COMBINING ROUGH SETS AND NEURAL NETWORK APPROACHES IN PATTERN RECOGNITION

Krzysztof Cyran[1]

Abstract. The paper focuses on problems which arise when two different types of AI methods are combined in one design. The first type is rule based, rough set methodology operating is highly discretized attribute space. The discretization is a consequence of the granular nature of knowledge representation in the theory of rough sets. The second type is neural network working in continuous space. Problems of combining these different types of knowledge processing are illustrated in a system used for recognition of diffraction patterns. The feature extraction is performed with the use of holographic ring wedge detector, generating the continuous feature space. No doubt, this is a feature space natural for application of the neural network. However, the criterion of optimization of the feature extractor uses rough set based knowledge representation. This latter, requires the discretization of conditional attributes generating the feature space. The novel enhanced method of optimization of holographic ring wedge detector is proposed, as a result of modification of indiscernibility relation in the theory of rough sets.

Keywords: Pattern recognition, neural networks, rough sets, hybrid methods, evolutionary optimization, holographic ring-wedge detectors

1 Introduction

The paper presents two types of artificial intelligence (AI) methods applied to the design of a hybrid opto-electronic pattern recognition system. Such systems have many advantages compared to 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 by computers, remarkably often with the use of 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 is the computer generated hologram (CGH), proposed by Casasent and Song [2]. CGH is essentially the holographic version of commercially available ring wedge detector (RWD). The application of RWD elements into image recognizers were pioneered by George et al. [3, 4]. Combining Casasent's idea of holographic RWD (HRWD) with results of preliminary studies of George and Wang, who applied RWD as a feature extractor to neural network based classifier, gave theoretical basis for building the complete and useful

[1] Institute of Informatics, Silesian University of Technology, ul. Akademicka 16, 44-100 Gliwice, Poland

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]. Proposed by them methodology was based on the theory of rough sets founded by Pawlak [6] and developed further by (among others) Mrózek [7, 8], Ziarko [9] and Skowron and Grzymała-Busse [10]. Rough set based methodology of optimization of the HRWD element was applied for diffraction image recognition problems by Jaroszewicz et al. [11].

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 [12-14]. Another examples include the systems designed for the monitoring of the engine condition [15, 16]. The purely optical version of this recognition system was considered by Cyran and Jaroszewicz [17]. This fully concurrent system is yet limited by the development of technology of optically implemented artificial neural networks. The critical issue is the obtaining of nonlinear activation function applied to process data from Stanford optical matrix-vector multiplayer.

In above works the ANN based classifiers were applied, however the optimization procedure in fact favored the rough set based classifiers, due to the same, discrete nature of knowledge representation used both for the definition of the objective function and in a classifier. The application of rough set based classifiers is presented in [18, 19].

Now, the purpose of this paper, being the extended version of a conference paper [20] is to discuss some important issues appearing when combining rough sets and neural network based information processing types. These issues result from inherent incompatibility of the knowledge representation in rough set theory and neural network studies.

The comparison of advantages and drawbacks of these two types of AI methodologies led the author to the idea of enhancement in the HRWD optimization. Postulated new method works in continuous feature space and yet does not resign from rough set based formal apparatus. To address properly these issues, all what is needed, is an appropriate modification of the notion of indiscernibility in the theory of rough sets. It is discussed in the next section.

2 Modification of indiscernibility relation

The indiscernibility relation plays crucial role in methodology of knowledge processing described by rough set formalism. In its classical form it requires the discretization of real valued attributes, performed independently of each other. However, an analysis of notions present in the theory of rough sets proves that vast majority of them do not require this specific, classical form. The only actual (from theoretical point of view) demand is that this relation should be an equivalence relation, *i.e.* it should be reflexive, symmetric and transitive. From practical perspective, this relation should make indiscernible such objects which belong to the same cluster in $\Re^n$. However clustering in $\Re^n$ cannot be done separately for each attribute belonging to $\Re$.

In particular classical form of indiscernibility relation is not required by consistency measure of a decision table. This coefficient is of special interest here because it has been used as the objective function in evolutionary optimization of HRWD dedicated for multimodal distribution of classes in a feature space.

The reasons supporting the choice of such criterion have been considered in [5]. Despite they seem to be reasonable, application of classical definition of indiscernibility relation, makes the result always sub-optimal. With this motivation, let us introduce the modified version of indiscernibility relation in a formal way.

