

# Comparison of Artificial Neural Network and Rough Set based Classifiers Applied to a Hybrid Pattern Recognition System

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

*Abstract:* The paper compares two types of classifiers in a hybrid opto-electronic pattern recognition system. The first type is rough set based classifier operating in highly discretized feature space. This is the consequence of the granular nature of knowledge representation in the theory of rough sets. The second type is artificial neural network, which processes information taken from continuous feature space. The paper deals with the issues which arise when these two types of feature space coexist in one pattern recognition problem. In particular these issues are illustrated in the example of system used for recognition of speckle images of intermodal interference. In both cases the feature extraction is performed with the use of holographic ring wedge detector, generating the continuous feature space. This is the feature space natural for application of the artificial neural network in a classification subsystem. However, the optimization of feature extractor proposed in earlier papers, uses rough set theory, requiring the discretization of conditional attributes generating the feature space. Therefore such optimization is more suitable for rough set classifiers. Advantages and drawbacks of both solutions are presented in the paper. As the conclusion the new method of optimization of holographic ring wedge detector is postulated.

*Key-Words:* Image processing, pattern recognition, hybrid systems, neural networks, rough sets, holographic ring-wedge detectors

## 1 Introduction

The paper presents the comparison of two types of classifiers in a hybrid opto-electronic pattern recognition system. Hybrid image recognition systems have many advantages if compared with pure optical or pure electronic solutions. They perform heavy computations (like transforming into frequency domain or feature extraction) in optical mode, practically contributing no time delays. The post-processing of optical results, is performed in electronic devices, often with the use of artificial intelligence (AI) methods. Presented in the paper system is an example of such hybrid pattern recognizer working in spatial frequency domain obtained by means of Fraunhofer diffraction [1].

The main element in optical part of the system presented is the computer generated hologram (CGH), and was proposed by Casasent and Song [2]. CGH is the holographic version of commercially available ring wedge detector (RWD). The application of RWD into image recognizers were pioneered by George et al. [3, 4]. Combining Casasent's idea of holographic RWD (HRWD) with preliminary results of George and Wang, applying RWD as a feature extractor to neural network based classifier, gave theoretical basis for building the complete and useful image recognition system. Despite this completeness, the system was lacking possibility of

adaptation. These problems were caused by the lack of optimization methods.

The issue of how to find a good objective function for the optimization of HRWD was discussed by Cyran and Mrózek in [5]. Their methodology was based on the rough set (RS) theory started by Pawlak [6] and developed further by Mrózek [7, 8]. This methodology of optimization of HRWD was applied for image recognition problems by Jaroszewicz et al. [9].

The method of optimization of HRWD was applied to artificial neural network (ANN) based system used for recognition of the type of subsurface stress in materials with embedded optical fiber [10 - 12]. Another examples include the systems designed for the monitoring of the engine condition [13, 14]. The purely optical version of this recognition system was considered by Cyran and Jaroszewicz [15]. This fully concurrent system was limited by the development of technology of optically implemented artificial neural networks. The critical issue was the obtaining of nonlinear activation function applied after Stanford optical matrix-vector multiplier.

In presented above works the ANN based classifiers were applied but the optimization procedure in fact favored the RS based classifiers, due to the discrete nature of knowledge representation in theory of rough sets used for the definition of the objective function. The application of RS based classifiers was presented in [16, 17].

As seen from this introductory material, numerous applications of the system considered were already published, but these publications were not dealing with the problem considered in this paper. Here, the purpose is the discussion of some important issues appearing in methodology used for the design of the system described above. These issues are connected with the incompatibility of the type of feature space. Such incompatibility is inherently present between the optimization procedure and the classifier (in the case of ANN based classifier) and between the feature extractor and the classifier (in the case of rough set based classifier).

The comparison of advantages and drawbacks of these two types of classifiers with conjunction of the optimization procedure, led to the idea of modification of the optimization strategy. Postulated new optimization should work in continuous feature space. However, it requires modifications in the meaning of discernibility relation in the theory of rough sets.

## 2 Optical Feature Extraction

Image recognition is the process opposite to image generation process. Objects belonging to some classes  $C_i$  in some physical phenomena produce their images  $I_i$  (Fig. 1).

