Proceedings IWISP '96; 4-7 November 1996; Manchester, U.K.
B.G. Mertzios and P. Liatsis (Editors)

An neural approach to invariant character recognition

Iraklis M. Spiliotis[1], Panagiotis Liatsis[2], Basil G. Mertzios[1]and Yannis P. Goulermas[2]

[1]Department of Electrical & Computer Engineering
Democritus University of Thrace
GR- 67100 Xanthi
Greece

[2]Control Systems Centre
UMIST
Manchester M60 1QD
United Kingdom

Abstract

Geometric transformations constitute a difficulty in *optical character recognition* (OCR) *systems*. This work describes the development of an intelligent OCR system based on *higher-order neural networks* (HONNs). These networks can be designed such that their outputs remain invariant to certain geometric distortions, such as translation, rotation and scale. The main obstacle in the practical application of HONNs is the explosion of the weights due to the number of input combinations. This problem is tackled using an efficient object representation scheme called *image block representation* (IBR).

I. Introduction

Invariant object recognition is a major research area in computer vision. A number of approaches have been proposed to address the issue of image correspondence once geometric transformations are applied [1]-[5]. The question remains to select the set of image features which will ensure an acceptable recognition rate.

Higher-order neural networks have developed over the past decade as an alternative to traditional object recognition approaches [6]-[10]. The basic concept of HONNs is the expansion of the input representation space using higher-order combinations of the input terms, such that the mapping from the input to the output space can become more readily obtainable. This idea has a certain appeal in object recognition systems since geometric feature extraction mechanisms can be incorporated within the structure of the HONN. For instance, it has been shown that features such as distances and line slopes defined by point pairs, and angles of similar triangles defined by point triplets are respectfully *invariants* to translation-rotation, translation-scale, and translation-rotation-scale transformations.

As mentioned above, these invariants can be obtained by enriching the input representation space with *all* possible point combinations (up to a certain order). Clearly, this constitutes a serious limitation for the application of HONNs in invariant image recognition, where images may have a spatial resolution of 512x512 pixels. In particular, for an MxN image and n-order point combinations, the number of input terms will be augmented by $(MxN)!/(MxN\text{-}n)!n!$. To allow the use of HONNs to object recognition problems, the technique of coarse coding has been proposed. This decomposes the image into a set of non-overlapping, offset images of coarse resolution, such that the number of input combinations is reasonably bounded. However, coarse coding does not ensure lossless image representation and thus does not allow perfect image reconstruction [11].

This research proposes the use of a simple yet effective object representation scheme called *image block representation* [12]-[14]. This method decomposes the object into a set of non-overlapping rectangular regions, which can then be used to extract the so-called *critical points*, i.e. points of interest on the object. The number of critical points is relatively small (when compared to the number of object pixels) and subsequently they can be used as direct inputs to the HONN architecture. The performance of the system is evaluated in the case of binary character recognition.

II. Higher-Order Neural Networks

A major criticism of single-layered perceptrons [15] was that they were unable to perform non-linear separation, an example being the XOR problem, due to the simplicity of the resulting decision boundaries. A way of dealing with this problem was to generalise the perceptron architecture such that it accommodates intervening layers of neurons, capable of extracting abstract features, thereby resulting to networks that could solve reasonably well any given input-output problem [16]. An alternative approach, based on recent studies of information processing exhibited by biological neural networks [17] as well as Group Method of Data Handling (GMDH) algorithms [18], was the expansion of input representation space by using multi-linear terms. This gave rise to a family of neural networks collectively known as higher-order neural networks.

In general, the output of a first-order neural network is defined by [9]

$$y_i^1 = \begin{cases} f\left(\sum_{n=0}^{1} T_n^{hid}(i)\right) = f\left(T_0^{hid}(i) + T_1^{hid}(i)\right), & \text{for hidden nodes} \\ f\left(\sum_{n=0}^{1} T_n^{out}(i)\right) = f\left(T_0^{out}(i) + T_1^{out}(i)\right), & \text{for output nodes} \end{cases} \quad (1)$$

where f(net) is a nonlinear threshold function such as the sigmoid function and $T_0^k(i)$ is the bias term for output i of the k-th layer (where k takes the values hidden and output) given by

$$T_0^k(i) = w_i^k \quad (2)$$

and $T_1^k(i)$ are the first-order terms for the i-th output unit of the k-th layer

$$T_1^k(i) = \sum_j w_{ij}^k x_j^{k-1} \quad (3)$$

where w_{ij}^k are the interconnection weights for each input x_j of layer $(k-1)$ and output node i of layer k.
Generalising to mixed n-th order networks gives [10]

$$y_i^n = f\left(\sum_n T_n^{hid}(i)\right) = f\left(T_0^{hid}(i) + T_1^{hid}(i) + T_2^{hid}(i) + \ldots + T_n^{hid}(i)\right) \quad (4)$$

for the nodes of the hidden layer, where $T_2^{hid}(i)$ and $T_n^{hid}(i)$ are given by

$$\begin{aligned} T_2^{hid}(i) &= \sum_k \sum_j w_{ijk}^{hid} x_j x_k \quad \text{(second-order terms)} \\ T_n^{hid}(i) &= \sum_n \cdots \sum_j w_{ij\ldots(k-n)}^{hid} x_j \ldots x_n \quad (n\text{th-order terms}) \end{aligned} \quad (5)$$

