Multiresolution locally expanded HONN for handwritten numeral recognition

Cheng-Lin Liu a,*, Jin H. Kim a, Ru-Wei Dai b

a Department of Computer Science, Korea Advanced Institute of Science and Technology, 373-1 Kusong-dong Yuseong-gu, Taejon 305-701, South Korea
b Institute of Automation, Chinese Academy of Sciences, Beijing 100080, PR China

Received 9 September 1996; revised 24 July 1997

Abstract

In this paper, we propose a neural network architecture, multiresolution locally expanded high order neural network (MRLHONN) to solve the problem of handwritten numeral recognition. In this recognition scheme, the multiresolution representation of character image is input into a high order neural network (HONN), while in each resolution, only neighboring pixels are expanded to produce high order input. The property of this architecture is that, the local expansion alleviate the problem of large connecting weight set, and the multiresolution representation remedy the inadequacy of local expansion. Two forms of multiresolution representations, quadtree representation and Gaussian pyramid, were used in experiments. The recognition results demonstrate the efficiency of the proposed architecture. © 1997 Elsevier Science B.V.

Keywords: Handwritten numeral recognition; High order neural network; Local expansion; Multiresolution representation

1. Introduction

Artificial neural networks (ANNs) have been widely used in pattern recognition problems and have achieved promising results. Neural network models that are frequently used in pattern recognition include multi-layer Perceptron (MLP), radial basis function (RBF) network, and learning vector quantization (LVQ) classifier, etc. To test the performance of neural network classifiers, the problem of handwritten numeral recognition is often experimented. Although many works have been done in feature extraction and network structure design to improve the recognition performance, we propose in this paper a new architecture by combining multiresolution representation and high order neural network (HONN).

Compared to other neural network models, HONN (Giles and Maxwell, 1987) is powerful for invariant object recognition and nonlinear approximation. It is reported that 3rd order HONN can recognize objects with translation, scaling and orientation invariance (Spirkovska and Reid, 1992). However, the HONN has two disadvantages, namely, too large number of connecting weights and insufficiency to tolerate distortion. To reduce the connecting weight set while retaining satisfactory recognition performance, strategies of partial connectivity were proposed (Spirkovska and Reid, 1990). An alternative way is to reduce the dimension of pattern description.
Coarse-coding (Spirkovska and Reid, 1992) is an example of dimension reduction, by which the image resolution is reduced and multiple shifted planes were used to remedy the information loss. In addition, Shin and Ghosh (1991) presented a novel architecture, Pi-Sigma network, which has high order nature but is structurally different from conventional HONN. It has much less connecting weights due to its structure but consequently, its approximation ability is compromised.

In recent years, there has been a trend of combining neural networks and image processing techniques in pattern recognition. Particularly, techniques of multiscale and multiresolution analysis are used to improve the approximation and generalization ability of neural networks. Davidson and Rock (1995) combined wavelet transform and HONN by expanding the coefficients of wavelet transform to produce non-linear terms to neural network so as to accelerate the learning process. Lee et al. (1996) used Haar wavelets to decompose the character image into multiple resolutions and orientations, and then used a cluster neural network to recognize handwritten numerals.

The major difficulty of handwritten numeral recognition is the great shape variation and distortion. In order to solve this problem using HONN with reduced connecting weights and improved distortion tolerance, we propose a new architecture, multiresolution locally expanded HONN (MRLHONN). This network has only 2nd-order expansion in local neighborhood. To remedy the recognition inadequacy caused by local expansion, we represent the input pattern in multiresolution, with linear values and high order expansions in each resolution used as inputs of a single-layer neural network. The multiresolution representation of character image is significant to alleviate the problem of locality because a pixel in lower resolution carries information of larger area in original image than a pixel in higher resolution, i.e., a pixel in lower resolution is more global so as to remedy the locality of higher resolution. The proposed architecture is also more economical than coarse-coded HONN because in multiresolution representation, the image of lower resolution has much less pixels.

The efficiency of the proposed architecture has been demonstrated in experiments of totally unconstrained handwritten numeral recognition. In experiments, two forms of multiresolution representations, quadtree representation and Gaussian pyramid, were used. It turns out that the result of Gaussian pyramid representation is better than that of quadtree representation. A database of totally unconstrained handwritten numerals collected in CENPARMI (Concordia University), was used to test the performance. Recognition rate of 97.05% has been achieved on testing data. The result is also compared favorably to some previous results on the same database.

