# Parallel distributed processing model with local space-invariant interconnections and its optical architecture

Wei Zhang, Kazuyoshi Itoh, Jun Tanida, and Yoshiki Ichioka

This paper proposes a parallel distributed processing model with local space-invariant interconnections, which is more readily implemented by optics and is able to classify patterns correctly, even if they have been shifted or distorted. Error backpropagation is used as a training algorithm. Computer simulation results presented indicate that the processing is effective and the network can deal with the shifted or distorted patterns. Moreover, the optical implementation architecture using matched filters for the model is discussed. *Key words:* Parallel distributed processing, space-invariant interconnection, matched filter, error backpropagation.

## I. Introduction

Pattern classification algorithms based on linear mappings,[1] such as the synthetic discriminant function (SDF) and least-squares linear mapping technique (LSLMT), have been widely investigated because they fit well the optical implementations. From the point of view of an artificial neural network or parallel distributed processing (PDP), such approaches are equivalent to the simple perceptron model,[2] a two-layer feed forward network with adaptive connection weights. It is well known that such a simple network cannot deal with nonlinear mapping tasks, such as the XOR problem.[3] Recent research on learning algorithms of the PDP models[4–6] allows us to build arbitrary networks to perform tasks that cannot be accomplished by networks based on linear mapping. The massive interconnectivity, parallelism, and analog nature of the optical architectures are good matches to the requirements of PDP models and have caused many workers to consider optical implementation for the PDP models.[7–10] In other words, a more sophisticated optical information processing system can be constructed utilizing the learning algorithms of the neural network.

In the pattern classification using PDP models, one problem is to deal with the position shift and pattern deformation. Because of the overlearning problem, test patterns which are shifted in position or deformed in shape fail to be recognized. One way to overcome this problem is to train the network with all the shifted and deformed patterns: this training obviously takes an extremely long time. An alternative approach is to design the networks to show position deformation-invariant responses. One of the choices in such a design is to constrain the PDP model to satisfy the following criteria:

(1) In the middle stages of the model, the units should show the same response pattern but shift in position when the stimulus pattern in the lower layer is shifted.

(2) A local deformation in the structure of the stimulus pattern should only cause a local deformation of the response pattern in the middle stages.

(3) For the decision units or grandmother cells in the last stage, precise positions, shapes, or orientations of stimulus patterns are not important; only densities matter.

According to the criteria above, we propose a multilayered feed-forward PDP model, in which the interconnections are designed to be local space invariant and the receptive field of each neuron in the lower layer is constrained. In the last layer, the decision unit only responds to the spatial average of the stimulus pattern. Details are in Sec. II.

The general structure of our model is similar to that of the Neocognitron model,[11] proposed by Fukushima based on early physiological studies of visual systems.[12,13] However, Neocognitron has a complicated microstructure and very complicated cells. As a result it does not fit the optical implementations, at least

The authors are with Osaka University, Department of Applied Physics, Yamadaoka 2-1, Suita, Osaka 565, Japan.

Received 13 April 1989.
0003-6935/90/324790-08$02.00/0.

© 1990 Optical Society of America.

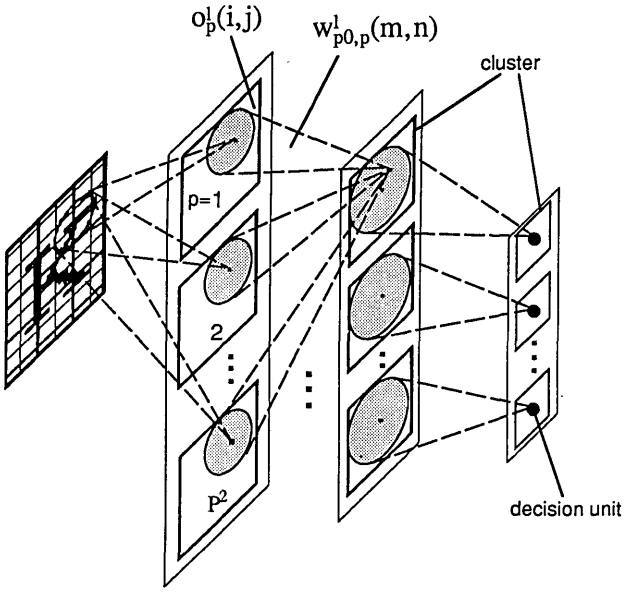



Fig. 1. PDP model with local space-invariant interconnection and constrained receptive fields.