Consider an object and any two non-identical points A, B on the object. Next an arbitrary translation and/or rotation of the object within the image is applied and points A, B become A' and B'. Since the invariant under translation and/or rotation is the relative distance between any two points on the object, the output of the HONN can be hand-crafted to be invariant to this set of transformations by considering only the second-order terms [9]

$$y_i^{hid} = f\left(\sum_k \sum_j w_{ijk}^{hid} x_j x_k\right) \quad (6)$$

and by constraining the input-hidden weights to satisfy

$$w_{iAB}^{hid} = w_{iA'B'}^{hid} \quad \text{if} \quad d_{AB} = d_{A'B'} \quad (7)$$

where d_{AB} and $d_{A'B'}$ are the Euclidean distances between points A, B and A', B' respectively.
The learning rule for the higher-order neural networks is the backpropagation algorithm, appropriately modified to accommodate the inclusion of the higher-order terms in the hidden layer. The updating rule for the weights of the hidden layer is then given by [10]

$$\Delta w_{ijk} = \eta \delta_i \sum_k \sum_j x_j x_k \quad (8)$$

where η is the learning rate, the δ's are calculated as in the classical backpropagation, while k, j take values which satisfy the invariance constraints.

III. Image Block Representation

A bilevel digital image is represented by a binary 2-D array. Without loss of generality, we consider that the object pixels are assigned to level 1 and the background pixels to level 0. Due to this kind of representation, there are rectangular areas of object value 1 in each image. These rectangles, called *blocks*, have their edges parallel to the image axes and contain an integer number of image pixels.
Consider a set that contains as members all the nonoverlapping blocks of a specific binary image, in such a way that no other block can be extracted from the image (or equivalently each pixel with object level belongs to only one block). It is always feasible to represent a binary image with a set of all the nonoverlapping blocks with object level and this information lossless representation is called *Image Block Representation* (IBR) [12]. Given a specific binary image, different sets of different blocks can be formed. Actually, the nonunique block representation does not have any implications on the implementation of any operation on a block represented image.

The IBR concept leads to a simple and fast algorithm, which requires just one pass of the image and simple bookkeeping process. In fact, considering a $N_1 x N_2$ binary image $f(x,y)$, x=0,1, ... ,N_1-1, y=0,1, ... ,N_2-1, the block extraction process requires a pass from each line y of the image. In this pass all object level intervals are extracted and compared with the previous extracted blocks.

As a result, a set of all the rectangular areas with level 1 that form the object. A block represented image is denoted as

$$f(x,y) = \{b_i : i = 0,1,....,k-1\} \quad (9)$$

where k is the number of the blocks. Each block is described by four integers, the coordinates of the upper left and down right corner in vertical and horizontal axes. The block extraction process is implemented easily with low computational complexity, since it is a pixel checking process without numerical operations. Fig. 1, illustrates the blocks that represent an image of the character d.

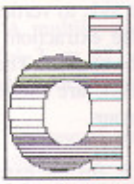

Figure 1. Image of the character d and the blocks.

IV. Critical Points Extraction

An object normalization procedure is first executed in order to facilitate rotation invariant descriptions of the objects. Specifically the maximal axis of the object is found and the whole object is rotated in such a way that the maximal axis has a vertical position and that the upper half of the image object contains the most of the object's maze. At this point, it is necessary to give the following definitions [14]:

1.Group is an ordered set of connected blocks, in such a way that all its intermediate blocks are connected with two other blocks, while the first and last blocks are connected with only one block.

2. Junction point is called a point that it is connected with two other points.

3. End point is called a point that it is connected with only one other point.

4. Tree point is called the point that it is connected with more than two other points.

5. Critical point is called a junction or an end or a tree point.

In this research, a fast non-iterative critical points detection method for block represented binary images is presented. The method has low computational complexity, extracts only critical points and to a degree appears to be immune to locality problems. This is achieved by the use of a suitable neighbourhood at each case. Specifically, groups of connected blocks are formed. Each group is terminated when an adjacent block does not exists for its continuation, or when two or more blocks exist for the continuation of the group.

