Optica Applicata, Vol. XXXI, No. 4, 2001

Concurrent signal processing in optimized hybrid CGH-ANN system

KRZYSZTOF A. CYRAN

Institute of Computer Science, Silesian University of Technology, ul. Akademicka 16, 44–100 Gliwice, Poland.

LESZEK R. JAROSZEWICZ, TADEUSZ NIEDZIELA, IDZI MERTA

Institute of Applied Physics, Military University of Technology, ul. Kaliskiego 2, 00–908 Warszawa 49, Poland.

The paper presents a concurrent hybrid pattern recognition system. The feature extractor of the system is based on optical properties of ring-wedge detector (RWD) or computer generated hologram (CGH) which serves as RWD. The classifier is made as artificial neural network (ANN). Since the feature extraction is an optical and thus fully concurrent process, hence such systems can be designed for real time pattern recognition if only the classifier of characteristic features works fast enough. In presented system the optimized by first author's original method CGH is used instead of widely described standard one. A comparison of recognition results for both types of feature extractors is also provided. Finally, a methodology of obtaining fully concurrent system with optimized CGH and optical ANN playing the role of a classifier is proposed.

1. Introduction

The data processing system presented in Fig. 1, which is a typical example of pattern recognition system, consists of two major parts: a feature extractor and a classifier [1]. Many papers describe hybrid optical-digital system in which the feature extractor is based on very attractive properties of ring-wedge detector placed in the back focal plane of the lens [2]–[4]. These properties result in shift, scale and rotation invariance of feature, and furthermore the whole process of obtaining such a set of features is a purely optic and thus fully concurrent one. Yet, despite of these unquestionably interesting potential possibilities of RWDs, they have drawbacks which eliminate them from many applications. The first disadvantage is their cost, since they are made up of expensive planar photodetectors. The second drawback is connected with the poor flexibility of RWD. Systems composed of such devices are therefore expensive and severe problems arise in practical applications of the laboratory setups.

Both disadvantages are eliminated by using computer generated holograms (CGH) [5] also referred to as diffractive optical elements (DOE) [4]. What is worth stressing

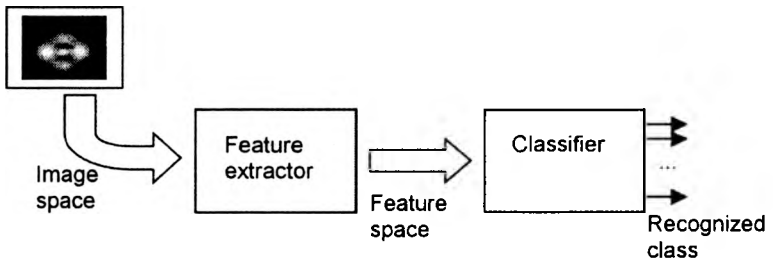

Fig. 1. Schematic diagram of pattern recognition system.

here is that applying CGHs instead of RWD makes it also possibile, to further improve the recognition ability of the system by the cost-effective optimization of the CGH. The stochastic evolutionary method of CGH optimization can be found in [6]. But this method was used only for computer simulations of DOE and therefore the concurrency inherently involved in optical processing has been lost.

This paper summarizes the above mentioned method of CGH optimization, and it also concerns the problem of obtaining real optimum optical element from computer simulation results. The methodology proposed yields better pattern recognition abilities of the system compared to a system with standard CGH and at the same time preserves the massively parallel processing characteristic for optical methods. Furthermore the possible use of optically implemented neural networks serving as the concurrent classifier of features is discussed. The objective here is to obtain pattern recognition system working as a whole concurrently and therefore with possible use for working in real time regime.

2. Concurrency in optical feature extraction

The main purpose of using the RWD as a feature extractor in pattern recognition system is to obtain a set of features that are invariant with respect to typical transformations. It is also important that all features are obtained in parallel, since all operations needed to generate them are purely optical ones.

The properties of RWD are the result of placing it in the back focal plane of the lens. It is well known that the image $F(u, v)$ generated in the back focal plane of the lens L is the Fourier transformed power spectrum of the image $f(x, y)$ passing through it. The above Fourier transformed image $F(u, v)$ has attractive properties such as: symmetric, rotational and scaling of the FT pattern [2], which means that the detector used may contain wedge- and ring-shaped elements. Thus the RWD used to extract characteristic features of the images in spatial frequency domain is a circular element divided into two halves. The first one is composed of half-rings (further on referred to as rings) with the same width i (*i.e.*, the difference between the outer and inner radii of the ring). The second half consists of pie-shaped wedges having the same angles. The standard RWD was patented and produced by ARC Incorporation. It is also commercially available and has got 32 rings and 32 pie shaped wedges [2].