The rest of this paper is organized as follows. Section 2 introduces the multiresolution representation of character image; Section 3 describes the architecture of MRLHONN; Section 4 gives experimental results followed by concluding remarks in Section 5.

2. Multiresolution representation of character image

The multiresolution representation of image is inspired by principles of human visual perception. It bridges the gap between global features and local features, as well as the gap between pixel-level analysis and region-level analysis (Rosenfeld, 1984). In multiresolution representation, the distance between pixels varies with the resolution, the pixels far apart in higher resolution become closer in lower resolution. This interesting property is significant to remedy the insufficiency of locally expanded HONN.

Two simple multiresolution representations, quadtree representation (Rosenfeld and Vanderburg, 1977) and Gaussian pyramid (Crowley and Stern, 1984; Burt, 1984), are used in this paper. The quadtree is the recursive averaging of $2 \times 2$ block of an image, it is much simple for computation. While Gaussian pyramid is the recursive lowpass filtering of an image, or the filtered version of original image with different scales. The Gaussian pyramid is less sensitive to object translation and rotation than quadtree, meanwhile its computation is not expensive. The computation of multiresolution representations are introduced in the following subsections.

Before multiresolution approximation, the size of a character image is normalized to alleviate the effect of size variation. The size of normalized image
is $32 \times 32$. The normalized image is again thinned by Hilditch’s algorithm (Hilditch, 1969) to eliminate the stroke-width variation. The normalization and thinning is essential in most methods for character recognition.

2.1. Quadtree representation

Denote the original image by $f^0(x,y)$, the size of the image is $2^n \times 2^n$ ($n=5$, in the context of this paper), then the pyramid has $n+1$ layers with the lowest layer being the original image and the top level being a single pixel. The resolutions of each layer from the bottom to top are $0, 2^{-1}, \ldots, 2^{-n}$, respectively, and the image of layer $j$ has $2^{n-j} \times 2^{n-j}$ pixels. In quadtree representation, the image of layer $j+1$ is computed from the image of layer $j$ by block averaging:

$$f^{j+1}(x,y) = \frac{1}{4} \left[ f'(2x,2y) + f'(2x,2y+1) 
+ f'(2x+1,2y) + f'(2x+1,2y+1) \right].$$

(1)

By this procedure, all subimages of layer 1 to $n$ are computed from original image recursively. Fig. 1 shows some examples of quadtree representation of numerals. In each row, the leftmost image is the (normalized and thinned) original image, while the other 4 images are subimages of resolutions $2^{-1}, 2^{-2}, 2^{-3}$ and $2^{-4}$, respectively. It turns out that numeral image in resolution $2^{-2}$ can be recognized with little ambiguity.

2.2. Gaussian pyramid representation

In Gaussian pyramid representation, the image of each layer is the lowpass filtered version of its lower level. The lowpass filter for Gaussian pyramid is a circularly symmetric Gaussian filter, with impulse response function of

$$h(x,y) = \frac{1}{2\pi\sigma^2} \exp\left( -\frac{x^2 + y^2}{2\sigma^2} \right).$$

(2)

The corresponding frequency transfer function is

$$H(u,v) = \exp\left( -\frac{u^2 + v^2}{2\sigma^2} \right),$$

(3)

where $\sigma = 2\pi\sigma^{-1}$. The bandwidth of Gaussian filter is proportional to $\sigma$ and inversely proportional to $\sigma$. Burt (1984), and Crowley and Stern (1984) presented fast computation algorithms of Gaussian pyramid, where the image of layer $j+1$ is the lowpass filtered version of $f'(x,y)$. The process of filtering is accomplished by convolution of $f'(x,y)$ and $h(x,y)$:

$$f^{j+1}(x,y) = \sum_{x'} \sum_{y'} f'(x',y') h(2x-x',2y-y').$$

(4)

