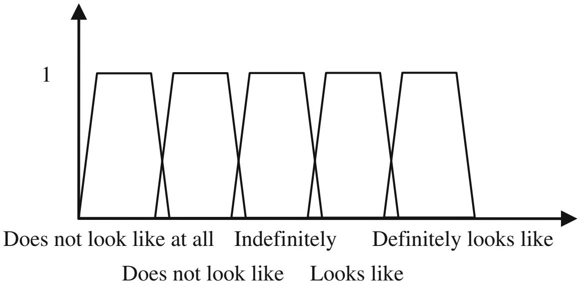


Fig. 6 Fuzzy representation of the results of pattern recognition

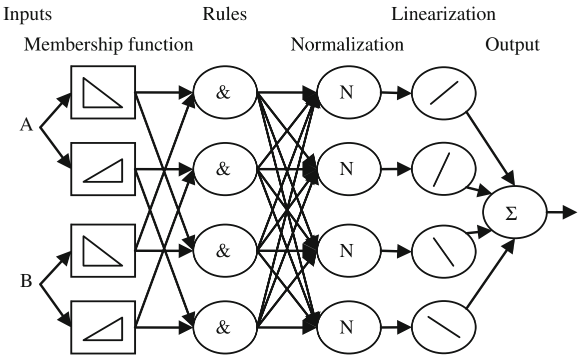


Fig. 7 The structure of applied ANFIS Network

To obtain satisfactory results of the decisions has been selected well-proven fuzzy neural network, which is built on the principle of Adaptive Neural Fuzzy Inference Systems (ANFIS), shown in Fig. 7 [13].

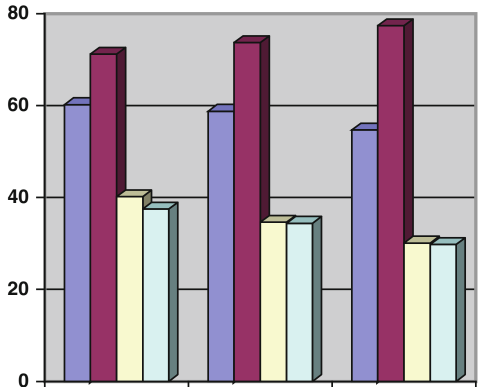
Decision-making procedure is repeated until the system is not sure what that one lexical marker characterizing the images as similar to test image, with the condition that the gap between the images of other lexical markers will be more than 10–15 % [16].

In general, a hybrid method of decision-making [26] produces the potential with equal chances, of selecting one object information from the three options, improving the quality of decisions made by approximately 10–15 %.

Thus, if the current average level of reliability in decision-making is around 70–75 %, neuro-fuzzy logic [20] will support this level at a drawdown to 55–60 % and it can be considered of quite satisfactory quality.

The necessity of application of the additional solving device is possibly due to the occurrence of disputable situations in which reliability of the most probable object of the information (A) differs from the second for probability (B) less than on 10 %.

Generally this means that the probability of needing a second solving device is relatively low and the determining factor is the presence of a 10 % area of the most likely one [25].



	1	2	3
Square	60,2	58,8	54,7
Rectangle	71,2	73,7	77,4
Parallelepiped	40,2	34,6	30,1
Trapeze	37,5	34,4	29,8

Fig. 8 Sequence from the three procedures to improve level of recognition of rectangle

Dynamic generation of the two (or more if needed) software implementations of fuzzy neural networks and their learning back propagation algorithm is the basic method of direct supervised learning of multilayer fuzzy neural directivity networks [12].