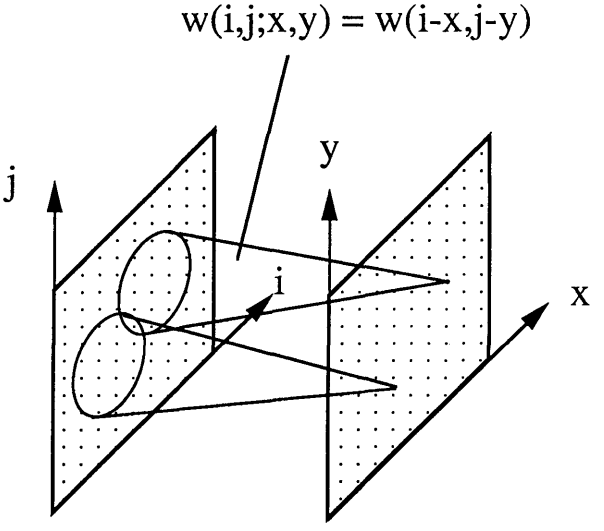



Fig. 2. Space-invariant interconnection.

in current component technologies. Our research emphasizes the optical implementation of the PDP models rather than the biological plausibility. Consequently, instead of the master/slave training algorithm, error backpropagation (EBP) is used to train the system. The reasons for using EBP are that, first, in our case we want to train the network to be able to classify patterns even when they rotate in a certain range, so that teacher signals are necessary. Second, in the EBP training procedure there is no need to design the internal representations of hidden units, which would be difficult in large scale networks. Last, in EBP training, lateral interconnection is not necessary, so that the structure and units of the network are simpler and may be easier to be implemented optically.

Computer simulation results presented in Sec. III debate whether the network can deal effectively with the shift and distortion of stimulated patterns. After training by the rotated patterns at some discrete angles, the network also shows rotation-invariant responses.

As described above, the pattern classification procedure of this model can be implemented by optics because of the character of the local space-invariant interconnection. An optical implementation architecture involving conventional optical matched filters is also discussed in Sec. IV.

## II. PDP Model with Local Space-Invariant Interconnections and its Learning Algorithm

### A. Model

The proposed model is a layered feed-forward network which consists of neuronlike units with semilinear activation functions, such as the sigmoid function. All the units have modifiable input synapses which are reinforced during the learning phase. As described previously, to meet the EBP algorithm mechanism and the optical implementation considerations, no lateral interconnection is used. The structure of this network is illustrated in Fig. 1. Units in any single layer are divided into clusters. Except for the last output layer, each unit in the above layers is connected with some of the units in every cluster of the layer below. In order that the learning that has occurred be independent of where on the input layer the input pattern appeared, all the units in the same cluster are constrained to learn exactly the same pattern of connectivity, while only the centers of the receptive fields are shifted in parallel depending on the position of the units. Figure 2 illustrates the space-invariant interconnection between two clusters. If we let $w(i,j;x,y)$ denote the connection weight, it can be formulated as follows except for the edge unit:

$$w(i,j;x,y) = w(i-x;j-y),$$

where $(i,j)$ and $(x,y)$ denote the unit coordinates in the lower layer and upper layer, respectively. Units in the same layer but in a different cluster have the same size receptive field but a different pattern of connectivity. The higher the layer is, the larger the receptive field becomes of each unit in that layer. The number of units in each cluster is so determined as to decrease in accordance with the increase in the size of the receptive fields. In the output layer, the receptive field of each unit becomes large enough to cover the whole input field, and each cluster is determined to have only one decision unit (grandmother cell). Each decision unit connects with only one exclusive cluster in the penultimate layer by an alike connection pattern.

Under the constraints of space-invariant interconnection, the pattern of connectivity between two clusters functions as a filter. In the first stage, the input image is correlated with a bank of filters followed by the sigmoidlike clippings to give the outputs of the