This equation shows that the pixel value of $f^{j+1}(x,y)$ is the sampled value of filtered $f'(x,y)$ in position of $(2x,2y)$. By tracking the iteration, the image of a layer is the filtered version of the original image, with different filter parameter for each layer. Specifically, the pixel value of $f'(x,y)$ is the sampled value of lowpass filtered version of $f^0(x,y)$ in position of $(2^j x, 2^j y)$. However, in our problem, we expect that the sampling position in each layer is in center of rectangular subregions for convenience of feature extraction and classification. That is to say, for layer $j$ (resolution $2^{-j}$), the sampling position of pixel $f'(x,y)$ should be $(2^j x + 2^{j-1} 2^j y + 2^{j-1})$ in original image. To this end, we compute $f'(x,y)$ directly from $f^0(x,y)$ by lowpass filtering with different parameters. Assuming the scale of Gaussian filter for computing $f'(x,y)$ from $f^0(x,y)$ is $\sigma'$ which corresponds to bandwidth $\sigma$, then the scale for computing $f'(x,y)$ directly from $f^0(x,y)$ is $2^{j-1}\sigma'$ corre-
responding to bandwidth $2^{j-1} \sigma_x$. The impulse response function of Gaussian filter with scale $2^{j-1} \sigma_x$ is

$$h(x,y) = \frac{1}{2\pi(2^{j-1} \sigma_x)^2} \exp \left( -\frac{x^2 + y^2}{2(2^{j-1} \sigma_x)^2} \right).$$

(5)

and the image $f^j(x,y)$ is computed by convolution of $f^0(x,y)$ and $h(x,y)$:

$$f^j(x,y) = \sum_{x'} \sum_{y'} f^0(x', y') \times h((2^j x + 2^{j-1} - x', 2^j y + 2^{j-1} - y)).$$

(6)

Now the remained task is to determine the parameter $\sigma_x$ of Gaussian filter. Burt (1984) gave a discrete Gaussian window for multiresolution image representation. By fitting this window function, the parameter $\sigma_x$ is approximately equal to 1. Empirically, in our problem the parameter should be a little smaller because we use thinned character image which has much smaller portion of foreground. The parameter used in our experiments is $\sigma_x = 2/3$.

3. Network architecture

The network designed for handwritten numeral recognition is a locally expanded HONN (MRLHONN) with multiresolution representation of pattern as inputs. The network has only one layer, if excluding the input layer. For the subimage of a resolution, the pixel values are used as linear inputs, and the production of neighboring pixels as high order inputs. In order to make the inputs have nearly same multitudes, we replace the production of two order inputs. Corresponding to 10 classes of numerals, the network has 342 inputs. For each pixel, a pair of neighboring pixels is connected by a bar in horizontal, vertical or diagonal direction. To construct the network, each pixel contributes to a linear input, and each connecting bar contributes to a higher-order input. By this expansion, for an image with size of $N \times N$, there are $(N-1) \times (N-1) \times 4 + 2 \times (N-1)$ high order inputs altogether. As the result, the image $f^j(x,y)$ produces 64 linear inputs and $7 \times 7 \times 4 = 182$ higher order inputs, $f^2(x,y)$ produces 16 linear inputs and $3 \times 3 \times 4 = 36$ high order inputs, $f^4(x,y)$ produces 4 linear inputs and 6 high order inputs. Totally, the single-layer network has 342 inputs. Corresponding to 10 classes of numerals, the network has 10 outputs. The number of connecting weights is $(342 + 1) \times 10 = 3430$. In comparison, if we use the image $f^2(x,y)$ of single resolution as inputs of polynomial classifier (2nd-order fully expanded HONN), the number of weights is $(64 \times 64 + 1) \times 10 = 40970$, which is a huge number compared to 3430. Therefore, the local expansion on multiresolution representation reduces connecting weights considerably.

With 342 inputs filled into a feature vector $x$, the output of the network is computed by

$$y_j = s(\mathbf{w}_j^T x + \theta_j) = s \left( \sum_{j=1}^{342} \mathbf{w}_{ji} x_i + \theta_j \right),$$

(8)

where $\mathbf{w}_j$ is the weight vector for the $j$th output, and $\theta_j$ is the bias value. The function $s(\alpha)$ is the sigmoid activation function which is given by

$$s(\alpha) = \frac{1}{1 + e^{-\alpha}}.$$  

(9)
Table 1
Experimental results on CENPARMI database

<table>
<thead>
<tr>
<th>Representation</th>
<th>Resolution</th>
<th>Train</th>
<th>Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quadtree</td>
<td>Single</td>
<td>97.53</td>
<td>96.20</td>
</tr>
<tr>
<td></td>
<td>Multiple</td>
<td>97.98</td>
<td>96.55</td>
</tr>
<tr>
<td>Gaussian</td>
<td>Single</td>
<td>98.40</td>
<td>96.65</td>
</tr>
<tr>
<td></td>
<td>Multiple</td>
<td>98.72</td>
<td>97.05</td>
</tr>
</tbody>
</table>