Let $S = <U, Q, v, f>$ be the information system composed of universe U, set of attributes Q, information function f, and a mapping v. This latter mapping associates each attribute $q \in Q$ with its domain V_q. The information function $f : U \times Q \to V$ is defined in such a way, that $f(x, q)$ denotes the value of attribute q for the element $x \in U$, and V is a domain of all attributes $q \in Q$, defined as a union of all domains of single attributes, *i.e.* $V = U_q \in Q, V_q$. Then each nonempty set of attributes $C \subseteq Q$ defines the classical version of indiscernibility relation $I_0(C) \subseteq U \times U$ of discrete attributes $q \in C$, as

$$\begin{aligned} &x \ I_0(C) \ y \quad \Leftrightarrow \\ &\forall q \in C, \ \ f(x,q) = f(y,q), \end{aligned} \tag{1}$$

where $x, y \in U$. If attributes are discrete this is very natural way of defining this relation. However if we originally deal with real valued attributes, then some kind of clustering and discretization of continuous attributes has to be performed before application of rough set theory. Let this process be denoted as a vector function $\mathbf{\Lambda}: \Re^{card(C)} \to \{1, 2, \ldots, \xi\}^{card(C)}$, where ξ is the discretization factor being the number of clusters covering the domain of the individual attributes $q \in C$. Analogously, let discretized value of any individual attribute $q \in C$ be the result of a scalar function $\Lambda: \Re \to \{1, 2, \ldots, \xi\}$. Then, the classical form of indiscernibility relation is defined as:

$$\begin{aligned} &x \ I_0(\Lambda[C]) \ y \quad \Leftrightarrow \\ &\forall q \in C, \ \ f(x,\Lambda[q]) = f(y,\Lambda[q]). \end{aligned} \tag{2}$$

However, the above form of the indiscernibility relation, proposed by classical theory of rough sets, as well as by its generalization named

variable precision model defined in [9], is not actually required for rough set based knowledge processing.
Let generalized version of the indiscernibility relation be dependent on a family (set) of sets of attributes. This allows us to introduce hierarchy of sets into originally unstructured set of attributes that the relation depends on.

Let $\mathbf{C} = \{ C_1, C_2, \ldots, C_N \}$ be such a family of disjoint sets of attributes $C_n \subseteq Q$, that unstructured set of attributes $C \subseteq Q$ is equal to the union of elements of the family $\mathbf{C}$. This means that $C = \bigcup_{C_n \in \mathbf{C}} C_n$. Observe that both $\mathbf{C}$ and C contain the same collection of single attributes, however $\mathbf{C}$ includes additional information about the structure of attributed. If this structure is irrelevant for the problem considered, it can be simply ignored and then we obtain, as a special case, the classical version of indiscernibility relation I_0. However it is possible to obtain also other versions of this modified relation for which the introduced structure is meaningful.

To satisfy practical requirement of proximity in a real valued space of indiscernible (in a rough set sense) objects, it is natural to define the modified indiscernibility relation $I_{\text{gen}}(\mathbf{C}) \in U \times U$ as

$$\begin{aligned} &x \; I_{\text{gen}}(C) \; y \quad \Leftrightarrow \\ &\forall C_n \in C, \quad \text{Clus}(x, C_n) = \text{Clus}(y, C_n), \end{aligned} \tag{3}$$

where $x, y \in U$, and $Clus\,(x, C_n)$ denotes the number of the cluster, that the element x belongs to. The two extreme cases of this relation are obtained when: (i) family $\mathbf{C}$ is composed of exactly one set of conditional attributes C, and (ii) when family $\mathbf{C}$ is composed of $card\,(C)$ sets, each containing exactly one conditional attribute $q \in C$.

