Outcomes of Neural and Rule-based Classifiers as Criterions in Bi-objective Evolutionary Optimization of Feature Space in Pattern Recognition

KRZYSZTOF A. CYRAN
Institute of Informatics,
Silesian University of Technology
16 Akademicka Street, 44-100 Gliwice
POLAND

Abstract: Neural networks are widely used as classifiers in many pattern recognition problems because of good generalization abilities, what is a crucial issue in any practical application. However, vast majority of neural network architectures demands a huge computational effort for the training process, what in turn limits such solutions from application in one important domain of pattern recognition, which is the optimization of feature extractors. The optimization requires iterative computation of the objective function, and therefore such computation should be univocal and computationally effective. The paper presents how these postulates are satisfied in a special neural architecture called probabilistic neural network, which can therefore be effectively used for calculation of criterion in an evolutionary optimization of the feature space. As the experimental verification of proposed methodology, the optimization of Fraunhofer diffraction based pattern recognition system is presented, and compared with alternative solution, i.e. application of the rule-based classifier outcomes in the same role. The optimized system is a class of hybrid, fast, opto-electronic image recognizers, and the paper presents the use of it in the recognition of three different domains of images: recognition of the type of the vehicles, recognition of the type of the road obstacle in infrared wavelength, and the recognition of the class of the subsurface stress in the optical fiber. The experimentally obtained results confirm, that probabilistic neural network based main criterion used in a bi-criterion evolutionary optimization, outperforms the rough set based criterion in the mentioned systems.

Key-Words: neural networks, rough sets, bi-criterion optimization, evolutionary optimization, holographic ring-wedge detectors, image processing system, pattern recognition, infrared imaging, hybrid systems

1 Introduction

Many critical arguments have been claimed against neural networks since the birth of this area of information processing. The widely known book of Minsky and Papert (1969) almost stopped this new paradigm, reserving its application only to linearly separable problems. Even if this critique is known to be false in the case of multi-layer nonlinear networks, another important issue is raised by opponents. It is the computational effort needed for training, which in the case of nontrivial application can be very high.

Another, perhaps even more consequential, drawback of many neural network architectures is the lack of stop criterion and strong dependence of the training results upon the starting point in the space of weights. This is why, even if neural networks are widely used as classifiers in many pattern recognition problems, they have also strong limitations in many practical applications.

One of such areas, in the domain of pattern recognition, is the optimization of feature extractors. The optimization requires iterative computation of the objective function, and therefore such computation should be univocal and computationally effective, both postulates hardly fulfilled in the case of neural networks. However, one specific architecture of neural network, called probabilistic neural network (PNN) can be used in this role, as is shown in the paper below.

To illustrate the problem the pattern recognition system based on the sampling of Fraunhofer diffraction images is considered. The system has well established origins, published by Casasent and Song [2], and by George and Wang [3].

The optimization of feature extractor in such a system was first proposed by Cyran and Mrozek[4] and by Jaroszewicz et al. [5]. The criterion used in the evolutionary optimization was based on theory of rough sets. This theory, originated by Pawlak [6] was further developed by (among others) Mrózek [7, 8], Skowron and Grzymała-Busse [9] and by Ziarko [10] in so called rough sets with variable precision rough set model.

In the rough set theory, the notion of indiscernibility relation, has been modified by Cyran [11] to fit better the problem of criterion definition in the optimization of holographic ring wedge detector (HRWD). The latter element is used in a mentioned system as a feature extractor, and since it defines real-valued feature space the discrete in nature rough set notions, should be modified in order to be applied. However, even the modified version of rough set based criterion seemed to be sub-optimal, and therefore the search for another criterion was started. As the result, Cyran [12] applied PNN in the same role in two pattern recognition systems.

The current paper tries to compare the properties of both classifiers: rule-based (namely, rough set based in modified form) and the PNN-based, in the role of criterion for HRWD evolutionary optimization. The experimental part is done in three domains: the first domain is the recognition of the type of the vehicle, the second domain is the recognition of the road obstacle, based on the infrared images, and finally the third domain is the recognition of subsurface stress in an optical fiber.