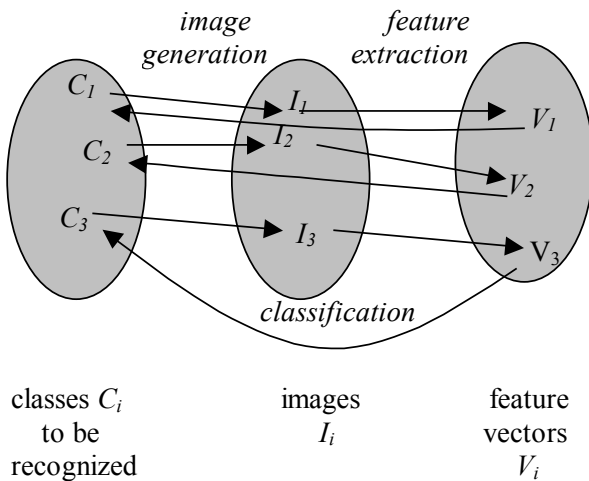


Fig. 1 Mappings present in image generation and recognition

The indirect approach, through the feature space, is favored due to the huge amount of information describing objects in image space. The feature space with reduced dimensionality describes images in more compact way, yet it should preserve all information required for the classification, being the mapping from feature space to space of classes (Fig. 1). The image

recognizer is therefore the system having processing the signal, as presented in the Fig. 2.

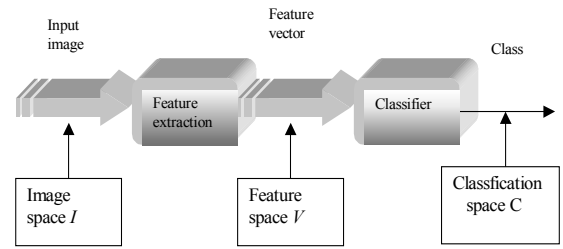


Fig. 1 Mappings present in image generation and

Hybrid opto-electronic solutions are systems composed of optical feature extractor and electronic (most often digital) classifier. One of the example is the system considered here, where the optical feature extractor uses HRWD element for integration of Fourier power spectrum over rings and wedges (Fig. 3). The Fourier spectrum is obtained by the Fraunhofer diffraction pattern brought by the spherical lens from infinity to back focal plane. The picture of optical setup is presented in Fig. 4. HRWD is a circular elements composed of rings and wedges, covered with the tiny grating (not visible in Fig. 3). Each region generates one feature, equal to the integral of the light intensity illuminating such region. Therefore the feature space is the  $N$ -dimensional space  $R^N$ , if  $N$  denotes the number of regions, i.e.,  $N$  is the sum of rings and wedges in HRWD.

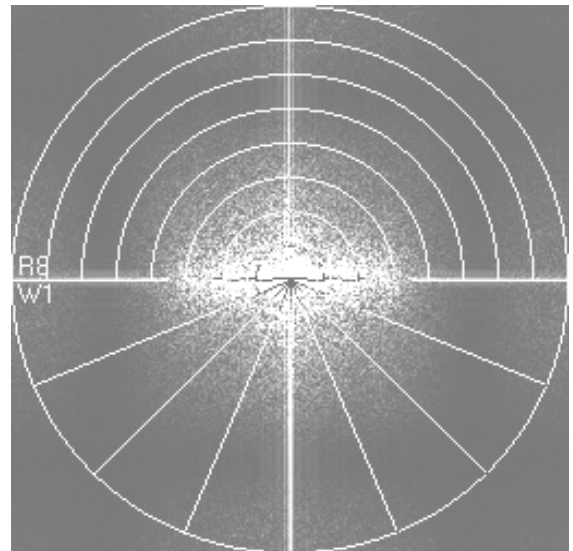


Fig. 3 RWD illuminated by Fourier power spectrum of the input image

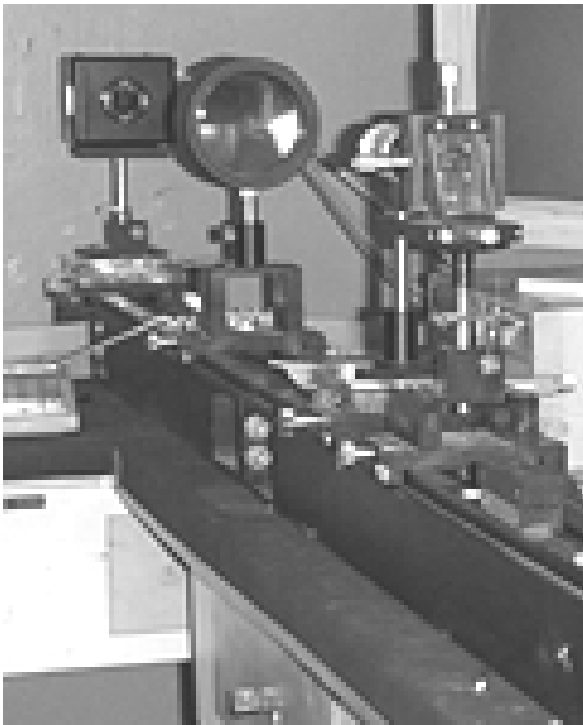


Fig. 4 Picture of the optical setup. The RWD is placed in back focal plane of the lens.