The division of the whole circle into two halves does not reduce the amount of information about light intensity in any half-circle, since the power spectrum is characterized by symmetry. Hence all information about image intensity is separately covered in the area consisting of rings and also in the area consisting of wedges. Each of the rings and wedges is built as a planar silicon photodetector that integrates the intensity of light passing through it. At the output of each photodetector the value of electrical signal corresponds to the value of the feature related with given RWD area (ring or wedge).

Features generated by photodetectors in the shape of rings are invariant to rotation since rings integrate the light intensity throughout the whole possible positions of the input image transformed by rotation. Similarly the features that correspond to the photodetectors in the shape of wedges are invariant with respect to the scaling transformation of the input image. This is because these photodetectors integrate light intensity of all possible sizes of input image. All features are also shift invariant, since light intensity of Fourier transform of the image is always shift invariant.

Despite of the relatively complex process of extraction of characteristic features, as described above, this process is made with full concurrency. The concurrency here should be viewed at a few levels. The first level concerns the concurrency of different operations. More precisely, the process of transforming the image into frequency domain is performed in a full parallelism with the process of light integration and signal conversion in planar photodetectors. To realize how complex operations they are, it is enough to inspect the formulae describing the two-dimensional Fourier transform $F(u, v)$ of the function $f(x, y)$ and integration of light intensity in RWD rings R_i and wedges W_i, where i stands for the number of ring or wedge $1, 2, \ldots, 32$, respectively. In formula below, F_{Ri} denotes the value of feature corresponding to ring R_i and in formula (3), F_{Wi} denotes the value of feature corresponding to wedge W_i

$$F(u, v) = \Im\{f(x, y)\} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x, y)\exp[-2\pi(ux + vy)]\mathrm{d}x\mathrm{d}y, \tag{1}$$

$$F_{R_i} = \iint_{u, v \in R_i} F^2(u, v)\mathrm{d}u\mathrm{d}v, \tag{2}$$

$$F_{W_i} = \iint_{u, v \in W_i} F^2(u, v)\mathrm{d}u\mathrm{d}v. \tag{3}$$

One should also realize that in this signal processing the first specialized processor, namely the Fourier transform processor, is the lens and the second specialized processor (*i.e.*, the integration and conversion processor) is the RWD itself. Both these processors co-operate forming together one concurrent lens-RWD-based feature extractor.

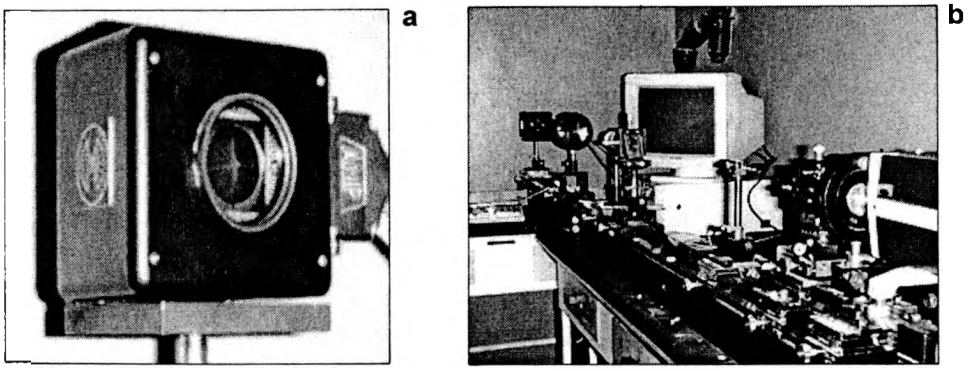

Fig. 2. RWD element (**a**) and lens-RWD-based feature extractor (**b**) [7].

The second level of concurrency in this extractor is clear if we consider that the co-operating system composed of the lens and RWD generates not just one feature but a set of 64 features. All of them are of course generated at the same time. The photo of RWD and of optical RWD-based feature extraction system is presented in Fig. 2.

2.1. Computer simulated processing of images obtained from optic-fiber sensor

Despite of all the afore mentioned advantages of lens-RWD-based feature extractors some papers [5], [7], [8] claim that the RWD can be replaced by CGH. This gives the possibility of building feature extractor that would be cheaper and more suitable for given application compared to universal device such as RWD. The key point here is the fact that rings and wedges of CGH are not made of expensive planar silicon photodetectors, but they are rather the binary diffraction grating [5] described by the equation

$$g(x) = \left[\mathrm{rect}\left(\frac{x}{\Delta x}\right) s\left(\frac{x}{d}\right)\right] \mathrm{rect}\left(\frac{x}{L}\right) \tag{4}$$

where the function rect: $\Re \to \{0, 1\}$ is defined as

$$\mathrm{rect}(x) = \begin{cases} 0 \leftarrow x \notin \langle 0, 1\rangle \\ 1 \leftarrow x \in \langle 0, 1\rangle \end{cases}, \tag{5}$$

and the function s: $\Re \to \Re$ is an ideal sampling function given by

$$s\left(\frac{x}{d}\right) = |d| \sum_{n=0}^{N-1} \delta(x - nd). \tag{6}$$

The function $g(x)$ describes the rectangular wave with unit amplitude, wavelength d and filling $(\Delta x)/d$. This wave is defined for the area of the length L and is composed

