

Evaluating a Zoning Mechanism and Class-Modular Architecture for Handwritten Characters Recognition

Sandra de Avila¹, Leonardo Matos¹, Cinthia Freitas²,
and João M. de Carvalho³

¹ Federal University of Sergipe, Comp. Depart., São Cristóvão, SE, Brazil

² Pontifical Catholic University of Parana, Curitiba, PR, Brazil

³ Federal University of Campina Grande, Campina Grande, PB, Brazil
{sandraeliza,lnmatos}@ufs.br, cinthia@ppgia.pucpr.br,
carvalho@dee.ufcg.edu.br

Abstract. In this article we propose a feature extraction procedure based on directional histograms and investigate the application of a nonconventional neural network architecture, applied to the problem of handwritten character recognition. This approach is inspired on some characteristics of the human visual system, as it focus attention on high spatial frequencies and on the recognition of local features. Two architectures were tested and evaluated: a conventional MLP (Multiple Layer Perceptron) and a class-modular MLP. Experiments developed with the Letter database produced a recognition rate of 93.67% for the class-modular MLP. Other set of experiments utilized the IRONOFF database resulting in recognition rates of 89.21% and 80.75% for uppercase and lowercase characters respectively, also with the class-modular MLP.

Keywords: Handwritten characters recognition, Class-modular architecture, Directional histogram.

1 Introduction

The recognition of handwritten characters has been a topic widely studied during the recent decades because of both the theoretical challenges in pattern recognition and its many possible applications, such as automatically processing postal ZIP codes from mail pieces and money amount in bankchecks. However, it proves to be a challenging problem due to the large variance the data may exhibit. Not only there are changes and distortions from one writer to another, but even for samples produced by the same writers [11].

In this paper we propose three steps approach for recognition of handwritten characters: (1) preprocessing, (2) feature extraction and (3) classification. The role of *pre-processing* is to segment the pattern of interest from the background. *Feature extraction* represents the image of each character by a set of significant features, i.e., each character is represented by a feature vector. *Classification* recognizes the class to which the analyzed character belongs, according to the provided features.

The experimental results we present were obtained with the IRONOFF database [13] from Nantes University, France, that contains off-line character images of the English alphabet, with 10685 instances of lowercase and 10681 instances of uppercase characters. In addition, some preliminary tests were performed with the Letter database from

UCI repository of machine learning [5]. The preliminary tests were used to investigate only the classifier effectiveness, since the feature extraction procedure is already performed for this database. Both databases are similar, as they contain equal number of classes (26) and approximately the same number of classes instance: 20000 (Letter) and 21366 (IRONOFF).

This paper is organized as follows. Section 2 presents the feature extraction procedure for handwritten characters recognition; Section 3 describes the architecture and training of the class-modular neural network; Section 4 presents experimental results and some discussions; finally, Section 5 draws some concluding remarks.

2 Feature Extraction Procedure

Studies in the area of vision show that human visual perception is strongly influenced by high spatial and temporal frequency features [1]. Contours or edges, where the contributions of high spatial frequencies dominate, plays a primordial role in the recognition of forms and interpretation of visual stimulus, as the sense of deepness. Since handwriting characters are visual representations of handmade strokes, most relevant information for the recognition task is extracted from contours.

In this article we present a feature extraction procedure based on directional and curvature histograms for the contour image. A histogram is a frequency distribution of some punctual feature, i.e. a feature of an isolated pixel. Some punctual features used by Shi et al. [10] are gradient and curvature.

Several existing operators can be used to calculate an image gradient. In order to make gradient computation more efficient, we create a gray scale image from the original binary one using a spatial lowpass filter. This grayscale image is used to obtain the gradient phase and curvature.

The directional histogram is a useful tool for emphasizing differences between segments of character contours. It captures the slant of other characteristics concerning their shape. Nevertheless the histograms does not point out where those features occur. An isolated histogram can detect an interesting feature, like a rounding stroke, but it does not reveal where it is located on the character contour. Hence, it does not provide sufficient information for distinguishing, for example, between digits six and nine, as it does not points out where the looping is. In order to mitigate this problem, we developed a zoning mechanism which allows associating directional and curvature features with the position where they occur.