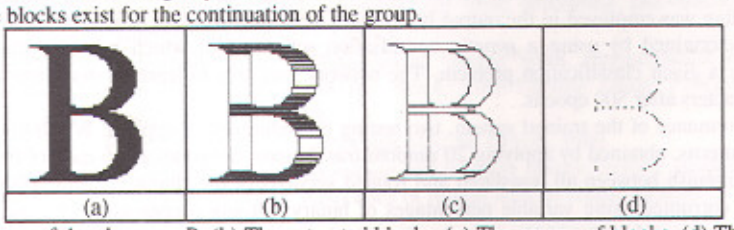

Figure 2. (a). Image of the character B. (b) The extracted blocks. (c) The groups of blocks. (d) The critical points.

Each group defines a local neighborhood and all the necessary processing takes place in this neighborhood. Using a few simple rules for the processing, the groups are checked and labeled to certain categories:

- *Vertical Elongated groups*. The absolute value of the angle of these groups with the horizontal axis is usually greater than 30°. The width of each block of a vertical elongated group should not exceed a threshold value. The connections among the blocks result to junction points, which belong to the thinned line that results from the group. For each pair of connected blocks, one junction point (the central point of the common line segment of two connected blocks) is extracted. For each block we check if the distance among its junction points and its extremities (i.e. the central points of the edges of the small dimension) of the block, exceeds a threshold value.
- *Horizontal Elongated groups*. The absolute value of the angle of these groups with the horizontal axis is smaller than 30°. The width of an horizontal elongated group is significantly greater than its height and also its height appears to have small variation. For the extraction of the junction points the algorithm starts from the left end of

an horizontal elongated group and moves to the right with constant width steps. At each step a junction point is extracted at the middle of the height of the group at this vertical position.

- *Angle groups*. The angle groups are connected with two other groups that lie on the same vertical or horizontal side of the angle group. The width and the height of an angle group are usually small. An angle group should not be connected to a noisy group. If a group has labeled as angle group and it is connected with a noisy group, then the label "angle" is replaced by the horizontal elongated label or the vertical elongated label. Three junction points are extracted from an angle group. The two junction points are extracted due to the connections with the two groups and another one for the formulation of an angle.
- *Noisy groups*. These are small and spurious branches of the object. The noisy groups have width and height less than a threshold and they are connected to only one group, which is not an angle group. In the most cases, the noisy groups are connected from the left or right side to vertical elongated groups or from the up or down side to horizontal elongated groups. In these cases the extraction of junction points from the noisy groups is not acceptable, otherwise a noisy end point would be created. The noisy groups are branches of the object that have small height and width and usually junction points are extracted from the noisy groups, if and only if the noisy group is connected at the ends of an elongated group.

Fig. 2 demonstrates (a) an image of the character B, (b) the extracted blocks, (c) the groups of blocks and (d) the critical points.

V. Results

In this work, we examine the application of IBR and HONNs techniques to the problem of recognising typed characters. The binary data consisted of 26 Latin letter characters (A-Z) and 10 digits (0-9) with a spatial resolution of 64x64 pixels. Since digits '6' and '9' are rotationally equivalent, they were considered as the same pattern. The font style selected for training the OCR system was 'Times New Roman'. Next, the techniques were applied to each of the 35 characters presented in 5 random translations and rotations, giving a total of 175 training patterns.

The first stage of the system was the pre-processing which resulted to the critical points extraction. Here, the rotation normalisation procedure is applied to ensure the success of the IBR scheme. Due to the discrete nature of the image grid, some noise was introduced to the characters, when their maximal axis was set to the vertical position. Next, each of the characters was decomposed into its resulting blocks, and the groups were labeled into vertical/horizontal elongated, angle and noisy. Finally, the critical points were found, using the procedure described in the last section.

The second stage of the system was the classifier. Here, the critical points were fed into a second-order neural network, which had a built-in feature extraction mechanism. This provided invariant classification with respect to translation and rotation. The input layer of the higher-order neural network had 256 inputs. This number was selected to correspond to the maximum number of critical points extracted from any one of the training images. Since a binary representation encoding was employed in the output layer, there were only 8 output units. The number of units in the hidden layer was determined by using a genetic optimisation scheme [10] which provides the minimal-optimal network topology for a given classification problem. The network was able to learn to discriminate between the 35 types of printed characters after 500 epochs.

To evaluate the performance of the trained system, two testing procedures were applied. We firstly tested the system with a set of 700 patterns, obtained by applying 20 random translations and rotations to each of the 35 characters. It was still able to distinguish between all translated and rotated versions of the characters with 100% accuracy. Next, the characters were corrupted using variable percentages of binary salt-and-pepper noise. It was observed that the system was able to distinguish with 100% accuracy for additive noise of up to 10%, and still had a satisfactory performance (recognition accuracy > 70%) for noise levels of 25%.