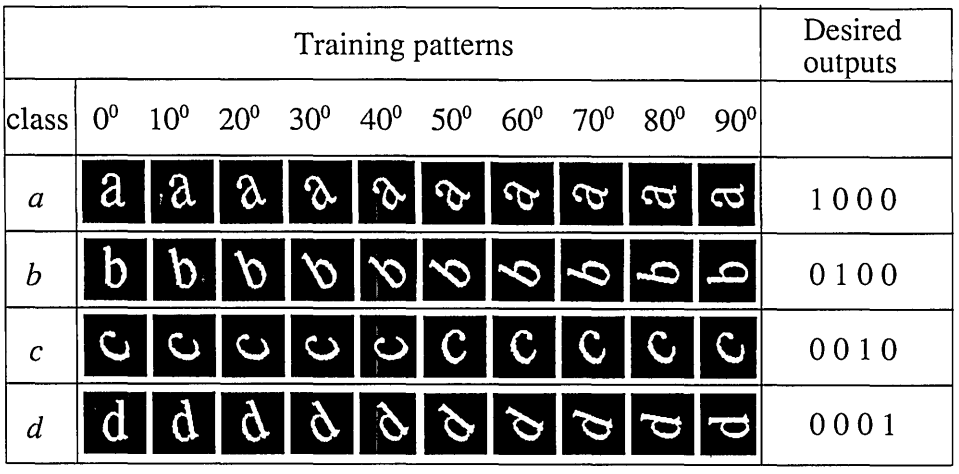



Fig. 3. Training set.

clusters in the second layer. In the middle stage, units of a cluster in the lower layer contribute their stimulations by correlating with its corresponding bank of filters. Units in the higher layer sum all the stimulations contributed from all clusters in the lower layer and give their outputs after the sigmoidlike clippings. In the last stage, outputs of a cluster in the lower layer are relayed to one exclusive decision unit through an alike filter of the same size as the cluster. It means that the decision unit distinguishes two stimulus patterns only by the different spatial averages regardless of their precise positions, shapes, or orientations.

### B. Learning Algorithm

Let $\text{net}_p^l(i,j)$ denote the net input of the unit in the $p$th cluster of the $l$th layer, $w_{p0,p}^l(m,n)$ is the connection pattern between the units in the $p$th cluster of the $(l+1)$th layer and the units in the $p0$th cluster of the $l$th layer, $o_p^l(i,j)$ is the output of the unit in the $p$th cluster of the $l$th layer, and $b_p^l$ is the bias of the unit in the $p$th cluster of the $l$th layer. The feed-forward propagations of the input signal can be formulated as

$$\text{net}_p^{l+1}(i,j) = \sum_{p0}\sum_{m}\sum_{n} w_{p0,p}^l(m,n) \times o_{p0}^l(i+m+s^l, j+n+s^l) + b_p^{l+1} \quad (1)$$

$$s^l = \tfrac{1}{2}(I^l - I^{l+1} - M^l),$$

$$o_p^{l+1}(i,j) = f[\text{net}_p^{l+1}(i,j)], \quad (2)$$

but for the last layer

$$w_{p0,p}^{L-1}(m,n) = \begin{cases} w^{L-1} & \text{if } p0 = p, \\ 0 & \text{else}, \end{cases} \quad (3)$$

where $l = (1,2,\ldots,L)$ layer number,
$L$ = total number of layers,
$p0,p = (1,2,\ldots,p^l)$ cluster number,
$p^l$ = total number of clusters in the $l$th layer,
$i,j = (1,2,\ldots,I^l)$ coordinate of units in a cluster,
$I^l$ = total number of units in a cluster of the $I$th layer,
$m,n = (1,2,\ldots,M^l)$ coordinate of the weight,

**Table I. Structure of the Network Used in Our Simulation**

| Layer (l) | : | 1 | 2 | 3 | 4 |
|---|---|---|---|---|---|
| Size (IxIxP) | : | 31x31 | 23x23x4 | 11x11x4 | 1x1x4 |
| Fan-in (MxMxP) | : | | 7x7x1 | 11x11x4 | 11x11x1 |

I: the number of units,
P: the number of clusters,
M: the size of receptive field.

$M^l$ = size of the receptive field of units in the $(l+1)$th layer, and
$f(x)$ = sigmoid function.

Let the change of weight be $\Delta w_{p0,p}^l(m,n)$ and the bias changes be $\Delta b_p^l$. The training algorithm in this model can be modified to

$$\Delta w_{p0,p}^l(m,n) = \eta \sum_i \sum_j \delta_p^{l+1}(i,j) \times o_{p0}^l(i+m+s^l, j+n+s^l), \quad (4)$$

$$\Delta b_p^l = \eta \sum_i \sum_j \delta_p^l(i,j), \quad (5)$$

$$\delta_{p0}^l(i,j) = f'[\text{net}_{p0}^l(i,j)] \sum_p \sum_m \sum_n \times [\delta_p^{l+1}(i-m-s^l, j-n-s^l) w_{p0,p}^l(m,n)]; \quad (6)$$

for the last layer,

$$\delta_p^L = f'(\text{net}_p^L)(t_p - o_p^L), \quad (7)$$

where $f'(x)$ denotes the derivative of $f(x)$ and $\eta$ denotes the learning rate.