The zoning mechanism adopted does not impose zones of fixed dimension, as it splits the image into non-overlapping regions, therefore avoiding the need to perform a scale normalization preprocessing. Thus, the image preserves all the original geometric properties. The association of directional and curvature histograms with zones of different sizes captures structural information about the character image. In a typical image of digit two, for instance, the pixels with high curvature usually occur in the upper right and bottom left corner, it is also usual that a straight line segment covers the bottom of the image and a rounding segment is present at the top. This kind of information is captured when zones of relative size are associated with the use of histograms.

2.1 Algorithm

The algorithm that summarizes the feature extraction process discussed in this section is stated as follows in two phases. The first performs the preprocessing operations and the second the features vector extraction.

Preprocessing

1. The contour of the input image is saved in a list of points. Both the inner and outer contours were used in this work.
2. A spatial lowpass filter is applied to the binary input image r times in order to yield a grayscale image. In the experiments described in this paper we used $r = 4$ (Fig. 1(a)).
3. A normalization procedure is applied to the image of step 2 in order to limit the grayscale variation to the interval $[0, 1]$.
4. Curvature (Fig. 1(b)) and gradient phase (Fig. 1(c)) are computed for the grayscale image from step 3, as described in Shi et al. [10], only for the pixels saved in step 1.

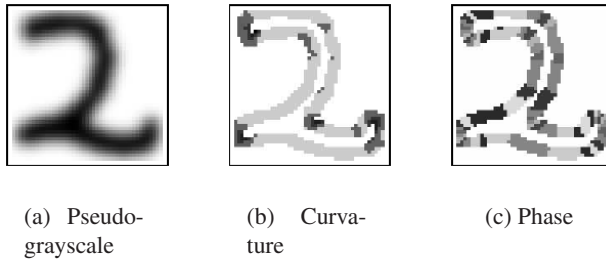


Fig. 1. Feature extracted from image

Feature Extraction Procedure

1. Image $F_{r \times c}$ is partitioned into zones with approximately the same size. If $r \approx c$ it is segmented into 16 zones (4×4). If $r > 1.25c$, F is segmented into 5×4 zones. If $c > 1.25r$, F is segmented into 4×5 zones (Fig. 2).
2. The interval $[0, \pi]$ is partitioned into 10 classes of equal amplitude $\pi/10$. If two angles differ from each other by π radians, they are considered the same.
3. The curvature is computed as in Shi et al. [10] and the interval $[-\infty, \infty]$ is divided into 5 classes.
4. For each zone the number of pixels from each class of direction and curvature is counted. This leads to histograms with 15 classes.
5. Each histogram is normalized.
6. The histograms are concatenated in order to compose the feature vector.

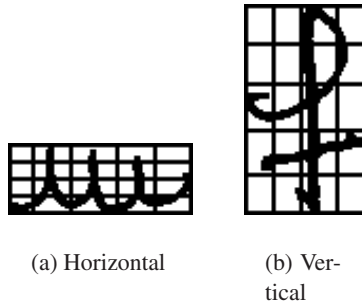


Fig. 2. Rectangular images zones

3 Classification

In character recognition systems, the type or format of the extracted features must match the requirements of the chosen classifier [12]. Multi-Layer Perceptron Neural networks (MLP-NN) classifiers have proven to be powerful tools in pattern recognition [14]. This work investigates two MLP architectures: Conventional and Class-modular, such as Oh and Suen [6] and Kapp et al. [3]. The class-modular architecture is described next.

3.1 Class-Modular Architecture

According to Oh and Suen [6], one property of the conventional MLP architecture is that all K classes share one single network. The essential task in designing a character recognition system is to choose features with good discriminative power and the network should segment the chosen feature space between into the K classes regions. However, determining the optimal decision boundaries for the K -classification network for character recognition in a high-dimensional feature space is very complex and can seriously limit the recognition performance of character recognition systems using MLP.

The conventional MLP-NN architecture has a “rigid” structure in which all the K classes are altogether intermingled. This architecture cannot be locally modified or optimized for each class. Principe et al. [8] and Oh and Suen [6] mention specific convergence problems with one large network, when the training set is not large enough compared with the classifier size, i.e., the number of free parameters in the classifier.