The network is trained by gradient descent rule, by which the modification of weight is
\[
\Delta w_{ji} = \eta (t_j - y_j) y_j (1 - y_j) x_i, \tag{10}
\]
where \( \eta \) is the learning rate, and \( t_j \) is the expected value for \( j \)th output. The connecting weights of the network are modified by iteratively inputting training patterns.

4. Experimental results

The recognition performance of the proposed architecture was tested on CENPARMI database of totally unconstrained handwritten numerals. This database was collected in CENPARMI, Concordia University, and has been frequently used in handwritten numeral recognition research. The database has 6,000 numeral images, 600 for each class. As usual, we use 4,000 images for training and the remained 2,000 for testing.

Two forms of multiresolution representation, quadtree and Gaussian pyramid with single and multiple resolutions were experimented. The recognition results on training and test sets are listed in Table 1. From the recognition results, it turns out that the recognition performance is obviously improved from single resolution to multiple resolution representation, which proves that the lower resolution indeed remedy the inadequacy of local expansion alone in higher resolution. Furthermore, the performance of Gaussian pyramid representation is better than that of quadtree representation. The reason lies in that Gaussian pyramid is less sensitive to translation and rotation. The best result is 97.05% on test data, which is given by MRLHONN with Gaussian pyramid representation. This result is promising enough by itself. To compare our result to that of previous works, we collected some previously reported results on the same database. The results are listed in Table 2, and briefly explained in the following.

The results of Lee et al. (1996), Gader and Khabou (1996), Cho (1997) are all achieved by neural network classifiers, so it is fair to compare them to our result. The basic ideas of their methods are outlined as follows. Lee et al. (1996) used wavelet decomposition of character image as input of a cluster neural network for recognition. The recognition rate on test data of CENPARMI database is 96.80%. Lee et al. (1996) presented an automatic feature generation method for neural network recognition of handwritten numerals. The feature detectors are some subimage masks which are generated randomly and evaluated by information measure or orthogonality measure. A multilayer feed-forward neural network is used in classification. The recognition rate based on orthogonality is 96.50%, and recognition rate of 98.00% was achieved by combining multiple classifiers. Cho (1997) designed three new neural network architectures for handwritten numeral recognition: multiple MLPs, HMM/MLP hybrid network and structure adaptive self organizing map (SOM). The

<table>
<thead>
<tr>
<th>Method</th>
<th>Correct</th>
<th>Substituted</th>
<th>Rejected</th>
<th>Reliability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lee et al. (1996)</td>
<td>96.80</td>
<td>3.20</td>
<td>0</td>
<td>96.80</td>
</tr>
<tr>
<td>Gader 1 (1996)</td>
<td>96.50</td>
<td>3.50</td>
<td>0</td>
<td>96.50</td>
</tr>
<tr>
<td>Gader 2 (1996)</td>
<td>98.00</td>
<td>2.00</td>
<td>0</td>
<td>98.00</td>
</tr>
<tr>
<td>Cho 1 (1997)</td>
<td>97.35</td>
<td>2.65</td>
<td>0</td>
<td>97.35</td>
</tr>
<tr>
<td>Cho 2 (1997)</td>
<td>96.55</td>
<td>3.45</td>
<td>0</td>
<td>96.55</td>
</tr>
<tr>
<td>Cho 3 (1997)</td>
<td>96.05</td>
<td>3.95</td>
<td>0</td>
<td>96.05</td>
</tr>
<tr>
<td>Suen et al. (1992)</td>
<td>93.05</td>
<td>0</td>
<td>6.95</td>
<td>100</td>
</tr>
<tr>
<td>Expert 4</td>
<td>93.90</td>
<td>1.60</td>
<td>4.50</td>
<td>98.32</td>
</tr>
</tbody>
</table>

* Multiple classifiers.
recognition rates produced by the three classifiers are 97.35%, 96.55% and 96.05%, respectively, which are listed in three rows in Table 2.