We would like to notice that the type of images used for recognition in the three mentioned domains are so different, that the obtained results can be generalized to many other image recognition systems. The observation of results of bi-criterion evolutionary optimization with the main criterion defined as a PNN and a rough set based coefficient, called the consistency measure of the decision table, can be used for formulation the wider conclusion that the PNN, contrary to typical cases of neural networks, can be effectively used in the optimization of a feature space in a pattern recognition problems.

2 Images used in a study

As it was mentioned in the introduction, the images which were used in the study belong to three different domains. The first recognition system was dedicated for classification of the vehicles into two classes: below and above 3.5 tons. The division was dictated by the road monitoring requirements. Some of vehicles used in the recognition are illustrated in the Fig. 1.

Fig.1. Exemple of vehicles above (upper row) and below (lower row) 3.5 tons.

The second system had the purpose of recognition of infrared images taken in a low visibility conditions (like fog) into two classes: living creatures and other obstacles (mainly vehicles). The examples are shown in Fig. 2.

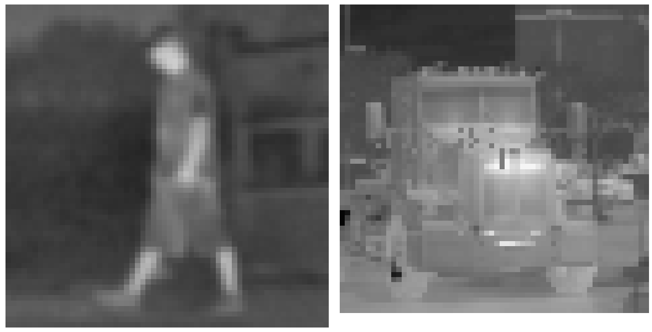


Fig. 2 Infrared images of living creatures: humans and animals or other obstacles (vehicles)

The third system considered is dedicated for the recognition of the subsurface stress in an optical fiber. The fiber is assumed to be fused in some material, and used as a detector of the stress in this material. The coherent light propagating in the fiber generates the speckle structures visible at the end of the fiber. The examples of such structures are given in the Fig. 3.

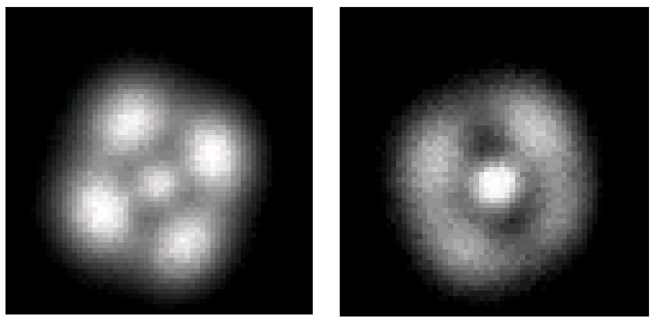


Fig. 3 Different speckle structures of the five modes propagating in the fiber (corresponding to various press conditions) can be used for the detection of the class on internal stress in the fiber.

The first two systems were to recognize two classes, whereas the purpose of the third system was to classify the input image into one of eight classes.

The good overall recognition abilities of the system in all three cases has already been reported in earlier works, dedicated for the specific systems (see for example [5, 11, 12]). The purpose of the current paper is to compare the results of the bi-objective optimization process with the use of PNN and consistency measure of decision table for all these systems. Both considered criterions are used as a main objective function in the optimization. The additional criterion in both cases is chosen as a normalized distance between one-class clusters in a feature space. The results of such evolution processes are presented in the next section.

3 Results and conclusions

The design of the presented system should be started with the optimization of a holographic ring-wedge detector. The evolutionary optimization procedure proposed in [3] has been further developed to become two-objective optimization.