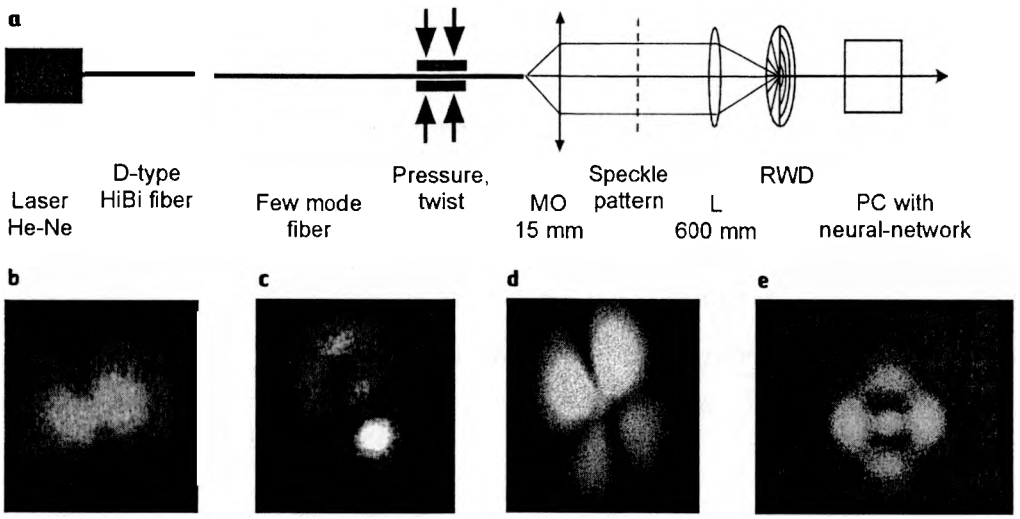

Fig. 3. Diagram of fiber optic sensor (**a**) and speckle patterns from its output for 2 (**b**), 3 (**c**), 4 (**d**), and 5 (**e**) modes propagating in few-mode fiber.

of N fringes. In other words, d is the distance between fringes, Δx is the width of the fringe, whereas L is the length (in the direction perpendicular to fringes) of the area covered with N fringes. Therefore, $u_h = 1/d$ is the space frequency of the diffraction grating defined by the wave $g(x)$.

The binary grating causes that the light passing through a given area is diffracted in a given direction and focused on a much cheaper normal quasi-dot photodetector. In this way, the process of light integration and conversion to electrical signal has been separated in space but not in time. Still these processes are performed in parallel and furthermore there is a possibility of applying high-speed and high-sensitivity photodetectors. Sample images of speckle patterns taken from the output of the optical fiber sensor [8] are shown in Fig. 3 together with the diagram of this sensor.

The feature extraction applied to such speckle patterns was performed in computer simulations of CGH-based system.

Separating the processes of light intensity integration and signal conversion it is also possibile to modify with relative ease the widths of rings and angles of wedges in CGH. Since the sizes of these elements could be easily modified, hence the opportunity to optimize them occurs. The first author's original method of optimizing the sizes of rings and wedges was reported in detail in paper [9]. This method was based on stochastic evolutionary optimization with repair algorithm and the objective function was defined in terms of rough set theory. The more precise rough set aspects of the problem were signalled in [10], [11], here only the objective function for evolutionary optimization will be introduced. For this purpose, the quality of approximation $\gamma_C(D^*)$ of classification family D with respect to the conditional attributes C (values of features corresponding to rings and wedges) was used.

The results obtained from the optimization were very promising, since it became clear that the standard CGH was outperformed by optimized one [12]. Yet this work was done only by computer simulation. The results of feature extraction were better in the optimized model of CGH compared to the standard one, but the processing was no more concurrent. Therefore, the approach was directly applicable only to those domains where real time recognition was not the case. But on the other hand, by even eliminating CGH and performing signal processing in computer software which simulated holographic element, yielded in maximum cost reduction and in many areas it is the cost and not the speed that is the most important criterion.

2.2. Concurrent signal processing in optimized CGH

In the case where the processing speed is not very important, the solution presented above is a good choice. But sometimes, real time operation is needed, and then computer simulated holographic elements are of no interest. Therefore, the next step should be taken to join promising results of computer optimization with concurrent processing of CGH. The task is to produce real CGH with sizes of rings and wedges calculated by computer simulations. Such a solution should give better recognition results compared to systems with standard CGH (or RWD, which, in addition, is a much more expensive device) and at the same time the work is performed with the same degree of concurrency as was discussed for RWD operation.