The classical form I_0 of the indiscernibility relation is obtained as the second extreme special case of modified version I_{gen}, because then clustering and discretization are performed separately for each continuous attribute. Therefore, it follows, that

$$\begin{aligned} &I_0(\mathbf{\Lambda}[C]) \;\equiv\; I_{\text{gen}}(\mathbf{C}) \Leftrightarrow \\ &\mathbf{C} = \{\{q_n\}: C = \bigcup_{q_n \in C} \{q_n\}\} \\ &\wedge\, Clus\,(x, \{q\}) = f\,(x, \Lambda[q])\;. \end{aligned} \tag{4}$$

3 Application to hybrid pattern recognizer

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

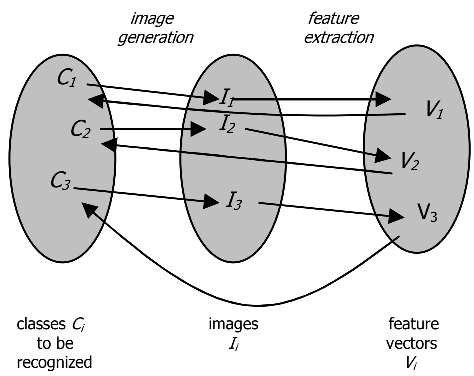

Figure 1. Mappings present in image generation and recognition

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. Hybrid opto-electronic solutions are systems composed of optical feature extractor and computer in the role of the classifier (Fig. 2).

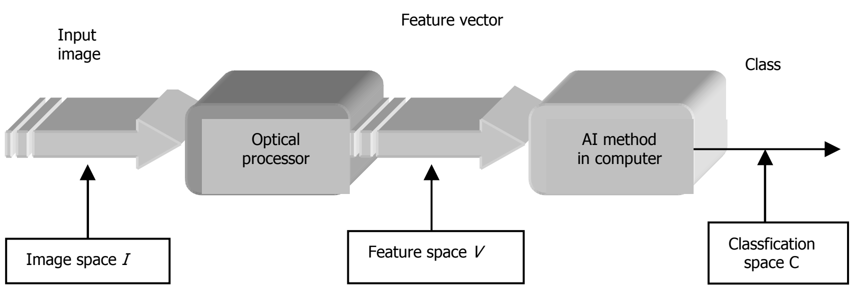

Figure 2. Opto-electronic image recognition system

One of the example is the system in which the optical feature extractor uses HRWD element for integration of Fourier power spectrum over rings and wedges (Fig. 3).

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

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.

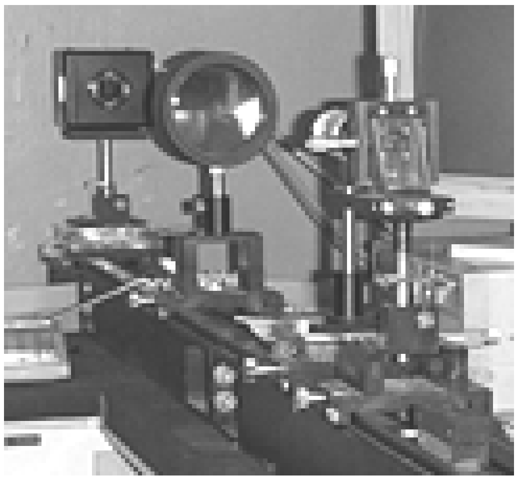

Figure 4. Picture of the optical setup.
The RWD is placed in back focal plane of the lens

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 $\Re^N$, if *N* denotes the number of regions*, i.e.* when N is the sum of number of rings and wedges in HRWD.

What is very important, is that features generated by rings are rotation and translation invariant, and features generated by wedges are size and translation invariant. These properties make possible to obtain the recognition system invariant with any of these transformation of input images.

The optimization of HRWD is a search for such sizes of rings and wedges, as well as their number, that the recognition abilities are maximized. For the measure of these abilities we used rough set based consistency measure of decision table $\gamma_C(D^*)$ with conditional attributes corresponding to rings and wedges, and decision attribute being the recognized class of the image. We argue that application of indiscernibility relation modified in (3) gives enhanced criterion as compared to criterion obtained with classical form of this relation (2) applied to real valued attributes. We demonstrate this by presenting optimization procedure and results of an experiment.