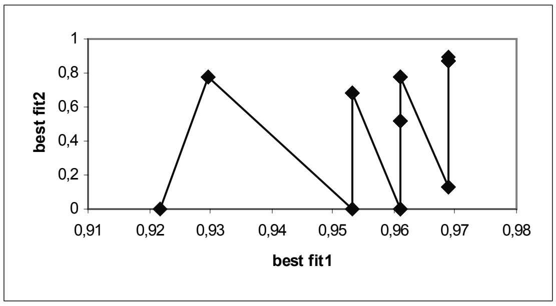
The first objective has been changed from the rough set based coefficient named the quality of approximation of the classification to the quality of classification done by probabilistic neural network with preset value of standard deviation. The second objective function remained the normalized distance between one-class clusters belonging to different classes in a feature space.

The evolutionary process was conducted for 1000 generations with 50 individuals in a population. The proportional selection was performed on the base of weighted average of two criteria (the main criterion was three times more influential as compared to the additional one).

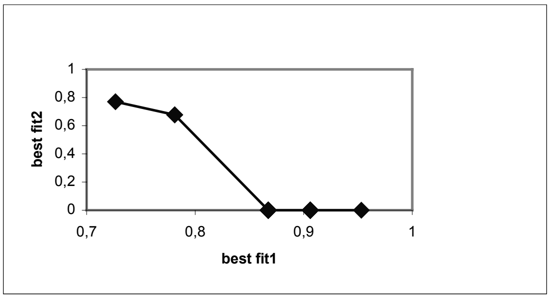
However, for the implementation of the elitist model, the lexicographic order scalarization was used with the main criterion standing at the first position. Such a choice made possible to observe the extreme mutual interaction of min and additional criteria, when the increase of the main objective leads to the minimum value of the additional criterion (this rule follows the search for Pareto-optimal solution).

In the case of PNN, the preset value of the standard deviation parameter had to be chosen. It had the same value in all experiments, after normalization with respect to the size of the input data. The results of the experiments are presented in the following figures. Figures 4 to 7 present the evolution of the best individual in a population with the use of the PNN and rule-based coefficient, indicated by (a) and (b) respectively.

The inspection of the phase planes (Fig. 4 and 6) clearly indicates that the criterion chosen as PNN – given in versions (a) of the figures – in general correlates positively with the second criterion, even if there are situations, when the increase of the main objective resulted in the minimization of the second one. These situations, however are not steady, and after a number of generations, the increase of the second criterion is observed. This in turn, results in the better chance for finding the better best value of the main objective. This is contrary to the generally negative correlation between the main and additional criteria, when the rule-based, consistency measure of the decision table is used as the main objective functions - see versions (b) of Fig. 4 and 6.



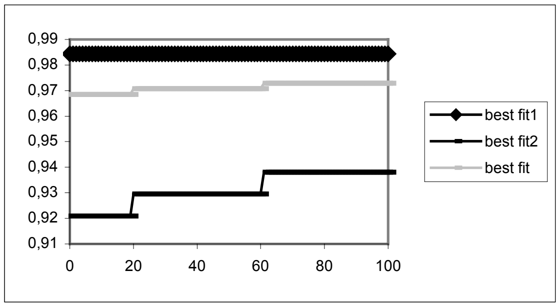
(a)



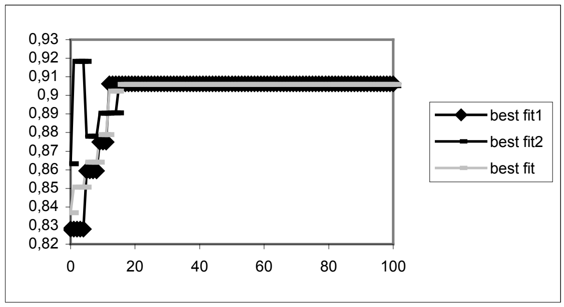
(b)

Fig 4. The phase plane of the main (best fit1) and additional (best fit2) criteria in the recognition of the vehicles problem

The Fig. 5 presents the outcome of the experiment with the recognition of the road obstacles – this case is difficult for the analysis, because the best value of the main criterion has been found in the firs generation, and the subsequent evolution led only to the increase of the additional criterion (and thus also the scalarized objective).