The process of producing CGH consists of two steps. In the first one, a mask of an optimized holographic element is generated by a computer and printed on a very high quality laser printer (1200 dpi or more). In the second step, the size of the printed mask is reduced by applying some photo-reduction method to obtain the intervals in gratings comparable with light length.

The mask of optimal DOE, generated by first author's program, is presented in Fig. 4. This figure also shows the difference in sizes of rings and wedges in standard RWD element and optimized DOE.

After printing on a high quality laser printer, the CGH mask is reduced by means of photographic method. Then the gratings covering its rings and wedges become the

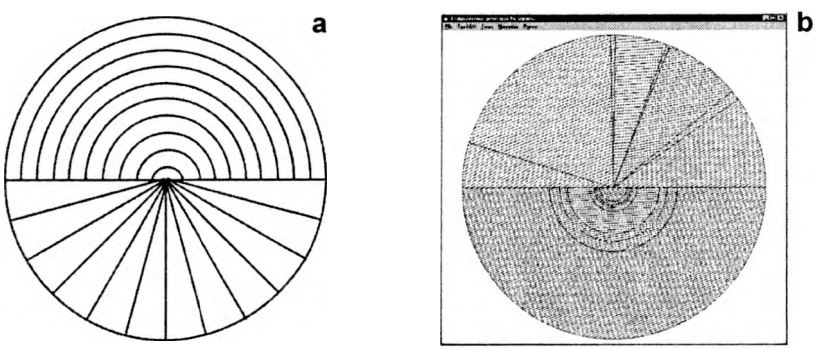

Fig. 4. Comparison of region shapes in standard RWD (**a**) and those produced by computer mask for optimized DOE (**b**).

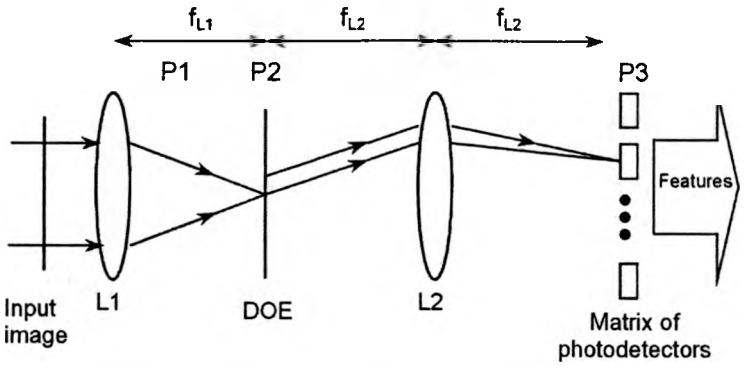

Fig. 5. Feature extraction in optimized CGH based system [10].

diffraction gratings described by Eqs. (4)–(6). The feature extraction system which uses the optimal DOE thus obtained is schematically presented in Fig. 5.

If the grating described by $g(x)$ is placed in the back focal plane of the lens L_2, then the distribution of its Fourier transform amplitude occurring in correlation plane P_3 is given by the equation

$$|G(u)| = L\Delta x d\left[\left|\frac{\sin \pi u \Delta x}{\pi u \Delta x}\right| s(ud)\right] \times \left|\frac{\sin \pi u L}{\pi u L}\right| \tag{7}$$

where $u = x_3/\lambda f_{L_2}$ (if the coherent light of the length λ is used). Therefore the binary grating characterized by space frequency $u_h = 1/d$ in Fourier plane P_2, gives in correlation plane P_3 fringes of n-th order. The strength of these fringes weakens with the growth of n, and the distance x_3 between them is given by

$$x_3 = n\frac{\lambda f_{L_2}}{d}. \tag{8}$$

Since binary grating generates diffracted waves of many orders (7), creating in P_3 fringes with spatial frequency equals f_{L_2}/d (8), hence in CGH design there is a need to satisfy the non-overlapping condition. This condition is fulfilled if diffracted waves of the first order from any CGH region (ring or wedge) do not overlap in P_3 with any other diffracted wave of higher order. If this condition is guaranteed, then instead of sinusoidal fringes the binary fringes can be used. Therefore, the complicated process of hologram design with the use of optical methods can be replaced by the process of binary hologram mask generating with the use of computer software. Such a mask, after being printed with a high quality laser printer is then photoreduced, so that the binary grating becomes diffraction grating. Then the process of light diffraction in CGH regions directs coherent beams passing through these regions to unique locations in the correlation plane P_3, where signal conversion is preformed by photodetector

array. The efficiency of such optical feature extractor is assured by results of optimization of CGH model in computer simulated system.