VI. Conclusions

A new approach to the problem of Optical Character Recognition was presented. The proposed system uses an efficient object representation scheme called image block representation, which decomposes the characters into non-overlapping rectangular regions, which are then used to find the critical points. Next, the critical points are fed into a higher-order neural network with invariances to translation and rotation. This alleviates the problem of the combinatorial explosion of the higher-order terms, associated with the use of HONNs. The optimal number of hidden units, for solving the character recognition problem, was determined using a Genetic Algorithms (GA) scheme. The structure of the neural network was selected to be 256 inputs -5 hidden -8 outputs. The system was able to identify the translated/rotated patterns with 100% recognition accuracy, while it demonstrated robustness to additive noise.

Future work will investigate the performance of the system using a number of font styles as well as handwritten characters. Another interesting application for this type of system is visual inspection. In particular, the problem of detecting blemishes in industrial workpieces, where the only discriminating feature between the two classes is the presence (or absence) of the defect.

Acknowledgments

The authors wish to acknowledge the British Council and the Greek General Secretariat for Research & Development for providing financial support for this research.

References

[1] M.W. Roberts, M. Koch and D.R. Brown, 'A multilayered neural network to determine the orientation of an object', *Proc. Int. Joint Conf. Neural Networks*, Vol. 2, pp. 421-424, 1990.

[2] K. Fukushima, 'A hierarchical neural network model for associative memory', *Biol. Cybern.*, Vol. 50, pp. 105-113, 1984.

[3] S.E., Troxel, S.K. Rogers and M. Kabrinsky, 'The use of neural networks in PRSI target recognition', *Proc. IEEE Int. Conf. Neural Networks*, Vol. 1, pp. 569-576, 1988.

[4] E. Barnard and D. Casasent, 'Invariance and neural nets', *IEEE Trans. Neural Networks*, Vol. 2, No. 5, pp. 498-508, 1991.

[5] N. Papamarkos, I. M. Spiliotis and A. Zoumadakis, 'Character recognition by signature approximation', *Int. Jour. Patt. Rec. Art. Intell.*, Vol. 8, No. 5, pp. 1171-1187, 1994.

[6] T. Maxwell, C.L. Giles, Y.C. Lee and H.H. Chen, 'Nonlinear dynamics of artificial neural systems', in *Neural Networks for Computing*, AIP Conf. 151, UT, pp. 299-304, 1986.

[7] C.L. Giles and T. Maxwell, 'Learning, invariance, and generalisation in higher-order neural networks', *Applied Optics*, Vol. 26, No. 23, pp. 4972-4978, 1987.

[8] M.B. Reid, L. Spirkovska and E. Ochoa, 'Simultaneous position, scale and rotation invariant pattern classification using third-order neural networks', *Neural Networks*, Vol. 1, No. 3, pp. 154-159, 1989.

[9] P. Liatsis, P.E. Wellstead, M.B. Zarrop and T. Prendergast, 'A versatile visual inspection tool for the manufacturing process', *Proc. CCA'94*, Vol. 3, pp. 1505-1510, 1994.

[10] P. Liatsis and Y.J.P. Goulermas, 'Minimal optimal topologies for invariant higher-order neural architectures using genetic algorithms', *Proc. ISIE'95*, Vol. 2, pp. 792-797, 1995.

[11] J. Sullins, 'Value cell encoding strategies', Tech. Rep. 165, CS Dept., Rochester Univ., New York, August 1985.

[12] I. M. Spiliotis and B.G. Mertzios, 'Real-time computation of two-dimensional moments on binary images using image block representation', accepted for publication in *IEEE Trans. Image Process.*

[13] I. M. Spiliotis and B.G. Mertzios, 'Fast algorithms for basic processing and analysis operations on block represented binary images' submitted to *Patt. Rec. Letters*.

[14] B. G. Mertzios, I.M. Spiliotis and N. Papamarkos, 'Image block representation and its applications to manufacturing and automation', accepted *in 5th Int. Work. Time-Varying Image Process. and Moving Object Recognition*, Florence, Italy, September 5-6, 1996.

[15] M. L. Minsky and S. Papert, *Perceptrons*, Cambridge MA: MIT Press, 1969.

[16] H. White, *Artificial Neural Networks: approximation and learning theory*, Oxford: Blackwell, 1992.

[17] D.A. Baylor, T.D. Lamb and K.W. Lau, 'Responses of retinal rods to single photons', *J. Physiol.*, No. 288, pp. 613-634, 1979.

[18] S. J. Farlow (ed.), *Self-organising methods in modeling: GMDH algorithms*, New York: Marcel Dekker Inc., 1984.