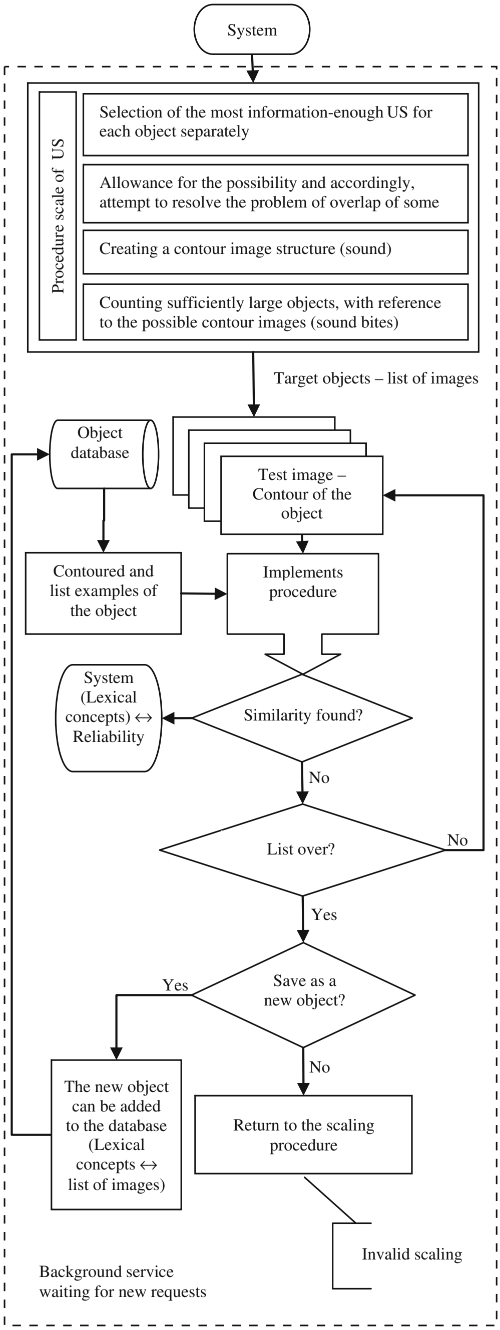
The size of a fuzzy neural network is determined by the volume of the object information. The main difference in application of fuzzy neural networks in the given work, is that each fuzzy neural network has learned only one of the patterns it is researching.

A feature of application is the absence of a priori given information in any form (usually put in logic of work of systems), and also the separation of procedure of decision-making at occurrence of disputable situations on two stages:

- One fuzzy neural network learning object (A), but the second object (B), while recognized object to the learning is not involved.
- At the end of training and the first and second fuzzy neural networks, the recognize objects send to their input, and holding for comparing of object proximity relevant of the visual or sound samples. The results of both of tests are applied to the second stage.
- The second stage is the processing of the fuzzy neural images as a of result of the reaction to the etalon object to the most probable objects of information that are transferred to the chart of implements in the form of specific fuzzy variables.

According to these values obtained by applying the algorithm of fuzzy comparisons {Definitely looks like $(+0.3, +0.5) = 10\ \%$; Looks like $(+0.1, -0.3) = 8\ \%$; Indefinitely $(-0.1, +0.1) = 4\ \%$; Does not look like $(-0.3, -0.1) = 0\ \%$; Does not look like at all $(-0.5, -0.3) = -3\ \%$}, choosing the closest indicator functions that are added to the competing object information (lexical concepts) (Fig. 8).

Fig. 9 Organizational block diagram of service modules in an infinite cycle of recognition



If after the decision making [21] the most likely object increases the gap with its competitors to a safe level (10–15 %), then the system decides in favor of the most probable object of information and accordingly, the associated lexical concepts, otherwise it is considered that the match object is not found and the algorithm passes to the interrupted work (search, record, etc.).

7 Algorithmic Model of Functioning

In Fig. 9, the operation of the system in standard conditions of the experiments. This block diagram of represents an algorithm of the main server, perform the coordination task management of the system modules. The organizational block diagram illustrates the overall cycle of functioning of the service modules that are involved in pattern recognition in an infinite cycle and in the case of successful recognition send results to the main service, and when new patterns are received automatically initiate a new task [10, 19, 20].

8 Conclusion

This article is the result of research and development of methods for recognition of object information and communication with the lexical concept without optimizing simplifications. It is dedicated to the research of creating of an automated system for the recognition of complex images and natural speech. When the learning is applied using a powerful computing system, the results will be able to recognize the visual and audio information, even at a high level of interference.

Suggested solution of the problem of allocation of objects of information from an overall picture and rather each other, is a method of local focuses for the visual information and a method of frequency density for the sound information is offered. The applied method is able to parse visual scenes and sound samples of high complexity, as well as to identify their constituent objects.

The algorithmic model of the functioning of the system is illustrated. Functional block diagrams describe the algorithm providing parse and recognition of the visual and sound information on the basis of architecture “the main service ↔ modular services”. The developed function of calculation of similarity of objects being the functional center of modular services, represents the unique algorithm which has shown satisfactory results and high stability of indicators of reliability.

The mechanism of a hybrid neuro-fuzzy method of decision-making reveals the occurrence of disputable situations. Defined the concept of the controversial situation, which is a consequence of issuing by the system of the two or more markers as a result recognition. Parameters and algorithms of functioning of hybrid system are also experimental factors which give satisfactory results on quality of accepted decisions.

The developed method of self-learning represents the mechanism of postponed learning, when the system learns to define the presence of new objects, self-learning, with the subsequent definition by the lexical information characterizing the given object.

The practical application of these scientific results is the ability to integrate visual and sound information around the lexical concepts that will be important in the further development of the system in the field of artificial intelligence.