3. Concurrency in classifier

After obtaining a fully concurrent and optimized for a given recognition feature extractor it is reasonable to assure concurrency in classifier as well. Artificial neural networks (ANNs) can be a good choice here, because they are characterized by relatively low levels of computation complexity, but very high degree of parallelism and interconnectivity [13]. Then the resulting pattern recognition system would be characterized by great speed and maximum overall accuracy.

Since ANNs are well known for their good classification abilities, they are very often used as classifiers of characteristic features obtained from CGH. The following subsection will give a short presentation of ANNs with respect to their potential parallel processing. The last subsection will discuss especially interesting part of neural computing, namely, the optical implementations of neural networks. They are of special interest in systems with CGHs mainly because of the physical nature of features extracted directly by CGH. Since the values of features are in fact encoded into intensities of light in given quasi-dot areas, therefore optical ANNs would eliminate the need of converting light intensities into electronic signals.

3.1. Neural classifiers

Good classification abilities of feed-forward ANN in implementation of any nonlinear mappings are presented in many papers [1], [14]. Such ANNs are especially widely used in pattern recognition problems [15], [16], since their operation is a result of the adaptive training process without a priori knowledge about the rules governing the classification of characteristic features of input images.

The most commonly used feed-forward ANNs consist of many very simple processors called neurons placed in layers, as shown in Fig. 6. Theoretically, there can be an arbitrary number of layers in a neural network, but practically three layers are enough.

The neurons (or the nodes of the network) generate on their outputs y_i a nonlinear sigmoid activation function h of weighted sum of inputs x_j

$$y_i = h(net_i), \; net_i = \sum_j w_{ij} x_j, \tag{9}$$

$$h(x) = \frac{1}{1 + \exp(-x/\theta)}. \tag{10}$$

In the above formulae, w_{ij} are referred to as weights from the output of neuron j in the previous layer to the input of neuron i in layer level and θ is a parameter influencing the slope of the curve $h(x)$.

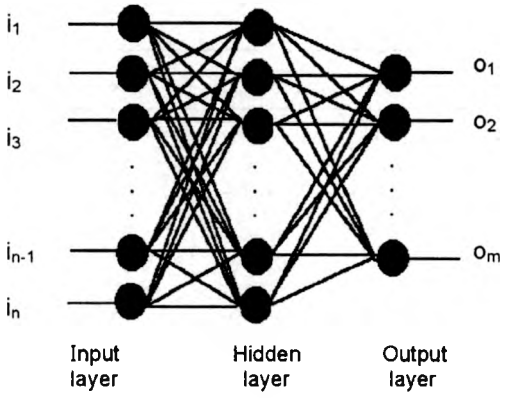

Fig. 6. Three layered feed-forward ANN.

In the experiment with recognition of 5-mode speckle patterns taken from the output of optical fiber sensor (Fig. 3) a set of 128 images was used. These images were taken for different values of external force acting on the fiber. This was the laboratory simulation of stress or strain of the fiber. Since such a stress has influence on images occurring at the output of the fiber, the recognition of the stress class can be done by analysing of the image. There were simulated (by different forces) eight different classes of stress, and for each class 16 sample images were taken (Fig. 7). Then 20% of images were used for testing and 80% for training of the ANN. For measuring the quality of recognition, the normalized decision error, defined as: $E_d = N_b/N_cN_p$, was used. In the above formula, N_b is the number of wrong classified images, N_c is the number of classes and N_p is the number of all images in the set. For images used for training this error was equal to zero, for images from the testing set it was greater than zero, but still small.

The more detailed results of experiment are presented in Tabs. 1 and 2. In these tables the following abbreviations are used:

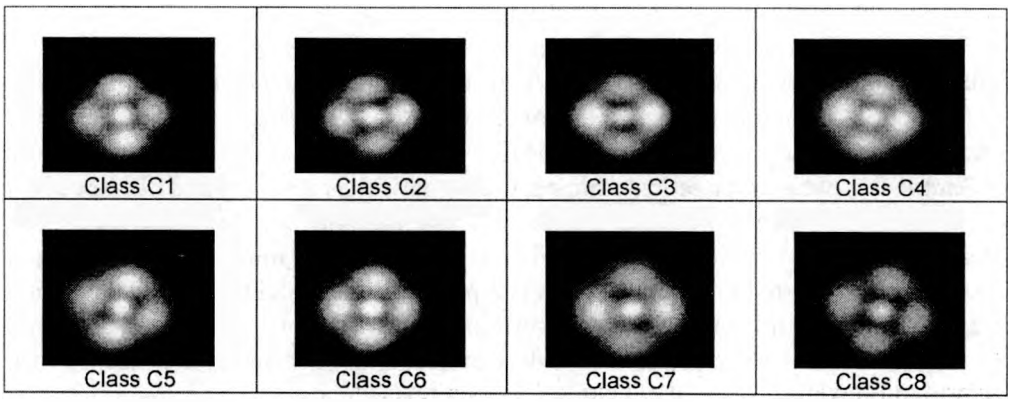

Fig. 7. Examples of images belonging to 8 classes to be recognized.