## III. Computer Simulation

A number of experiments are carried out using the model described previously for multiclass pattern rec-

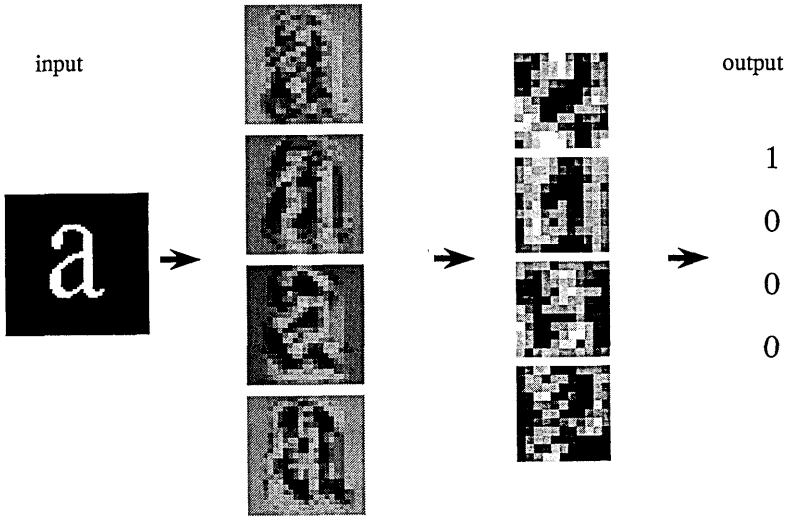



Fig. 4. Responses of each layer to the input pattern **a**.

ognition. The central issues are to determine whether the system can be trained under the constraint of the local space-invariant interconnection and is able to deal with shift or distortion of input patterns. The structure of the system used is illustrated in Table I.

As shown in Fig. 3, four letters of the Roman alphabet are the objects. The bright pixels have a level of 1, and the dark pixels have a level of 0. To train the system to be invariant over rotation ranges up to 90°, images of each letter separated by rotations of 10° are used as training patterns. The desired outputs are designed so that for the input of a stimulus pattern, regardless of the rotation, only one of the exclusive output units indicates positive, i.e., 1, while the other units indicate a negative output, i.e., 0.

In the experiments, all the weights start with random values between −0.3 and 0.3. The values of 0.1

Fan-out field

$$\sum_{m,n} w^{l}_{p0,p}(m,n) \cdot w^{l}_{p0,p'}(m,n), \quad p' \le p$$

l=1, p0=1

| p' \ p | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| 1 | 1.00 | 0.02 | -0.06 | -0.13 |
| 2 | | 1.00 | 0.04 | 0.10 |
| 3 | | | 1.00 | 0.10 |
| 4 | | | | 1.00 |

l=2, p0=1

| p' \ p | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| 1 | 1.00 | -0.10 | 0.07 | -0.06 |
| 2 | | 1.00 | -0.11 | 0.16 |
| 3 | | | 1.00 | -0.03 |
| 4 | | | | 1.00 |

l=2, p0=2

| p' \ p | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| 1 | 1.00 | -0.09 | 0.02 | -0.00 |
| 2 | | 1.00 | -0.09 | -0.01 |
| 3 | | | 1.00 | -0.02 |
| 4 | | | | 1.00 |

l=2, p0=3

| p' \ p | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| 1 | 1.00 | 0.06 | -0.14 | -0.06 |
| 2 | | 1.00 | -0.08 | -0.04 |
| 3 | | | 1.00 | -0.01 |
| 4 | | | | 1.00 |

l=2, p0=4

| p' \ p | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| 1 | 1.00 | -0.14 | 0.08 | -0.04 |
| 2 | | 1.00 | -0.11 | 0.07 |
| 3 | | | 1.00 | -0.06 |
| 4 | | | | 1.00 |

(a)

Fan-in field

$$\sum_{m,n} w^{l-1}_{p0,p}(m,n) \cdot w^{l-1}_{p0',p}(m,n), \quad p0' \le p0$$

l=3, p=1

| p0' \ p0 | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| 1 | 1.00 | -0.09 | -0.09 | -0.02 |
| 2 | | 1.00 | -0.01 | 0.12 |
| 3 | | | 1.00 | -0.06 |
| 4 | | | | 1.00 |

l=3, p=2

| p0' \ p0 | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| 1 | 1.00 | -0.05 | -0.12 | 0.07 |
| 2 | | 1.00 | -0.17 | 0.04 |
| 3 | | | 1.00 | 0.01 |
| 4 | | | | 1.00 |

l=3, p=3

| p0' \ p0 | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| 1 | 1.00 | -0.03 | -0.04 | 0.00 |
| 2 | | 1.00 | 0.21 | -0.07 |
| 3 | | | 1.00 | -0.05 |
| 4 | | | | 1.00 |

l=3, p=4

| p0' \ p0 | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| 1 | 1.00 | 0.02 | 0.14 | 0.04 |
| 2 | | 1.00 | 0.05 | 0.09 |
| 3 | | | 1.00 | 0.01 |
| 4 | | | | 1.00 |

(b)

Fig. 5. Inner products between each pair of the connection patterns in (a) the fan-out field and (b) the fan-in field.