Since the defined above enhanced objective function is not differentiable, gradient-based search methods must be excluded. Therefore we optimize HRWD in a framework of evolutionary algorithm as presented in pseudo-code below:

```
t ← 1;
POPULATION ← Initialize;
Evaluate (Q);
ξ ← 2^Q;
do for x in POPULATION
  do for i = 1 to card (U)
    C_x[i] ← χ (image_i );
    d_x[i] ← C_j ;
  od;
  F_x ← Evaluate (γ_C (D*));
od;
do while (ξ ≥ NumOfClasses) and (t < MaxGenNum)
  FOUND ← FALSE;
  Select (POPULATION);
  Crossover (POPULATION);
  Mutate (POPULATION);
  Repair (POPULATION);
  do for x in POPULATION
    do for i = 1 to card (U)
      C_x[i] ← χ (image _i );
      d_x[i] ← C_j ;
    od;
    F_x ← Evaluate (γ_C (D*));
    if F_x = MaxValue then
      FOUND ← TRUE;
```

```
        xopt  ← x;
      fi;
    od;
    if FOUND then
      ξ ← ξ / 2;
    fi;
    t ←  t + 1;
  od;
```

In the above algorithm t is the generation number, $\mathbf{x}$ is the chromosome (representing the HRWD) in the POPULATION and $\mathbf{x_{opt}}$ is the chromosome representing genotype of the optimum HRWD. $C_x[i]$ are discrete conditions of decision rule *i* generated by HRWD for image $image_i$. Similarly $d_x[i]$ denotes the decision attribute of mentioned decision rule and C_j is the abstract class the image $image_i$ belongs to.

The algorithm has two flow control parameters: *MaxGenNum* (specifying maximum number of epochs for evolution) and *MaxValue,* indicating the maximum required value of the objective function. Normally *MaxValue* should be set to 1 – to obtain fully consistent decision table but sometimes this could be to strong demand – then one should reduce this parameter.

This algorithm resembles that applied for criterion calculated based on classical definition of indiscernibility relation. The difference is in the meaning of ξ parameter. Previously it was the discretization factor required by rough set theory, now it is the number of clusters in clustering procedure. This change influences the initial value of ξ and the termination of presented program. The initial value of ξ now is calculated as 2^Q for such minimum Q for which $\xi \geq Card\,(U)$. The program is stopped after achieving the maximum value of $\gamma_{\mathbf{C}}(D^*) = MaxValue$ for the value of parameter $\xi = NumOfClasses$ (where *NumOfClasses* denotes the number of classes to be recognized), as opposed to previous version, terminating when $\gamma_{\mathbf{C}}(D^*) = MaxValue$, for $\xi = 2$. As genetic operators, the proportional selection in elitist model, one point cross-over and uniform mutation were used. The repair algorithm was used to handle the constraints given by the possible structures of HRWD. As the result of operation of the algorithm the parameters describing optimized HRWD are obtained (they are phenotypic features encoded in chromosome $\mathbf{x_{opt}}$)

The system considered we applied to recognition of the class of intermodal interference visible as the speckle structures (Fig 5).

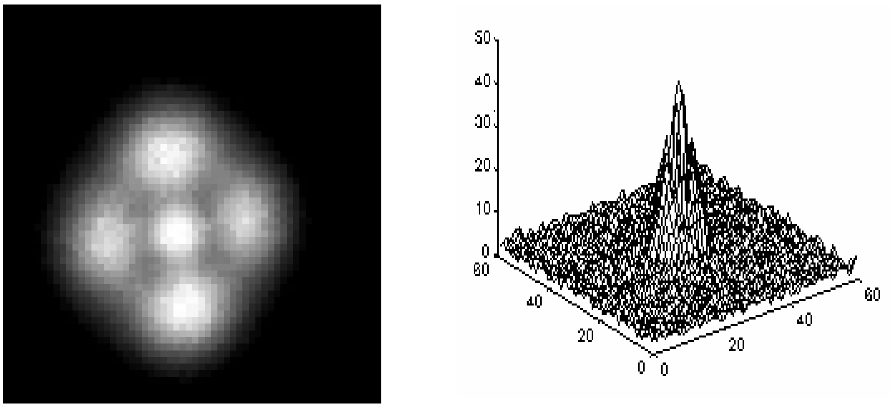

Figure 5. Intermodal interference image taken from the output of optical fiber, and 3D plot of its power spectrum

The layout and intensity of the speckle strucures are dependent on the type of subsurface stress in the optical fiber illuminated by the coherent light from the laser.

The experiments were conducted for a set of 128 images of speckle patterns resulting from intermodal interference occurring in optical fiber. The images taken from the outpu of the fiber belonged to 8 classes taken in 16 sessions. The training set was composed of 120 images taken out in 15 sessions. The testing set contained 8 images belonging to different classes, representing one session. The process of training and testing was repeated 16 times, according to jack-knife method, *i.e.* for each iteration another session was used for testing set, and all but one sessions were used for training set. This procedure gave good basis for reliable cross-validation with reasonably large number of images used for training.