Below we present the application of AI methods, namely ANN and rough set theory to process the information from the output of optical part. The classification of relatively short feature vectors (composed of several numbers, corresponding to several regions of HRWD) are easy to perform in electronic machines.

### 3 Properties of Neural Classifiers

ANN are widely used for classification purposes. In our system, we looked for an architecture having reduced interference and thus extensive locality.

The interference in ANN is present when the learning in one point of the feature space causes the forgetting in other point of this space. This is certainly the undesirable phenomenon. ANNs which are less sensitive to interference are called spatially local networks [18]. So it is important to assure enough plasticity of the network, so it was able to learn new facts, and at the same time to remember old facts.

Formally, the interference is a measure of influence of learning at point  $x$  upon the mapping (from feature space to classification space) performed by ANN at point  $x' \neq x$ .

Let this mapping be defined by the equation  $y = f(\mathbf{x}, \mathbf{w})$ , where  $y \in [0, 1]$  denotes the output (associated with posterior probability of one class),  $\mathbf{x} \in X$  is the feature vector,  $\mathbf{w} \in W$  denotes the weight vector,

function  $f: X \times W \rightarrow [0, 1]$  is continuous mapping implemented by ANN,  $X \subset \mathcal{R}^N$  is the feature space, and  $W \subset \mathcal{R}^m$  is the space of weights. The purpose of learning of the network is to choose the weight vector  $\mathbf{w}$ , so the ANN approximates desired function  $y^* = f^*(\mathbf{x})$ . Let the learning algorithm has the general form:  $\Delta \mathbf{w} = \alpha \mathbf{H}(\mathbf{x}, \mathbf{w}, e)$ , where  $\Delta \mathbf{w}$  denotes the change of the weight vector,  $\alpha$  is the learning rate,  $\mathbf{H}$  is direction of the weight change and  $e = y - y^*$  is the error of approximation. Then the interference at point  $x'$  caused by learning at point  $x$  is denoted by  $I_{f, \mathbf{w}, \mathbf{H}}(x, x')$  and defined as [18]:

$$I_{f, \mathbf{w}, \mathbf{H}}(\mathbf{x}, \mathbf{x}') = \begin{cases} \lim_{\alpha \rightarrow 0} \frac{f(\mathbf{x}', \mathbf{w}) - f[\mathbf{x}', \mathbf{w} + \alpha \mathbf{H}(\mathbf{x}, \mathbf{w}, 1)]}{f(\mathbf{x}, \mathbf{w}) - f[\mathbf{x}, \mathbf{w} + \alpha \mathbf{H}(\mathbf{x}, \mathbf{w}, 1)]} & \text{if lim exists} \\ 0 & \text{else} \end{cases} \quad (1)$$

The interference is defined for given pairs of points in feature space. The locality  $L_{f, \mathbf{w}, \mathbf{H}, X}$  is the feature of the whole feature space and is computed as the reciprocal of average squared interference over the feature space:

$$L_{f, \mathbf{w}, \mathbf{H}, X} = \left[ \int_X \int_X I_{f, \mathbf{w}, \mathbf{H}}(\mathbf{x}, \mathbf{x}')^2 d\mathbf{x} d\mathbf{x}' \right]^{-1} \quad (2)$$

It is worth to point out that both radial basis functions (RBF) networks and multilayer perceptrons (MLP) may arbitrarily close approximate the posterior probabilities of recognized classes and be arbitrarily largely local if adequately large number of modifiable weights is assured.

### 4 Rough Set based Classifier

On contrary to neural networks, the rough set based classifiers work in the discrete space. The application of such classifier in the system considered is dictated by the HRWD optimization method. The criterion in this optimization has been chosen as the consistency measure of the decision table. This notion is defined in theory of rough sets.