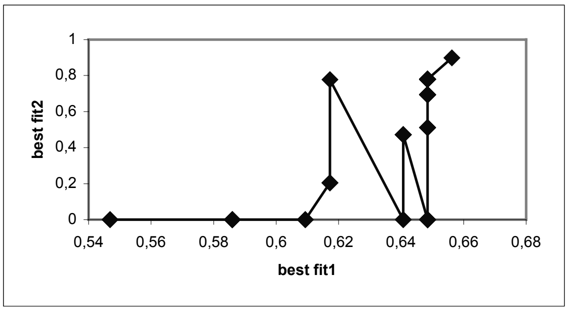


(a)

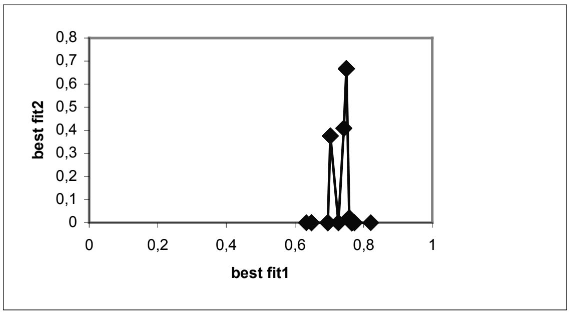


(b)

Fig 5. The course of the main (best fit1), additional (best fit2) and scalarized (best fit) objective functions over first 100 generations in a recognition of road obstacles in infrared light.



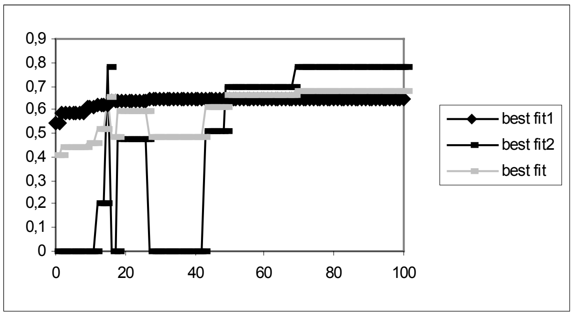
(a)



(b)

Fig 6. The phase plane of the main and additional criteria in the recognition of speckle structures

Fig. 7 presents basically the outcome of the same experiment as Fig. 6, however instead of plane phase, the time course of the evolution is shown. Comparison of the Fig. 6 and 7 gives more in-depth view on the type of mutual influence between main and additional criteria, both in version (a) and (b). Once more, the general negative correlation is visible in Fig. 7 in the (b) version, what results in the minimum value of the second criterion at the end of the evolution.



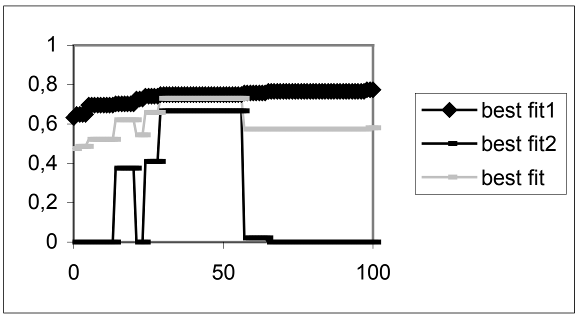
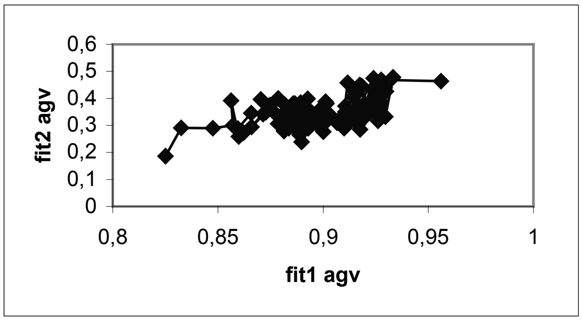
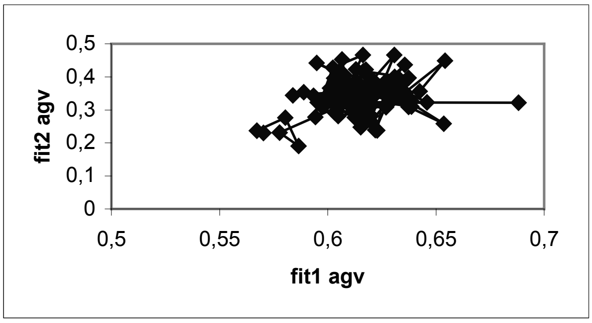