Table 1. Results of recognition of the stress class for $T = B$.

T_{CGH}	ε_l	ε_t	T_t	H	N_r	E_d	E_{dc}
S	0.1	0.400	*N/N*	13	360	4.8%	4.8%
O	0.1	0.400	*N/N*	13	343	5.8%	3.8%
O	—	0.499	*Y/Y*	7	3826	3.4%	2.8%
O	—	0.499	*Y/Y*	6	9950	3.4%	2.8%

Table 2. Results of recognition of the stress class for $T = E$.

ε_l	ε_t	T_t	H	N_r	E_d	E_{dc}
—	0.499	—	13	130.000	4.8%	4.8%
—	0.499	—	11	100.000	5.8%	4.8%
—	0.499	—	10	100.000	5.3%	3.8%
—	0.499	—	7	70.000	5.3%	7.7%

T – training method (possible values are: B for backpropagation gradient-descent method and E for stochastic evolutionary method),

T_{CGH} – type of the CGH used: S – standard, O – optimized,

ε_l – training (learning) tolerance of ANN used as a classifier,

ε_t – testing tolerance of ANN,

T_t – tolerance tuning/testing while training techniques has been applied in training of the ANN (possible values of this option are: Y or N for both techniques),

H – number of neurons in hidden layer,

N_r – number of runs (epochs) in training,

E_d – normalized decision error for testing set,

E_{dc} – normalized decision error for testing set after using competition mechanism.

In theory of neural networks the operation of all neurons of the same layer is performed in parallel. The problem of time consuming learning of ANN (see the Appendix for back propagation training mathematics) is unimportant for speed of the operation of already trained network. Therefore, the total time of operation depends only on the number of layers (and, of course, on the time τ of processing in one neuron). Since the number of layers is almost always equal to three and the first layer is only a buffering one (*i.e.*, not performing the calculations, see Fig. 4), therefore the time of ANN's response is 2τ and is almost independent of the size of task performed by ANN.

In the last sentence the word "almost" is necessary since the size of the task performed has influence on the number of terms to be added in formula (4) and therefore the time τ is longer for more complicated calculations. Nevertheless, the speed-up of parallel operations of all neurons in a layer is impressive. However, such excellent results of parallel processing in neural networks are possible only if the neural chips are used (in neural accelerators in standard computers or even in neural computers). In the majority of situations this is not the case and then the computer simulation of neural network is performed. Certainly, the operations performed in all simulated neurons are then executed sequentially. Therefore in the concurrent system under consideration, usual computer simulated neural networks should be replaced by neural chips. But this is not the only solution. It is even more suitable to apply here optically implemented neural networks. In this case, as was already mentioned, the conversion from optical representation of characteristic features into electronic signals is not needed anymore. Furthermore, as it becomes clear after presentation of operation in optical ANN, the time of data processing in purely optical ANN is independent of the complexity of the problem to be solved if only the latter fits into three-layered architecture. In the majority of cases it does fit, but if not, the total time is increased to 3τ instead of 2τ, since all mappings used for classification can be done by four-layer architectures [17]. The complexity of the problem influences only the number of neurons in a layer but has no effect on the time of calculations.

3.2. Optical neural networks

In electronic processors the information channels are made of conducting material on a two-dimensional surface. Hence, surface area and power dissipation concerns limit of very high interconnectivity. Optics offers the promising alternative of exploiting the third dimension by allowing free-space (*i.e.*, three-dimensional) interconnections. Non-interference among intersecting optical channels and essentially instantaneous transport over short distances are inherent advantages in choosing optics. Therefore, optical technology in ANNs is very promising in real-time speech and vision processing problems [13]. As was already said, this technology is especially useful for building classifiers of features generated by CGHs, since these features are of optical nature and it is very natural to process them further in optics. The basic process of calculating the weighted sum of the inputs is done in a "Stanford" vector-matrix multiplier shown in Fig. 8.

A linear array of light sources, each encoding the input intensity value, are fanned out vertically by a cylindrical lens. In this way, each input is smeared across a column of a two-dimensional array. By adjusting the transmission of each pixel of the two-dimensional array, often implemented by spatial light modulator (SLM), a unique weighted path or interconnection from each source to detector is defined. Note that these interconnections are defined in a three-dimensional space. Then the second cylindrical lens does fan-in along the horizontal direction giving the total weighted summation at each detector in the array.