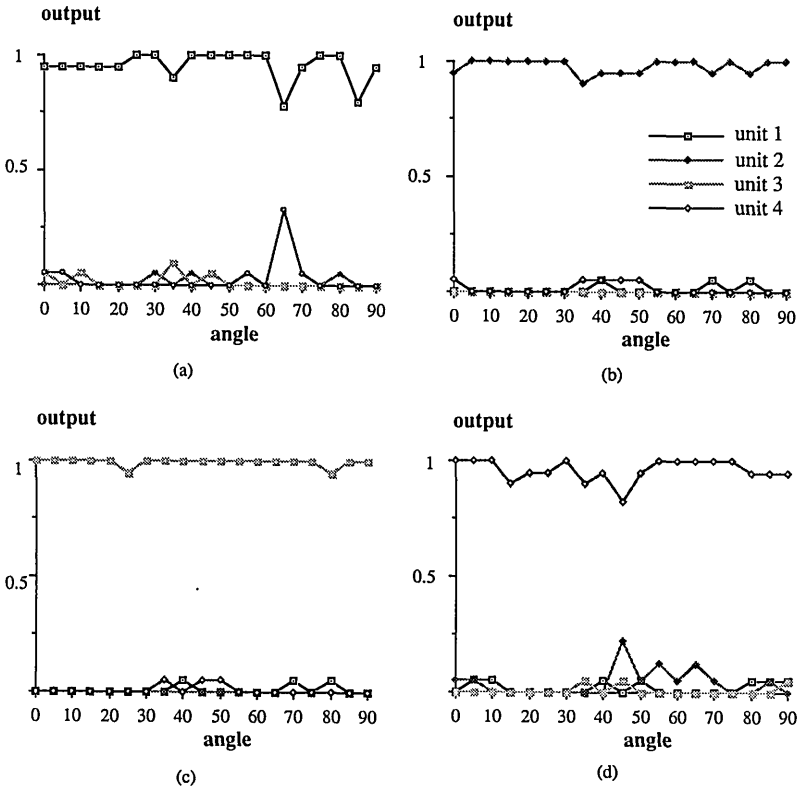



Fig. 6. Responses of the system measured at every 5°. Unit $i$ denotes each decision unit in the last layer: (a) to the class of **a**; (b) to the class of **b**; (c) to the class of **c**; (d) to the class of **d**.

and 0.9 are typically used as the convergence target of 0 and 1, respectively. The convergence state is reached after 127 presentations of each pattern with $\eta = 0.05$. To increase the learning rate without oscillation, Eq. (4) is modified to include a momentum term.[6] In our simulation, the constant factor of the momentum term is 0.9.

The internal representations of hidden units are investigated after training. It should be mentioned that in the case of the EBP model, because no competition exists and the sigmoid function is used as the unit activity function, for a particular input stimulus pattern each unit gives an analog output of response. As a result, the minute investigations of hidden layers become relatively difficult. Figure 4 shows an example of the activities of the units when image **a** at 0° angle is input. All the connection patterns are normalized, and the inner products between each are calculated to show their relationships. The results are illustrated in Fig. 5. It can be seen that each cluster has almost orthogonal fan-in and fan-out connection patterns. It means that each fan-in or fan-out connection pattern maps different feature of the stimulus input. Finally, the system is tested with various input images. Figure 6 shows the responses of the system to the images spanning their entire design range at 5° internals including views of the images between those used in the training set. As can be seen in Fig. 6 the system is rotation invariant in the 0–90° range. Figure 7 shows some examples of the test images that the system correctly recognizes. Because of the effect of edges,