Fig 7. The course of the main (best fit1), additional (best fit2) and scalarized (best fit) objective functions over first 100 generations in a recognition of speckle structures.

Figures 8 to 10 present the evolution of the average individual in a population (contrary to figures 4 – 7, which were focused on the best individual) in all three experiments. The comparison of all versions (a) with all versions (b) indicates that the evolution with the PNN used as a main criterion - version (a) – is a one directional process, process towards better values of both criteria.

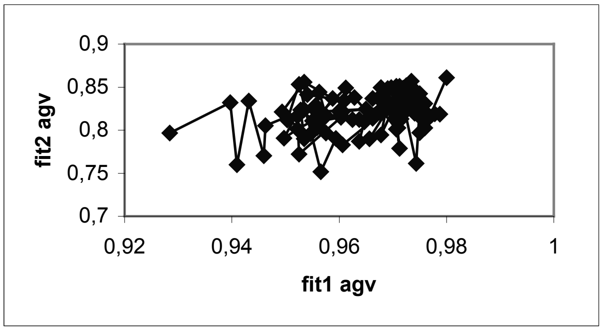


(a)

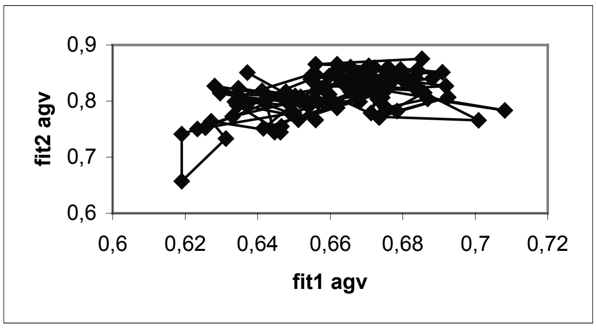


(b)

Fig 8. The phase plane of the main (fit1 avg) and additional (fit2 avg) criteria in the recognition of the vehicles problem



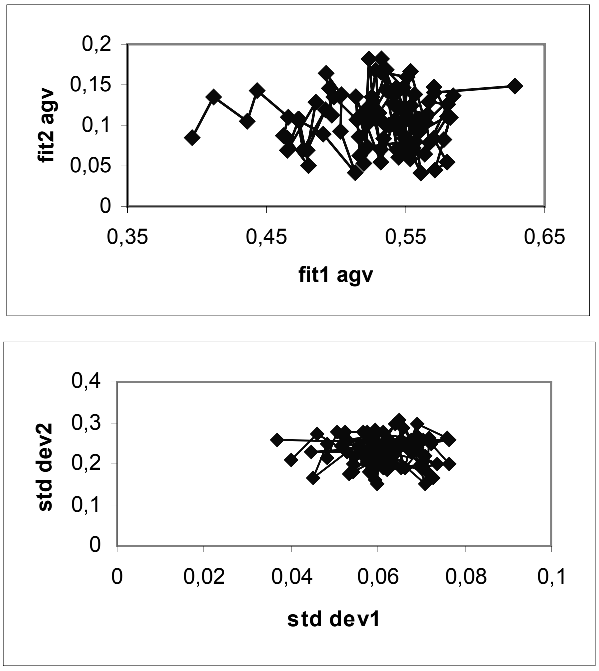
(a)



(b)

Fig 9. The phase plane of the main (fit1 avg) and additional (fit2 avg) criteria in the recognition of the type of road obstacle in infrared images

At the same time evolution with the rule-based coefficient used as a main criterion – versions (b) of Figures 8 to 10- represents the behavior similar to Brownian movements. In these cases, even the beginning and the end of evolution is hardly visible.