The results of testing with probabilistic neural network (PNN) classifying images in feature space obtained from standard, optimized, and optimized with modified indiscernibility relation HRWDs, are presented in Table 1. One can observe that the normalized decision error (*NDE*) is small even for system with standard HRWD, indicating good overall properties of the presented diffraction pattern recognizer. However, *NDE* is further reduced when optimization is included, reaching its minimal value for optimization with proposed in (3) modification of indiscernibility relation.

The percentage of improvement with respect to system with standard HRWD, indicated by the coefficient *k* in Table 1 is 20% for optimization with classical definition of indiscernibility relation, and 25% for the modified version.

Table 1. Results of testing the classification abilities of the system. The classifier is a probabilistic neural network having Gaussian radial function with standard deviation s = 0.125. NDE and k stands for normalized decision error, and for percentage of decrease of

normalized decision error (with respect to results for standard HRWD), respectively.

System with:	*NDE* [%]	*k* [%]
Standard HRWD element	2.0	0
HRWD optimized with standard indiscernibility relation	1.6	20
HRWD optimized with modified indiscernibility relation	1.5	25

The increase of *k* from 20% to 25% does not look very impressive, however its proper interpretation requires referring to Fig. 6.

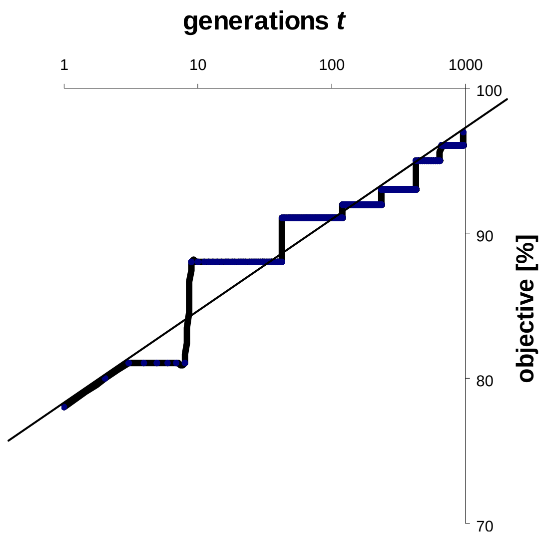

Figure 6. Evolutionary optimization presented in a horizontal log scale for number of generations t, reveals that (on average) the objective function increase is proportional to the exponent of t

Mentioned above figure presents the course of objective function in evolutionary optimization. What is striking, is the fact, that when the horizontal axis (denoting the number of generations) is drawn in a log scale, then the average increase of the objective is remarkably well mimicked by a straight line. This means that the objective grows like a logarithm of *t*, due to intuitively understandable fact, that optimization of already optimized solution becomes harder and harder.

The difficulty of finding better solution than solution already optimized, gives some flavor of the corresponding difficulty concerning the decrease of *NDE*, represented by the increase of k. In this light, the increase of k from 20 to 25% reflects the effect of significant improvement of the HRWD generated feature space. The computer generated mask of optimized HRWD element (Fig. 7) is obtained from its representing chromosome $\mathbf{x}_{\mathbf{opt}}$ by the application of decoding function and formulas defining gratings of HRWD regions.

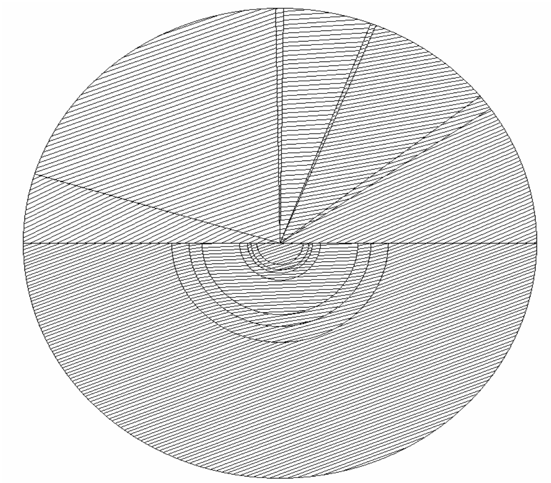

Figure 7. Mask of optimized HRWD

4 Discussion

It is a well known drawback of rough set theory, that it deals with continuous attributes not in a natural way. To overcome this disadvantage, the modified form of indiscernibility relation can be used (3). It introduces the structure into unstructured collection of attributes that the relation depends on. Since the classical relation is a special case of the modified version, therefore this modification can be considered as a step towards generalization. Remarkably, the generalization is equally valid both for classical theory of rough sets, and for the variable precision model, most often used in processing of knowledge obtained from huge data sets. In the case of real-valued attributes, the modified relation allows for performing multidimensional cluster analysis, contrary to multiple one-dimensional analyses required by classical form.