To overcome such limitations Oh and Suen [6] introduce the class modularity concept to the MLP classifier. To implement this concept, the original K -classification problem is decomposed into K 2-classification subproblems, one for each of the K classes. A 2-classification subproblem is solved by the 2-classifier specifically designed for the corresponding class. The 2-classifier is in charge of only one specific class, discriminating that class from the other $K - 1$ classes.

The MLP for a 2-classifier is illustrated in Fig. 3(a). The modular MLP classifier consists of K subnetworks, M_{wi} for $0 \leq i < K$, each responsible for one of the K classes. The specific task of M_{wi} is to classify two groups of classes, Ω_0 and Ω_1 where $\Omega_0 = \{w_i\}$ and $\Omega_1 = \{w_k | 0 \leq k < K \text{ and } k \neq i\}$.

Designing subnetworks M_{wi} follows the same way as for the conventional MLP. In each of the subnetworks, the three layers are fully connected. The input layer has d nodes to accept the d -dimensional feature vector. The output layer has two output nodes, denoted by O_0 and O_1 for Ω_0 and Ω_1 , respectively.

The architecture for the entire network constructed by K subnetworks is shown in Fig. 3(b). The feature module extracts a feature vector X which is shared by all K classes. X is applied to the input layer of all subnetworks and each M_{wi} uses its own weights set to produce an output vector $D = (O_0, O_1)$. The values of O_0 constitutes the final decision [6] [3].

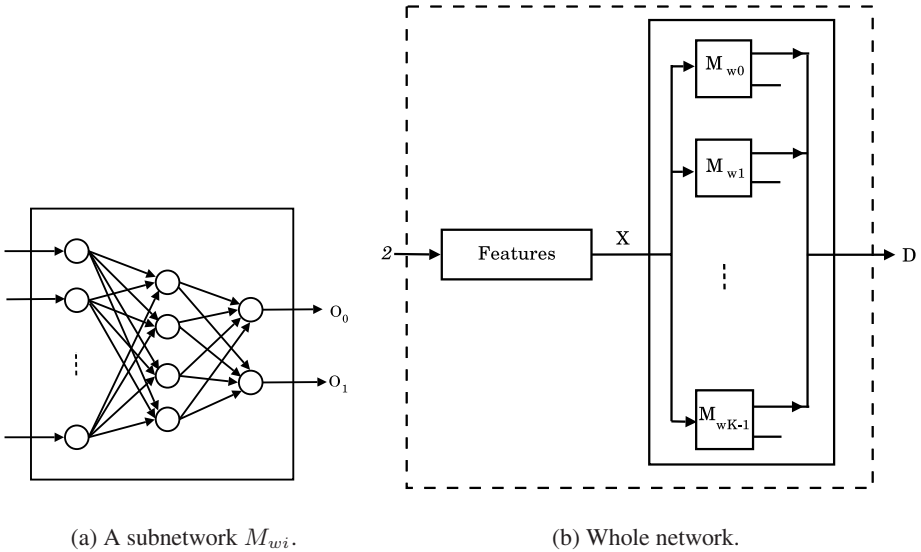


Fig. 3. Class-modular MLP architecture [6]

For training, each of the K 2-classifiers (subnetworks) is trained independently of the other classes, the error-backpropagation algorithm being applied in the same way as for the conventional MLP. A distinct training set is built for each subnetwork, as follows: the samples in the original training set are organized into two groups, Z_{Ω_0} and Z_{Ω_1} such that Z_{Ω_0} has the samples from classes in Ω_0 and Z_{Ω_1} the ones from classes in Ω_1 . For recognition of the input character patterns, the class decision module considers only the values of Z_{Ω_0} from each subnetwork and employs a simple winner-takes-all scheme to determine the final class. Therefore, the winner class is associated with the subnetwork which provides the largest output to the desired pattern.

Oh and Suen [6] presented the effectiveness and superiority of the class-modular MLP compared to the classical nonmodular approach, as shown in Table 1.

Kapp et al. [3] compared the conventional and class-modular architectures for handwritten recognition applied to bankchecks. Experiments confirm that the modular network shows a superior recognition performance. The recognition rates obtained were 77.08% for the nonmodular network and 81.75% for the modular network.