References

1. Abakumov, V.G., Krilov, V.N., Antoshuk, S.G.: Detection and recognition of attributes of the objects using a spherical model of the visual analyzer. Electron. Commun. **8**(2), 211–212 (2000)
2. Belyavchev, V.G., Voskoboiniov, Y.E.: Local adaptive filtering algorithms for digital images. Sci. Bull. NQTU. **3**, 21–32 (1997)
3. Breiman, Friedman: Classification and Regression Trees, Olshen and Stone. Wadsworth and Brooks, Monterey (1985)
4. Dai, Y., Nakano, Y.: Recognition of facial images with low resolution using a Hopfield memory model. Pattern Recogn. **31**, 159–167 (1998)
5. Drouin, M., Abou-Kandil, H., Mariton, M., Duc, G.: Une nouvelle methode de decompozision-coordinasion 1 re partie: Principe et mise en ocuvre. APII. **3**, 205–226 (1985)
6. Duc, G., Drouin, M., Mariton, M., Abou-Kandil, H.: Une nouvelle methode de decomposition-coordination. 2 e partie: Application a la compensation des systemes multivariables. APII. **3**, 227–242 (1985)
7. Frei, W.: Image enhancement by histogram hyperbolization. Comput. Graph. Image Proc. **6**(3), 286–294 (1977)
8. Gonzalez, R.C., Fittes, B.A.: Gray–level transformation for interactive image enhancement. Mech. Mach. Theory. **12**, 111–112 (1975)
9. Gonzalez, R.C., Wintz, P.: Digital Image Processing, p. 505. Addison–Wesley, Reading (1987)
10. Gutta, S., Wechsler, H.: Face recognition using hybrid classifiers. Pattern Recogn. **30**, 539–553 (1997)
11. Hall, E.L.: Almost uniform distribution for computer image enhancement. IEEE Trans. Comput. **23**(2), 207–208 (1974)
12. Jacobsen, X., Zscherpel, U., Perner, P.: A comparison between neural networks and decision trees. Machine learning and data mining in pattern recognition. Lecture Notes in Artificial Intelligence, pp. 144–158 (1999)
13. Jang, J.S.R.: ANFIS: Adaptive-Network-Based Fuzzy Inference System. IEEE Trans. Syst. Man Cybern. **23**(5/6), 665–685 (1993)
14. Jang, J.S.R., Gulley, N.: The Fuzzy Logic Toolbox for Use with MATLAB. The MathWorks Inc, Natick (1995)
15. Lee, Y.: Handwritten digit recognition using K neares-neighbor, radial-basis function, and back-propagation neural networks. Neural Comput. **3**, 440–449 (1991)
16. Michie, D., Spiegelharter, D.J., Taylor, C.C (eds.): Machine Learning, Neural and Statistical Classification. Ellis Horwood, New York (1994)
17. Petrou, M.: Learning in pattern recognition. Machine learning and data mining in pattern recognition. Lecture Notes in Artificial Intelligence, pp. 1–12 (1999)
18. Ranganath, S., Arun, K.: Face recognition using transform features and neural networks. Pattern Recogn. **30**, 1615–1622 (1997)

19. Sergeev, V.V.: Application of a pattern recognition tasks in digital image processing. Autosummetry. **2**, 63–76 (1998)
20. Shahbazova, S.H.N., Suleymanov, M.R.: Stages of recognition probability images. In: Proceedings of AzTU (2004)
21. Shahbazova, S.H.N., Bonfig, K.V., Suleymanov, M.R.: Questions of construction of system of recognition of non-standard image and speech in the architecture of a hybrid vehicle based on neuro-fuzzy logic. In: 2nd international conference on application of fuzzy systems and software computing and analysis system, decision and control. Barcelona, Spain (2004)
22. Sabah, G.: Knowledge representation and natural language understanding. AI Commun. **6**(3–4), 155–186 (1993)
23. Valentin, D., Abdi, H., O'Toole, A.J., Cottrell, G.W.: Connectionist models of face processing: a survey. Pattern Recognit. **27**, 1209–1230 (1994)
24. Vetter, T., Poggio, T.: Linear object classes and image synthesis from a single example image. IEEE Trans. Pattern Anal. Mach. Intell. **19**, 733–742 (1997)
25. Yoon, K.S., Ham, Y.K., Park, R.-H.: Hybrid approaches to frontal view face recognition using the hidden markov model and neural network. Pattern Recogn. **31**, 283–293 (1998)
26. Zadeh, L.A.: Fuzzy sets and information granularity. In: Gupta, M., Ragade, R., Yager R. (eds.) Advances in Fuzzy Set Theory and Applications, pp. 3–18. Kluwer Acadamic Press, Amsterdam (1979)