Suen et al. (1992) reviewed some excellent works of structural methods, and listed some representative results. Since none of an individual classifier can achieve very high reliability (i.e., low substitution rate under rejection) in totally unconstrained handwritten numeral recognition, strategy of multiple expert system (MES) is advocated where a classifier is regarded as an expert and multiple classifiers are combined to overcome insufficiency of individual classifiers. The recognition rate is 93.05% when eliminating substitution. Among the individual classifiers combined, “Expert 4” performs the best, whose result is appended in the last row of Table 2.

To fairly compare our result to the previous ones, the reliability of MRLHONN is also tested by permitting rejection. To obtain reliable recognition, it is required that the difference between the maximum output and the second maximum of the network is larger than a threshold, otherwise the pattern is rejected. The result under rejection is given in Table 3. Compared to previous results of individual classifiers listed in Table 2, we can see that our result is the best among all the individual classifiers. Firstly, recognition rate of 97.05% without rejection is better than 96.80% of Lee et al. (1996), 96.50% of “Gader 1” (1996), 96.55% of “Cho 2” and 96.05% of “Cho 3” (1997). Under threshold 0.2, the performance of our method is again better than that of “Expert 4” in the sense that both substitution and rejection rates are lower. The performance of our method is not so good as that of MES. However, since the mechanism of our method is much different from the previous classifiers, combining MRLHONN and other classifiers is expected to achieve high reliability.

To analyze the reason of misrecognition by MRLHONN, we provide some examples of misrecognition in Fig. 3. In Fig. 3, 3 images of each misrecognized numeral are displayed, which are the original image, thinned image and lowpass filtered image $f(x, y)$, respectively. The numbers under the images denote the true class label and recognition result, respectively. We summarize that the misrecognition is caused by 3 reasons: inherent ambiguity of character shape, information loss caused by preprocessing, and serious distortion. These problems are expected to be solved by more sophisticated methods, such as knowledge-based recognition or multiple classifier combination.

**Table 3**
Recognition reliability of MRLHONN under rejection

<table>
<thead>
<tr>
<th>Threshold</th>
<th>Correct</th>
<th>Substituted</th>
<th>Rejected</th>
<th>Reliability</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1</td>
<td>95.75</td>
<td>2.15</td>
<td>2.10</td>
<td>97.80</td>
</tr>
<tr>
<td>0.2</td>
<td>94.40</td>
<td>1.40</td>
<td>4.20</td>
<td>98.54</td>
</tr>
<tr>
<td>0.3</td>
<td>93.25</td>
<td>0.90</td>
<td>5.85</td>
<td>99.04</td>
</tr>
<tr>
<td>0.4</td>
<td>92.00</td>
<td>0.70</td>
<td>7.30</td>
<td>99.24</td>
</tr>
<tr>
<td>0.5</td>
<td>90.35</td>
<td>0.45</td>
<td>9.20</td>
<td>99.50</td>
</tr>
</tbody>
</table>

To analyze the reason of misrecognition by MRLHONN, we provide some examples of misrecognition in Fig. 3. In Fig. 3, 3 images of each misrecognized numeral are displayed, which are the original image, thinned image and lowpass filtered image $f(x, y)$, respectively. The numbers under the images denote the true class label and recognition result, respectively. We summarize that the misrecognition is caused by 3 reasons: inherent ambiguity of character shape, information loss caused by preprocessing, and serious distortion. These problems are expected to be solved by more sophisticated methods, such as knowledge-based recognition or multiple classifier combination.

**Fig. 3.** Some of misrecognized samples by MRLHONN.
5. Conclusion

A neural network architecture, multiresolution locally expanded HONN (MRLHONN) is proposed to recognize unconstrained handwritten numerals. In this network, the input pattern is represented in multiresolution by quadtree or Gaussian pyramid. In each resolution, the pixel values are used as linear inputs, and only expansion of neighboring pixels are used as high order inputs. The local expansion reduces connecting weights considerably, while the multiresolution representation remedy the insufficiency of local expansion. The recognition performance of this method was demonstrated on a totally unconstrained handwritten numeral database. Comparison of our result to that of previous works proves the efficiency of the proposed architecture.

Acknowledgements

The authors would like to thank the anonymous referees for their valuable comments which led to the improvement of this paper.

References