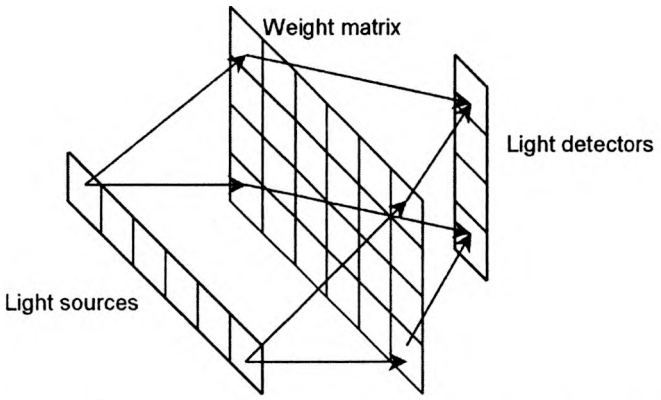

Fig. 8. One layer linear optical ANN based on "Stanford" optical matrix-vector multiplier.

Optical interconnections may be fixed or adaptive. Fixed interconnections are determined in advance by simulation and implemented in some permanent medium such as transparencies or fixed holograms. Adaptive paths demand more complicated hardware and can be implemented by SLMs or adaptive holographic interconnections [13].

The more serious problem in optical implementations of ANNs is how realize the nonlinearity property of the activation function h. The most promising solution is a liquid crystal light valve-based system. In these types of devices the incoming light beam causes an electric field to be generated which controls the light modulating material to allow a nonlinear response approximating a soft threshold representing activation function h.

4. Conclusion

In the paper, a fully concurrent version of the pattern recognition system, used as a computer simulated model for speckle pattern classification, has been proposed. The first part of the system, *i.e.*, feature extractor, is based on CGH optimized by the first author's original method. The second part, the classifier, is a massively parallel neural network, preferably in optical implementation.

Computer simulations of optimized CGH-based extractor produced features that were classified, by neural networks with normalized classification error of 2.8% for the testing set, compared to the error of 4.8% in the case of features obtained for the standard CGH. However, the better classification results were occupied by the non -concurrency in data processing by simulating optimal CGH. This paper presents the next step, which consist in obtaining the optimized, yet optical CGH, joining the optimum feature generation with the concurrency present in standard optical CGH devices.

The last part of the article gave the introduction to possible application of optical neural networks as classifiers of characteristic features generated by optimized CGH

devices. It should, however, be stressed here that optical implementations of ANNs, though being undoubtedly very attractive because of their massively parallel nature, are still at the immature phase of development.

Appendix

This appendix gives the mathematics of the backpropagation algorithm used for ANN learning. The complex, time consuming operations of this algorithm, though very important for the performance of the system, are executed only at the training stage, and therefore have no influence on the overall speed of trained network, and the whole pattern recognizer. However, since training is so important for the behaviour of the ANN (whether simulated in computer, hardware implemented or optical one) this basic training algorithm of feed-forward networks is given below.

Backpropagation method modifies array of weights $\mathbf{W}$ according to the direction of the gradient of error occurring at the outputs of the network [18]. Its aim is to find a network with minimal functional of the root mean square error $e(\mathbf{W})$ for the whole training set consisting of M training facts. Each training fact is a pair $(\mathbf{x}^m, \mathbf{t}^m)$, where $\mathbf{x}^m \in \Re^K$ is an input vector and $\mathbf{t}^m \in \Re^I$ is a vector of expected answers for m-th training fact. If we denote a vector of real answers of the ANN by $\mathbf{y}^m(\mathbf{W}) \in \Re^I$, then for the i-th output neuron and the m-th training fact an error $e_i^m(\mathbf{W})$ is given by

$$e_i^m(\mathbf{W}) = \frac{1}{2}(t_i^m - y_i^m(\mathbf{W}))^2. \tag{A1}$$

The total error e^m for the m-th training fact is

$$e^m(\mathbf{W}) = \sum_{i=1}^{I} e_i^m(\mathbf{W}) = \frac{1}{2}\sum_{i=1}^{I}(t_i^m - y_i^m(\mathbf{W}))^2. \tag{A2}$$

The total error e for all training facts is given by

$$e(\mathbf{W}) = \sum_{m=1}^{M} e^m(\mathbf{W}) = \frac{1}{2}\sum_{m=1}^{M}\sum_{i=1}^{I}(t_i^m - y_i^m(\mathbf{W}))^2. \tag{A3}$$

The training of the network consists in minimization of this error. It is the search for the minimum of the scalar field over the vector space. Such search can be done by a gradient descent method

$$\Delta\mathbf{W} = -\eta\nabla e(\mathbf{W}). \tag{A4}$$

For elements w_{ij} of array $\mathbf{W}$ we have

$$\Delta w_{ij} = -\eta\frac{\partial e(\mathbf{W})}{\partial w_{ij}}. \tag{A5}$$