(a)

(b)

Fig 10. The phase plane of the main (fit1 avg) and additional (fit2 avg) criteria in the recognition of the speckle structures.

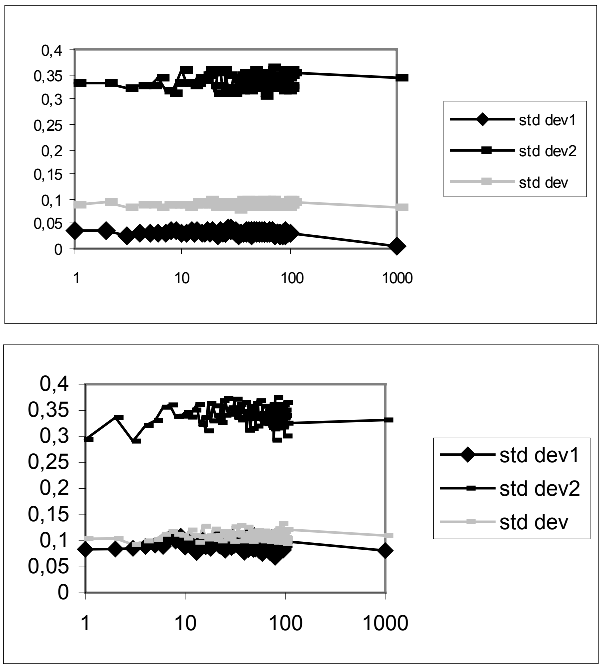


Fig 11. The course of the standard deviations of the main (std dev1), additional (std dev2) and scalarized (std dev) objective functions over 1000 generations in a recognition of vehicles. Notice that the time axis is in logarithmic scale.

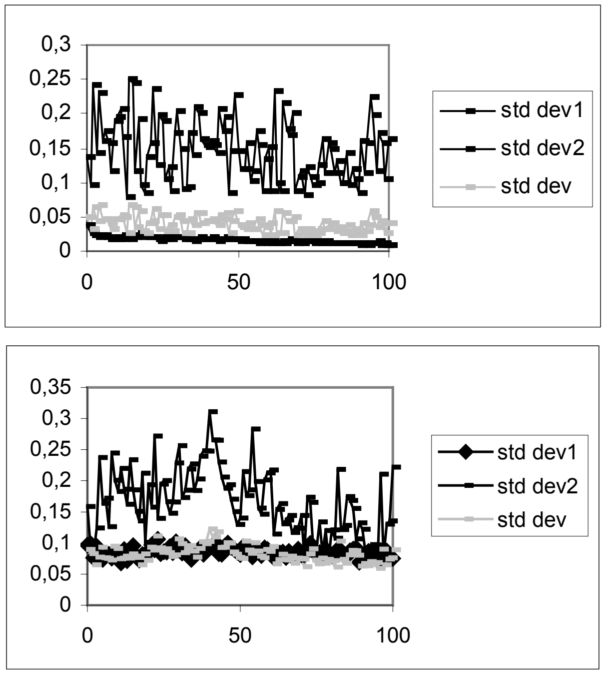


Fig 12. The course of the standard deviations of the main (std dev1), additional (std dev2) and scalarized (std dev) objective functions over first 100 generations in a recognition of the type of road obstacle in infrared light.