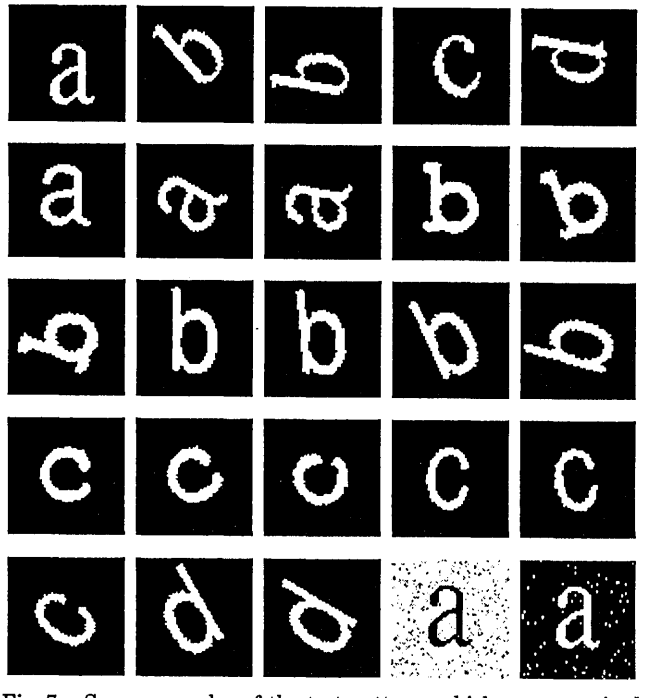


Fig. 7. Some examples of the test patterns which are recognized correctly. The last two patterns are deformed by the impulse noises whose levels are bright and Gaussian variable, respectively. To make it clear, the penultimate one is displayed reversely.

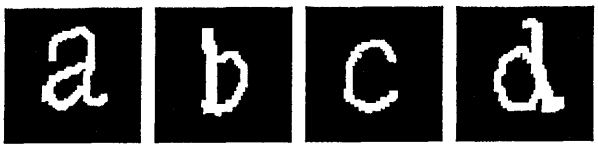



Fig. 8. Some examples of the test patterns which are not recognized correctly by our model.

the system is not completely shift-invariant. In our simulation, the shift-invariant ranges are 4 pixels in the horizontal direction and 8 pixels in the vertical direction for trained images. The letters deformed in shape, shown in Fig. 7, are fonts different from the trained letters. We tested them at every 5° in the range of 0–90°, and more than 75% of them were correctly recognized. The effects of noise are also investigated by convolving bright (with level 1) and Gaussian (the level is a Gaussian variable with a standard deviation of $\sigma$) impulse noise to the trained patterns. In our simulations, the system can recognize the patterns correctly in the case of bright noise when the occurring probability is <0.02 and in the case of Gaussian noise with $\sigma = 0.2$ when the occurring probability is <0.25. On the other hand, Fig. 8 shows some examples of the images which cannot be correctly recognized at any rotation angle.

## IV. Optical Implementation Architecture

Because of its local space-invariant structure, the model described previously can be implemented by using the conventional optical matched filters. Figure 9 shows the architecture for the use of a four-class pattern classification. The system has four unit layers; except for the input layer, the units in each layer are divided into four clusters corresponding to the different areas of the spatial light modulators (SLMs). In the last layer each cluster includes only one unit for the decision output. The space-invariant interconnections between each pair of clusters are implemented by the hologram arrays which are synthesized in the fashion of the VanderLugt matched filter.[14] The sigmoidlike clipping of the unit activity function is performed by the SLMs. The bias light of each SLM permits the interferometric detection.

Let us consider the signal path from the input layer to the second layer. As described in Sec. II, the net input distribution of a cluster in the second layer is the cross correlation of the stimulus pattern in the input layer and its corresponding connection pattern. We define a filter function $w^1(m,n)$ as

$$w^1(m,n) = w^1_{1,1}(m - m0,n + n0) + w^1_{1,2}(m + m0,n + n0) + w^1_{1,3}(m - m0,n - n0) + w^1_{1,4}(m + m0,n - n0), \quad (8)$$

where $w^1_{1,i}(m,n)$ denotes the connection pattern between the input layer and the second layer as defined in Eq. (1). As illustrated in Fig. 10, if $m0,n0 > 0.5$ ($I^1 + M^1$), the net inputs of the units in the second layer can be obtained by correlating the stimulus pattern with the filter function $w^1(m,n)$ without crosstalk. In Fig. 9, $L1$ and $L2$ are Fourier transform lenses, hologram $H$ restores filter function $w1$ in the fashion of a matched filter which can be synthesized by either optics or computer. Here $\lambda$ denotes the wavelength of the incident light and $f$ the focal length of the Fourier transform lenses. The intensity distribution stored in the hologram $H$ can be written as

$$I(x_2,y_2) = \left| W^1\left(\frac{x_2}{\lambda f}, \frac{y_2}{\lambda f}\right) + \exp[-i2\pi(\alpha x_2 + \beta y_2)] \right|^2, \quad (9)$$

$$\alpha = \frac{\cos\theta_x}{\lambda}, \quad \beta = \frac{\cos\theta_y}{\lambda},$$

where the first term is the Fourier transform of $w^1(m,n)$, and the second term is a tilted reference plane