In the experimental study we used a family $\mathbf{C} = \{C_R, C_W\}$ composed of two sets C_R (representing feature space generated by rings) and C_W (representing feature space generated by wedges), each of them containing 8 features (treated as real valued conditional attributes). It allowed us to enhance the recognition abilities by reducing the normalized decision error by 5% compared to system optimized with classical indiscernibility relation. One should notice that this improvement is achieved on already optimized solution, which made any further improvement extremely difficult.

Despite the fact, that proposed here enhanced optimization uses modified version of indiscernibility relation, the classical form is also useful (see for example [5]). Therefore, in order to maximize even more the recognition abilities of the system one can propose a system in which the two types of optimization procedures are used. The optimized HRWD structures (say HRWD_discrete and HRWD_continuous), obtained from these procedures, can be combined in one image recognition system as presented in Fig. 8. Such system (described also in [21]) is in fact composed of two separate feature extractors, however the speed of recognition remains the same, due to parallel processing of optical signals. The spatial light modulators (SLM1 and SLM2, Fig. 8) present the same transparent image of speckle structures, recorded on CCD camera and transmitted from the host computer. The coherent laser beam is divided to two beams, each illuminating one transparency. Due to separate optimization procedures, each optical path produces optimal feature space for given classifier. The final decision is based on two local classifications in any known data fusion algorithms.

The advantage of the system presented in the paper compared to similar systems described in earlier works is two-fold.

First, it uses HRWD optimization methods dedicated for the given type of classifier subsequently used in a system. For rough set based classifier the optimization procedure works in a discrete feature space searching for the optimal structure of HRWD fitted for work with discrete type of classification. Similarly, for artificial neural network based classification working with continuous valued feature vectors, the procedure of HRWD optimization is also able to work in real valued attribute space. This has been made possible by the author's modification of discernibility relation defined in rough set theory.

Secondly, the application of both subsystems in a one two-way solution should result in further improvements in recognition abilities. However further studies are necessary to obtain numerically the improvement. Obvious limitation of presented system is its cost. Even if holographic version of RWD is much cheaper than commercially available RWD, the need of usage of two HRWD in a two-way system, as compared to only one such element in a traditional one-way system, makes presented solution not extremely cost effective, and sometimes the trade-off between cost and efficiency could favour one-way solution. However, when the speed is not crucial, the hybrid system can be simulated by a computer program, and then the expenses do not increase with introduction of the second way of image processing.

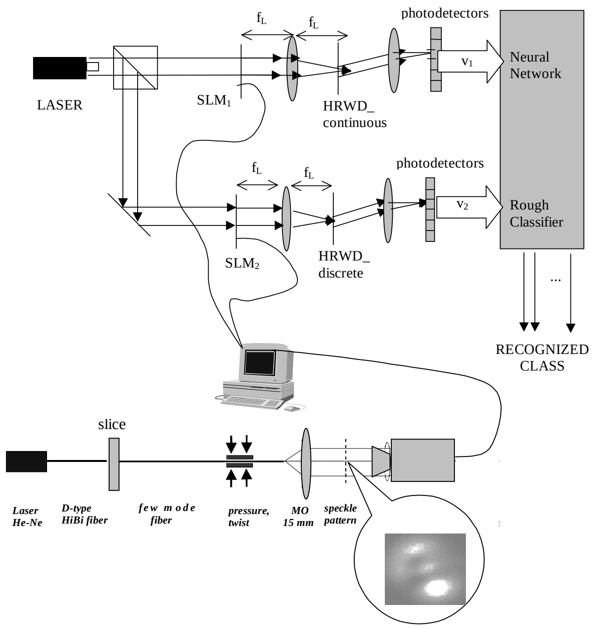

Figure 8. Integration of neural network and rough classifier in combined system composed of two-way optical feature extractor.