Formally, the decision table  $T$  is an ordered 5-tuple:

$$T = \langle U, C, D, v, f \rangle, \quad (3)$$

where  $U, C, D$  are finite and nonempty sets. The set  $U$  is referred to as the universe. The elements of the universe  $U$ , in the decision table  $T$ , are the numbers of decision rules. Set  $C$  is the set of conditional attributes, and set  $D$  is the set of decision attributes. In the case of classification considered in the paper, the set  $D$  consists

of only one element  $d$ . Mapping  $v$  associates set  $V_q$ , called the domain of attribute  $q$ , with each element  $q \in C \cup D$ . Finally, the function  $f: U \times C \cup D \rightarrow V$ , where  $V$  is a union of sets  $V_q$ , is called the decision function. The decision function is defined in such a way, that  $f(x, q)$  means the value of attribute  $q \in C \cup D$  for each element  $x \in U$ . The decision table  $T = \langle U, C, D, v, f \rangle$  is often presented in tabular form, shown in the Table 1.

Table 1. The example of decision table  $T = \langle U, C, D, v, f \rangle$  for classification

Rule Number	Conditional attributes					Decision $d$
	$C_1$	...	$c_n$	...	$c_N$	
$I$	$v_{c_1 I}$	...	$v_{c_n I}$	...	$v_{c_N I}$	$v_{dI}$
...	...	...	...	...	...	...
$M$	$v_{c_1 M}$	...	$v_{c_n M}$	...	$v_{c_N M}$	$v_{dM}$
...	...	...	...	...	...	...
$M$	$v_{c_1 M}$	...	$v_{c_n M}$	...	$v_{c_N M}$	$v_{dM}$

Let us define the function  $f_x: Q \rightarrow V$ , as  $f_x(q) = f(x, q)$ . Such function represents single line in table 1. Formally, the decision rule in decision table  $T = \langle U, C, D, v, f \rangle$  is defined as a function  $g: C \cup D \rightarrow V$ , iff there exists in  $T$  an element  $x \in U$ , such that  $g = f_x$ . Therefore, each line in Table 1 represents the separate decision rule. It is evident, that each decision rule indicates also the class that should be recognized for given values of conditional attributes. Hence, the  $m$ -th decision rule of the decision table  $T$  can be uniquely associated with the if-then rule given below:

$$\text{if } (c_1 = v_{c_1 m} \ \& \ \dots \ \& \ c_n = v_{c_n m} \ \& \ \dots \ \& \ c_N = v_{c_N m}) \text{ then } d = v_{dm}$$

Since conditional attributes in a decision table must be discrete (or the decision table could not be used for other examples than those used for knowledge gathering), so the corresponding feature space is discrete. In our case each conditional attribute is discretized version of the feature obtained from HRWD region. This required discretization is highly non linear transformation of feature space, which can degrade the ability of recognition of the resulting system.

Each decision table can be decomposed into fully consistent table and table composed of contradicting rules. The consistency measure of the decision table  $T$  is then ratio of the number of decision rules in the fully

consistent table to the number of the decision rules in the table  $T$ . This coefficient can be used as a criterion in the method of optimizing the HRWD. The purpose is to obtain the decision table with consistency measure equal to one.

The formalism of rough sets gives also the tools for determining the minimal number of attributes that describe the process of the classification with the same accuracy as the original table  $T = \langle U, C, D, v, f \rangle$ . The notion of relative reduct is appropriate for this purpose. Furthermore, for any decision rule there exist some attributes that are necessary and unnecessary for classification. The notion of relative value reduct can be applied for determining these attributes. The problem of looking for relative reducts and relative value reducts can be simplified by using the discernibility matrices and discernibility functions.

### 5 Illustrative example

Considered here system we applied to recognition of the class of intermodal interference visible as the speckle structures (Fig 5). The layout and intensity of the speckles are dependent on the type of subsurface stress in the optical fiber illuminated by the coherent light from the laser.

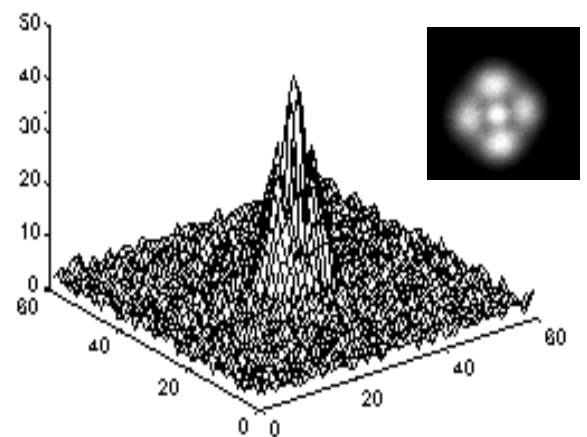


Fig. 5 Intermodal interference image taken from the output of optical fiber, and 3D plot of its power spectrum