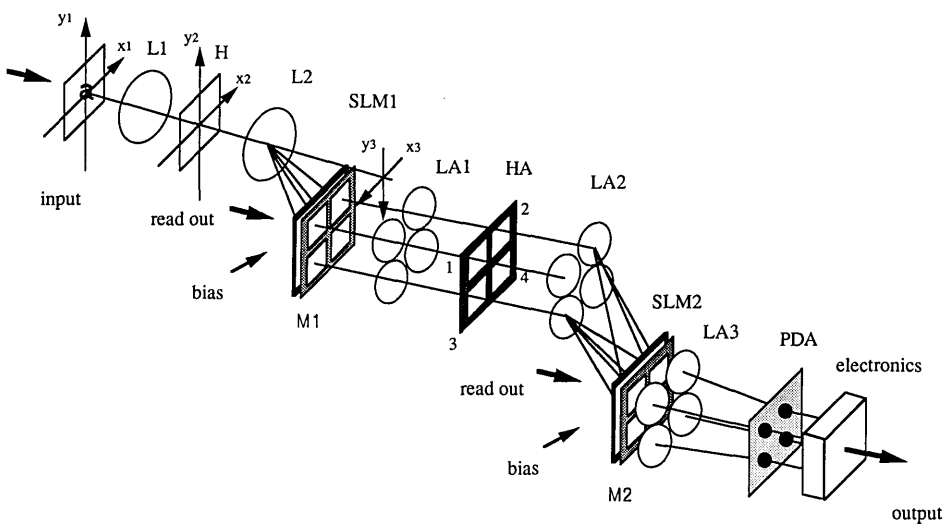
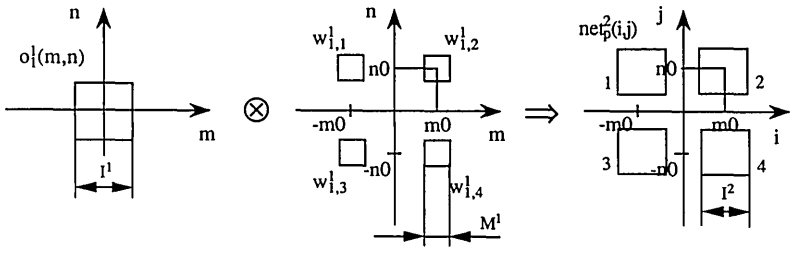



Fig. 9. Optical architecture: $L$, Fourier transform lenses; $H$, hologram; $SLM$, spatial light modulators; $LA$, Fourier transform lens arrays; $HA$, hologram array; $M$, masks; PDA, photodiode array.




input pattern filter function $w^1(m,n)$ net input of the second layer

Fig. 10 Getting the net input of the second layer by taking the cross correlation between the input and connection patterns without crosstalk. The symbol ⊗ denotes operation of the cross correlation.

wave with unity amplitude. Variables $\theta_x,\theta_y$ denote the angle of the normal of the reference plane wave with respect to the $x_2,y_2$-axis, respectively. As described in Ref. 14, to obtain cross correlation without crosstalk with useless terms, the reference wave should be introduced at a sufficient angle. In our case the angles are set so that

$$\alpha > \frac{1}{\lambda f}\left[\frac{3(I^1+2M^1)}{2}+I^1\right], \quad \beta = 0. \tag{10}$$

As a result, when a stimulus pattern appears in the input layer, the net input to the second layer is obtained around the coordinate $(x_3,y_3) = (\alpha\lambda f,0)$. The $SLM1$ performs the sigmoidlike clipping of the incident light including the net input and bias, then relays the readout light onto the next stage through mask $M1$, which constrains the size of the cluster.