Acknowledgements

The author would like to acknowledge the financial support of this work by the grant no 5 T12C 005 25 sponsored by the Polish Ministry of Education and Science (MNiI)

References

1. Kreis T., 1996: *Holographic interferometry – principles and methods*, Akademie Verlag Series in Optical Metrology, 1.
2. Casasent D., Song J., 1985: A computer generated hologram for diffraction-pattern sampling, *Proc. of SPIE*, *523*, 227-236.
3. George N., Wang S., 1994: Neural networks applied to diffraction-pattern sampling, *Applied Opt*ics, *33*, 3127-3134.
4. George N., Wang S., Venable D.L., 1989: Pattern recognition using the ring-wedge detector and neural network software, *Proc. of SPIE*, *1134*, 96-106.
5. Cyran K.A., Mrózek A., 2001: Rough sets in hybrid methods for pattern recognition, *International Journal of Intelligent Systems*, *16*, 149-168.
6. Pawlak Z., 1991: *Rough sets – theoretical aspects of reasoning about data*, Kluwer Academic Publishers.

7. Mrózek A., 1992: Rough sets in computer implementation of rule-based control of industrial processes. In *Intelligent decision support. Handbook of applications and advances of the rough sets*, Słowiński R. (Ed.) Kluwer Academic Publishers, 19-31.
8. Mrózek A., 1992: A new method for discovering rules from examples in expert systems, *Man-Machine Studies, 36*, 127-143.
9. Ziarko W., (1993). Variable Precision Rough Set Model, *Journal of Computer and System Sciences*, 40, 39-59.
10. Skowron A., Grzymała-Busse J.W., 1994: From Rough Set Theory to Evidence Theory. In *Advances in Dempster Shafer Theory of Evidence* Yager R.R., Ferdizzi M., Kacprzyk J. (Eds), John Wiley & Sons.
11. Jaroszewicz L.R., Cyran K.A., Podeszwa T., 2000: Optimized CGH-based pattern recognizer, *Optica Applicata*, *30*, 317-333.
12. Cyran K.A., Jaroszewicz L.R., Niedziela T., 2001: Neural network based automatic diffraction pattern recognition, *Opto-electronic Rev.*, *9* (3), 301-307.
13. Cyran K.A., Stańczyk U., Jaroszewicz L.R., 2002: Subsurface stress monitoring system based on holographic ring-wedge detector and neural network. In *Quality, Reliability and Maintenance*, McNulty G. J. (Ed.), Professional Eng. Publishing, 65-68.
14. Cyran K.A., Niedziela T., Jaroszewicz L.R., Podeszwa T., 2002: Neural classifiers in diffraction image processing, *Proc. of International Conf. On Computer Vision and Graphics*, Zakopane, Poland, 223-228.
15. Podeszwa T., Jaroszewicz L.R., Cyran K.A., 2003: Fiberscope based engine condition monitoring system, *Proc. of SPIE*, *5124*, 2003, pp. 299-303.
16. Jaroszewicz L.R., Merta I., Podeszwa T., Cyran K.A., 2002: Airplane engine condition monitoring system based on artificial neural network. In: *Quality, Reliability and Maintenance*, McNulty G.J. (Ed.), Professional Engineering Publishing, 179-182.
17. Cyran K.A., Jaroszewicz L.R., 2001: Concurrent signal processing in optimized hybrid CGH-ANN system, *Optica Applicata*, *31*, 681-689.
18. Cyran K.A., Jaroszewicz L.R., 2000: Rough set based classifiction of interferometric images. In *Interferometry in speckle light. Theory and applications*, Jacquot P., Fournier J.M. (Eds.) Springer, 413-420.
19. Cyran K.A., 2003: PLD-based rough classifier of Fraunhofer diffraction pattern, *Proc. of International Conf. On Computer, Communication and Control Technologies,* Orlando, FL, 163-168.
20. Cyran K.A., 2005: Combining rule based and connectionist approaches in a diffraction pattern recognition, *Proc. of artificial Intelligence Studies, Special Issue, vol. 2(25)*, 149-157.
21. Cyran K.A., 2005: Integration of classifiers working in discrete and real valued feature space applied in two-way opto-electronic image recognition system, *Proc. of the fifth IASTED International Conference on Visualization, Imaging, and Image Processing (VIIP 2005)*, Benidorm, Spain, 592-597.