To fulfil the strict requirements of gradient descent method the weight array modification should be performed once after the whole training set. In practice, however, more often weights are modified after each training fact. For this case, the changes of weights are done according to the following formulae:

$$\Delta w_{ij} = -\eta \frac{\partial e^m(\mathbf{W})}{\partial w_{ij}} = -\eta \frac{\partial e^m(\mathbf{W})}{\partial net_i^m} \frac{\partial net_i^m}{\partial w_{ij}^m}. \tag{A6}$$

If we denote by [19]

$$\delta_i^m = -\frac{\partial e^m(\mathbf{W})}{\partial net_i^m} \tag{A7}$$

and taking into consideration (4) we finally have

$$\Delta w_{ij}^m = \eta \delta_i^m O_j^m. \tag{A8}$$

For output neurons

$$\delta_i^m = -\frac{\partial e^m(\mathbf{W})}{\partial O_i^m} \frac{\partial O_i^m}{\partial net_i^m} = h'(net_i^m)(t_i^m - y_i^m). \tag{A9}$$

However, for hidden neurons the value of t_i^m is directly unknown, therefore there is a need to backpropagate errors from the output layer according to the chain rule

$$\frac{\partial e^m(\mathbf{W})}{\partial O_i^m} = \sum_{n=1}^{I} \frac{\partial e_n^m(\mathbf{W})}{\partial net_n^m} \frac{\partial net_n^m}{\partial O_i^m} = \sum_{n=1}^{I} \frac{\partial e^m(\mathbf{W})}{\partial net_n^m} \frac{\partial net_n^m}{\partial O_i^m} = -\sum_{n=1}^{I} \delta_n^m w_{ni}^m. \tag{A10}$$

And finally, for hidden layer we have

$$\delta_i^m = h'(net_i^m) \sum_{n=1}^{I} \delta_n^m w_{ni}^m. \tag{A11}$$

Acknowledgments – The authors would like to acknowledge financial support of the Polish State Committee for Scientific Research (KBN) under the grant No. 0-T00A-021-17 in the year 2000.

References

[1] ŻURADA J., BARSKI M., JĘDRUCH W., *Artificial Neural Networks* (in Polish), Wydawnictwo Naukowe PWN, Warszawa 1996.

[2] GEORGE N., WANG S., VENABLE D. L., Proc. SPIE **1134** (1989), 96.

[3] GEORGE N., WANG S., Appl. Opt. **33** (1994), 3127.
[4] MARSHALL M., BENNER R., Opt. Engin. **31** (1992), 947.
[5] CASASENT D., SONG J., Proc. SPIE **523** (1985), 227.
[6] CYRAN K.A, JAROSZEWICZ L.R., MRÓZEK A., Proc. SPIE **4238** (1999), 234.
[7] JAROSZEWICZ L.R., MERTA I., KIEŻUN A., Proc. SPIE **3555** (1998), 337.
[8] JAROSZEWICZ L.R., MERTA I., Proc. SPIE **3730** (1998), 13.
[9] JAROSZEWICZ L.R., CYRAN K.A., PODESZWA T., Opt. Appl. **30** (2000), 317.
[10] CYRAN K.A., Proc. SPIE **3744** (1999), 241.
[11] CYRAN K.A., *Computer Generated Holograms and Rough Sets in Pattern Recognition*, Proc. Internat. Workshop: *Control and Information Technology*, Ostrava, Czech Rep., 1999, 213–219.
[12] JAROSZEWICZ L.R., CYRAN K.A., KŁOSOWICZ S.J., MRÓZEK A., Proc. SPIE **3744** (1999), 386.
[13] SAXENA I., HORAN P.G., *Optical implementations*, [In] *Handbook of Neural Computation*, [Eds] E. Fiesler, R. Baele, E1.5:1-E1.5:20, Institute of Physics Publishing and Oxford University Press, New York 1997.
[14] DENOEUX T., *Pattern classification*, [In] *Handbook of Neural Computation*, [Eds] E. Fiesler, R. Baele, F1.2:1-F1.2:8, Institute of Physics Publishing and Oxford University Press, New York 1997.
[15] LISBOA P.J.G., *Image classification using Gabor representations with a neural net*, [In] *Neural Networks for Vision, Speech and Natural Language*, 112–127, [Eds] R. Linggard, D.J. Myers, C. Nightingale, Chapman & Hall, London 1992.
[16] SCHALKOFF R., *Pattern Recognition – Statistical, Structural and Neural Approaches*, Chapt. 10–12, Wiley, Inc., Singapore, 1992.
[17] TADEUSIEWICZ R., *Neural Networks* (in Polish), Akademicka Oficyna Wydawnicza RM, Warszawa, 1993.
[18] DE LEONE R., CAPPARUCCIA R., MERELLI E., IEEE Trans. Neural Networks **9** (1998), 381.
[19] LAWRENCE J., *Introduction to Neural Networks – Design, Theory, and Applications*, California Scientific Press, 1994.

Received September 7, 2000
in revised form February 19, 2001