In the second stage, the light paths are divided into four independent channels by lens arrays $LA1$ and $LA2$ and hologram array $HA$. Each channel is constructed in the same way as the first stage. Subholograms in hologram array $HA$ store the fan-out connection pattern of its corresponding cluster in the second layer. As a result, the stimulus pattern from a cluster of the second layer will be mapped to the input of $SLM2$ through its own channel. To get the net input contributed from all the clusters in the second layer in the same position on $SLM3$, the angles of the reference plane wave of each subhologram $(\alpha_i,\beta_i)$ should be modified individually as

$$\begin{aligned}
(\alpha_1,\beta_1) &= \left(\alpha+\frac{m0}{\lambda f},\beta-\frac{n0}{\lambda f}\right),\\
(\alpha_2,\beta_2) &= \left(\alpha-\frac{m0}{\lambda f},\beta-\frac{n0}{\lambda f}\right),\\
(\alpha_3,\beta_3) &= \left(\alpha+\frac{m0}{\lambda f},\beta+\frac{n0}{\lambda f}\right),\\
(\alpha_4,\beta_4) &= \left(\alpha-\frac{m0}{\lambda f},\beta+\frac{n0}{\lambda f}\right),\\
\alpha &> \frac{1}{\lambda f}\left[\frac{3(I^2+2M^2)}{2}+I^2\right], \quad \beta = 0,
\end{aligned} \tag{11}$$

where subscript $i$ denotes the channel number as shown in Fig. 9. Here $I^2, M^2$ are the sizes of the cluster in the second layer and its fan-out receptive field, respectively. (Superscripts 2 denote the layer number, not signs of second power.) After the sigmoidlike clipping of the sum of the net input and bias light by the $SLM2$, the stimulus patterns of the third layer are relayed to the last stage by the readout light of the $SLM2$. In the last stage, each lenslet in lens array $LA3$ gathers the light from its own cluster in the third layer into a photodetector. Electronic circuits perform the sigmoidlike clipping and create the final decision output.

## V. Conclusion

A multilayered feed-forward PDP model with local space-invariant interconnections has been investigated. This model is able to classify patterns correctly, even if they have shifted, distorted, or rotated, in a certain range.

Under the constraint of the local space-invariant interconnection, the EBP training algorithm can be used to design an optical multiple matched filter system for pattern classification use. Using this algorithm, a more sophisticated system can be designed by just representing the input and desired output patterns without knowledge of the internal representations within the system. We anticipate that this layered system will be more powerful than a conventional optical filtering system based on linear mapping.

## References


1. Q. Tian, Y. Fainman, Z. H. Gu, and S. H. Lee "Comparison of Statistical Pattern-Recognition Algorithms for Hybrid Processing. I. Linear-Mapping Algorithms," J. Opt. Soc. Am. A **5**, 1655–1669 (1988).
2. M. Minsky and S. Papert, *Perceptrons* (MIT Press, Cambridge, MA, 1969).
3. B. Widrow and R. Winter, "Neural Nets for Adaptive Filtering and Adaptive Pattern Recognition," Computer **21**, 25–39 (1988).
4. T. Kohonen, "The "Neural" Phonetic Typewriter," Computer **21**, 11–22 (1988).
5. G. A. Carpenter and S. Grossberg, "The ART of Adaptive Pattern Recognition by a Self-Organizing Neural Network," Computer **21**, 77–88 (1988).
6. D. E. Rumelhart *et al.*, *Parallel Distributed Processing* (MIT Press, Cambridge, MA, 1986).
7. K. Wagner and D. Psaltis, "Multilayer Optical Learning Networks," Proc. Soc. Photo-Opt. Instrum. Eng. **752**, 86–97 (1987).
8. A. D. Fisher, W. L. Lippincott, and J. N. Lee, "Optical Implementations of Associative Networks with Versatile Adaptive Learning Capabilities," Appl. Opt. **26**, 5039–5054 (1987).

9. M. R. Feldman, S. C. Esener, C. C. Guest, and S. H. Lee, "Comparison Between Optical and Electrical Interconnects Based on Power and Speed Considerations," Appl. Opt. **27,** 1742–1751 (1988).
10. D. Casasent, "Optical Associative Processors for Visual Perception," Proc. Soc. Photo-Opt. Instrum. Eng. **882,** 47 (1986).
11. K. Fukushima, S. Miyake, and T. Ito, "Neocognitron: A Neural Network Model for a Mechanism of Visual Patter Recognition," IEEE Trans. Sys. Man Cybernet. **SMC-13,** 826–834 (1983).
12. D. H. Hubel and T. N. Wiesel, "Receptive Fields, Binocular Interaction and Functional Architecture in Cat's Visual Cortex," J. Physiol. London **160,** 106–154 (1962).
13. D. H. Hubel and T. N. Wiesel, "Receptive Fields and Functional Architecture in Two Nonstriated Visual Area (18 and 19) of the Cat," J. Neurophysiol. **28,** 229–289 (1965).
14. J. W. Goodman, *Introduction to Fourier Optics* (Wiley, New York, 1968).