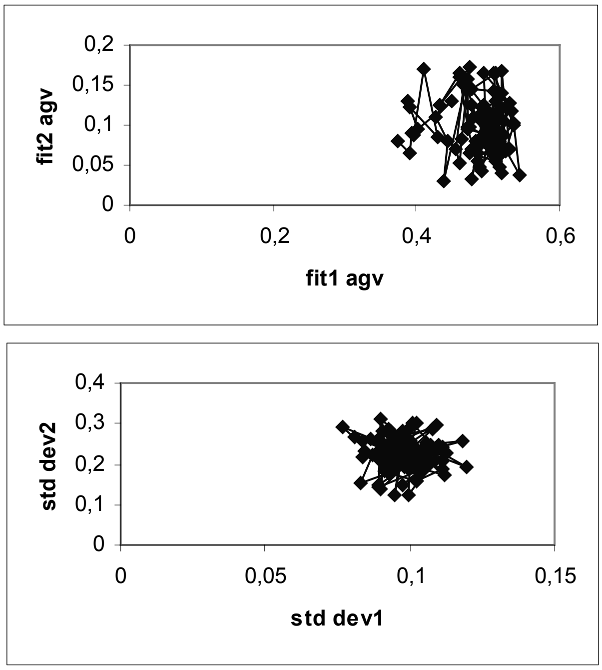


Fig 13. The phase plane of the standard deviation of the main (std dev1) and additional (std dev2) criteria in the recognition of the speckle structures.

Analysis of the Figures 11 to 13 (illustrating the behavior of the standard deviation of the objectives during the evolution in all three experiments) reveals that variation of the fit values is similar for PNN and rule based criterions, and in both cases much smaller than that of the additional criterion. This observation is easy for understanding, having in mind three times larger selection pressure for the main criterion, as compared to the additional one.

Summarizing the whole analysis, we would like to point out that the criterion based on the outcomes of probabilistic neural network with preset standard deviation parameter, has better properties with the conjunction with the second criterion (normalized distance between one-class cluster belonging to different classes) than the main criterion based on the rough set coefficient, called consistency measure of the decision table. The last conclusion remains true even if modified version of indiscernibility relation is used in rough sets, instead of standard one.

4 Acknowledgements

The author would like to acknowledge here the financial support of this work by Institute of Informatics, at Silesian University of Technology in the framework of the BW2006 activities.

References:

[1] M. L. Minsky & S. A. Papert, *Perceptrons* (Cambridge, MA: MIT Press), 1969.

[2] D. Casasent & J. Song, A computer generated hologram for diffraction-pattern sampling, *Proceedings of SPIE*, Vol. *523*, 1985, pp. 227-236.

[3] N. George & S. Wang, Neural networks applied to diffraction-pattern sampling, *Applied Opt*ics, Vol. *33*, 1994, pp. 3127-3134.

[4] K.A. Cyran & A. Mrózek, Rough sets in hybrid methods for pattern recognition, *International Journal of Intelligent Systems*, Vol. *16*, 2001, pp. 149-168.

[5] L.R. Jaroszewicz, K.A. Cyran, & T. Podeszwa, Optimized CGH-based pattern recognizer, *Optica Applicata*, Vol. *30*, 2000, pp. 317-333.

[6] Z. Pawlak, Rough *sets – theoretical* aspects *of reasoning about data*, London: Kluwer Academic Publishers, 1991.

[7] A. Mrózek, Rough sets in computer implementation of rule-based control of industrial processes, in: R. Słowiński (ed.) *Intelligent decision support. Handbook of applications and advances of the rough sets*, Dordrecht: Kluwer Academic Publishers 1992, pp. 19-31.

[8] A. Mrózek, A new method for discovering rules from examples in expert *systems*, *Man-Machine Studies,* Vol. *36*, 1992, pp. 127-143.

[9] A. Skowron & J. W. Grzymała-Busse, From Rough Set Theory to Evidence Theory. In: Yager RR, Ferdizzi M, Kacprzyk J. (eds): Advances in Dempster Shafer Theory of Evidence. John Wiley & Sons, NY, 1994.

[10] W. Ziarko, Variable Precision Rough Set Model, *Journal of Computer and System Sciences*, vol. 40, 1993, pp. 39-59.

[11] K. A. Cyran, Combining rough sets and neural approaches in pattern recognition, *Studia Informatica*, vol. 2(6), pp. 7-20, 2005.

[12] K. A. Cyran, Image processing system used for recognition of the type of vehicle and the road obstacles in visible and infrared light, *WSEAS Transactions on Information Science and Applications*, vol. 3(10), pp. 2021-2027, 2006.