The set of images (composed of 128 elements) was decomposed into two parts: training set (with  $N_t = 102$  images, i.e. 80% of 128) and testing set (with  $N_t = 26$  images, i.e. 20% of 128). The number of neurons  $N_i$  in input layer was equal to the number of features  $N$  obtained from the HRWD. The number of hidden neurons  $N_h$  was also equal to  $N_i$ , and the number of output neurons  $N_o$  was equal to the number of classes to be recognized  $K$ . As the learning rule the modified

backpropagation method was applied to sigmoidal neurons, whose operation is described by:

$$net_i^m = \sum_j w_{ij}^m O_j^m + \theta_i^m$$

$$O_i^m = \frac{1}{1 + \exp(-\beta net_i^m)} \tag{4}$$

In above formulae  $w_{ij}^m$  are weights associated with connection from neuron  $j$  to  $i$  (superscript  $m$  indicates in all cases the  $m$ -th step of the training),  $O_i^m$  and  $O_j^m$  are outputs from neuron  $i$  and  $j$  respectively,  $net_i^m$  is a network excitation and  $\theta_i^m$  represents the threshold value of neuron  $i$ . In the Table 2, presenting the results of the recognition on testing set, the following abbreviations are used:

- in the column indicated as HRWD, the value  $S$  indicates standard HRWD and value  $O$  indicates optimized HRWD,
- $B_d$  is the number of bad decisions for testing set
- $B_{dc}$  is the number of bad decisions if the competition among output neurons is applied,
- $E_d$  is a normalized decision error (without competition of output neurons) given by  $E_d = B_d / (N_t K)$
- $E_{dc}$  is a normalized decision error (after competition of output neurons) given by  $E_{dc} = B_{dc} / (N_t K)$

Table 1 Results of classification by ANN using standard and optimized HRWD

HRWD	$B_d$	$B_{dc}$	$E_d$ [%]	$E_{dc}$ [%]
$S$	7	5	3.4	2.4
$O$	2	1	1	0.5

The mask of optimized HRWD is presented in Fig. 6. Observe not equal sizes of rings and wedges as opposed to standard RWD (compare Fig. 6 with Fig. 1). The improvement of recognition abilities (Table 2) is almost 5-fold if optimized structure is used.

If RS based classifier is used then the normalized decision error  $N_d$  is equal to 1.7%, somewhat worse result as compared to results obtained with ANN based classifier. This is caused by the need of discretization of feature space for RS based classifier. However also system with ANN based classifier cannot be claimed as fully optimal. It is due to the need of discretization of feature space for obtaining the optimal structure of HRWD. The natural is the postulate to avoid this discretization, however then classical rough set theory

cannot be applied. Recently, the author proposed the modification of the discernibility relation, which makes possible to deal with continuous attributes in some domains of rough set theory. This modification uses the cluster analysis and is described in other works sent for publication.

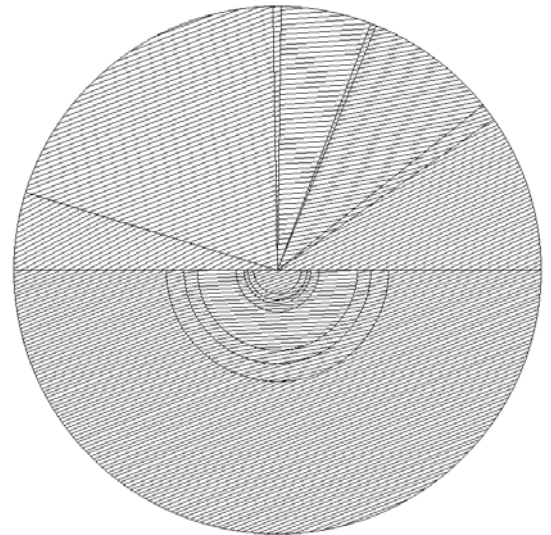


Fig. 6 Mask of optimized HRWD

## 6 Conclusion

In the paper two types of classifiers were considered. Rough set based classifier, being the example of the classifier operating in highly discretized feature space was slightly outperformed by the ANN based classifier. This is mainly the consequence of the granular. The paper focused on issues of coexisting of these two types of feature space in one pattern recognition problem. In both cases the feature extraction is performed by optimized HRWD generating the continuous feature space. However, the optimization of feature extractor used rough set theory, requiring the discretization of conditional attributes generating the feature space.

Such discretization transforms the search of optimal solutions in the discrete space into the search of sub optimal solutions in the continuous space. The resulting subsystem is therefore always sub optimal no matter whether ANN or RS based classifier is used. Even if sub optimal solution in this case is a few-fold better than standard one, the need of obtaining actually optimal HRWD structure can be postulated as the consequence of presented here analysis of continuous and discrete aspects present in the recognition system. Therefore to further improve the optimization of feature space dedicated for continuous-valued ANN classifiers Cyran (results in press) proposed the modification of the notion of discernibility relation, fundamental for rough sets

theory. This modification made possible to avoid, highly non linear transformation of feature space.

## 7 Acknowledgements

The author would like to acknowledge here the financial support of this work, performed partially in the frame of the grant no 5 T12C 005 25 sponsored by the Polish State Committee for Scientific Research (MNiI) and partially as statutory activities of the Institute of Informatics, Silesian University of Technology BK2005, sponsored as well by the Polish State Committee for Scientific Research (MNiI). Special acknowledgments are also for L. R. Jaroszewicz for supporting the author with the images of speckle structures used here as the illustrative example.

### References:

- [1] T. Kreis, *Holographic interferometry – principles and methods*, Berlin: Akademie Verlag Series in Optical Metrology, 1, 1996.
- [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 Optics*, Vol. 33, 1994, pp. 3127-3134.
- [4] N. George, S. Wang, & D.L. Venable, Pattern recognition using the ring-wedge detector and neural network software, *Proceedings of SPIE*, Vol. 1134, 1989, pp. 96-106.
- [5] K.A. Cyran & A. Mrózek, Rough sets in hybrid methods for pattern recognition, *International Journal of Intelligent Systems*, Vol. 16, 2001, pp. 149-168.
- [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] L.R. Jaroszewicz, K.A. Cyran, & T. Podeszwa, Optimized CGH-based pattern recognizer, *Optica Applicata*, Vol. 30, 2000, pp. 317-333.
- [10] K.A. Cyran, L. R. Jaroszewicz, & T. Niedziela, Neural network based automatic diffraction pattern recognition, *Opto-electronic Review*, Vol. 9, No. 3, 2001, pp. 301-307.
- [11] K.A. Cyran, U. Stańczyk, & L.R. Jaroszewicz, Subsurface stress monitoring system based on holographic ring-wedge detector and neural network, in: G.J. McNulty (ed.): *Quality, Reliability and Maintenance*, Bury St Edmunds, London: Professional Engineering Publishing. 2002, pp. 65-68.
- [12] K.A. Cyran, T. Niedziela, L.R. Jaroszewicz, & T. Podeszwa, Neural classifiers in diffraction image processing, *Proc. International Conf. On Computer Vision and Graphics*, Zakopane, Poland, 2002, pp. 223-228.
- [13] T. Podeszwa, L.R. Jaroszewicz., & K.A. Cyran, Fiberscope based engine condition monitoring system, *Proceedings of SPIE*, Vol. 5124, 2003, pp. 299-303.
- [14] L.R. Jaroszewicz, I. Merta, T. Podeszwa, & K.A. Cyran, Airplane engine condition monitoring system based on artificial neural network, in: G.J. McNulty (ed.) *Quality, Reliability and Maintenance*, Bury St Edmunds, London: Professional Engineering Publishing, 2002, pp. 179-182.
- [15] K.A. Cyran & L.R. Jaroszewicz, Concurrent signal processing in optimized hybrid CGH-ANN system, *Optica Applicata*, Vol. 31, 2001, pp. 681-689.
- [16] K.A. Cyran & L.R. Jaroszewicz, Rough set based classification of interferometric images, in: P. Jacquot & J.M. Fournier (eds.) *Interferometry in speckle light. Theory and applications*, Berlin, Heidelberg, NY: Springer, 2000, pp. 413-420.
- [17] K.A. Cyran, PLD-based rough classifier of Fraunhofer diffraction pattern, *Proc. International Conf. On Computer, Communication and Control Technologies*, Orlando, FL, 2003, pp. 163-168.
- [18] S. Weaver, L. Baird, & M.M. PolyCarpou, An analytical framework for local feedforward networks, *IEEE Transactions on Neural Networks*, Vol. 9, No. 3, 1998, pp. 473-482.