Table 1. Recognition rates for test set of [6]

	Classes	Conventional (%)	Modular (%)
Numerals	10	94.15	97.30
English capital letters	26	81.03	91.11
Touching numeral pairs	100	57.06	75.18
Hangul characters	352	22.46	68.75

Freitas et al. [2] investigated many non-symmetrical zoning mechanisms for hand-written characters recognition, based on analysis of confusion matrix of individuals classifiers (class-modular). The recognition rates obtained were: 4 zones = 82.89%, 5Horizontal zones = 81.75%, 5Vertical zones = 80.94%, 7 = 84.73% and 5Horizontal-5Vertical-7 zones = 85.9%. The database used in experiments was IRONOFF, specifically uppercase characters.

The partitioning strategy of the training set discussed in this article is known in literature as “one-vs-all” classification. This is a common way to deal with binary classifiers, such as SVM (Support Vector Machine), to solve the problem of multiclass classification. Rifkin and Klatau [9] reviewed extensively the existing literature concerning one-vs-all approach and they also defended the superiority of that approach, as we did in this section.

Therefore, there is considerable evidence regarding the effectiveness and superiority of the class-modular architecture, compared to the conventional architecture, in terms of recognition performance. The experiments and results that we describe next, support this conclusion.

4 Experiments and Results

The conventional and modular networks were implemented via the SNNS¹ simulator. For training, both for the Letter and the IRONOFF databases, we used the same parameter values, shown in Table 2. The results obtained for each database are presented next.

4.1 Experiments for Letter Database

For the conventional network, the MLP was composed by 16 neurons in the input layer, one hidden layer with 64 neurons and 26 neurons in the output layer. Training terminated at the 1000-th epoch. The achieved recognition rate was 83.10%.

For the class-modular networks, each one of the K 2-classifiers consists of an input layer of 16 neurons, one hidden layer of 64 neurons and an output layer of 2 neurons. The maximum number of training epochs was 100.

The metrics for classifier performance evaluation were *sensitivity*, i.e., the percentage of positive instances that were classified as positive (true positive), and *specificity*, i.e., the percentage of negative instances that were classified as negative (true negative). These metrics were extracted from the confusion matrix for a two-class classification problem.

¹ Available in <www-ra.informatik.uni-tuebingen.de/SNNS/>

Table 2. Parameter values applied for training of the conventional and the modular networks

Parameter	Value
Layers	3
Learning algorithm	<i>backpropagation</i>
Learning rate	0.02
Activation function	sigmoid
Initial weights	random $[-1,1]$

The average rates obtained for the set of K 2-classifiers were 90.19% (sensitivity) and 99.81% (specificity). The higher specificity rate is justified by the fact that training and test sets for a 2-classifier are not balanced between the two classes, as the number of negative samples is much higher than the number of positive samples.

The global classifier output is achieved by combining the 2-classifiers individual outputs. In this work we used a simple winner-takes-all scheme to determine the final class. The global recognition rate by using the Letter database was 93.67%. This is a significant improvement compared to the 83.1% recognition rate produced by the conventional neural network, obtained despite having used a winner-takes-all strategy in decision module.

4.2 Experiments for IRONOFF Database

The IRONOFF database contains a large diversity of handwritten character images, provided from several authors with distinct age, sex and social level. It captures the diversity of writing style that is a major source of difficult in handwritten character recognition. In Fig. 4 we show some samples from uppercase and lowercase off-line character images from IRONOFF database.

For the lowercase and uppercase characters sets we used the same nonmodular MLP. The network consists of an input layer of 375 neurons, which corresponds to length the feature vector, one hidden layer of 256 neurons and an output layer of 26 neurons. Training terminates at the 1000-th epoch. The recognition rates obtained were 56.14% to lowercase characters and 63.49% to uppercase characters.

The class-modular networks were composed by 375 neurons in the input layer, 256 neurons in an one hidden layer and 2 neurons in the output layer. It was use 250 epochs for training.

The same metrics applied to the Letter database to classifier performance evaluation were used to IRONOFF database. Therefore, the sensitivity and specificity average rates were for lowercase characters 70.35% and 90.84% and for uppercase characters 84.29% and 99.86%, respectively. The global recognition rate for lowercase characters was 80.75% and for uppercase characters was 89.21%. These results match the best available thus far in the literature for the IRONOFF database, as illustrated in Table 3.

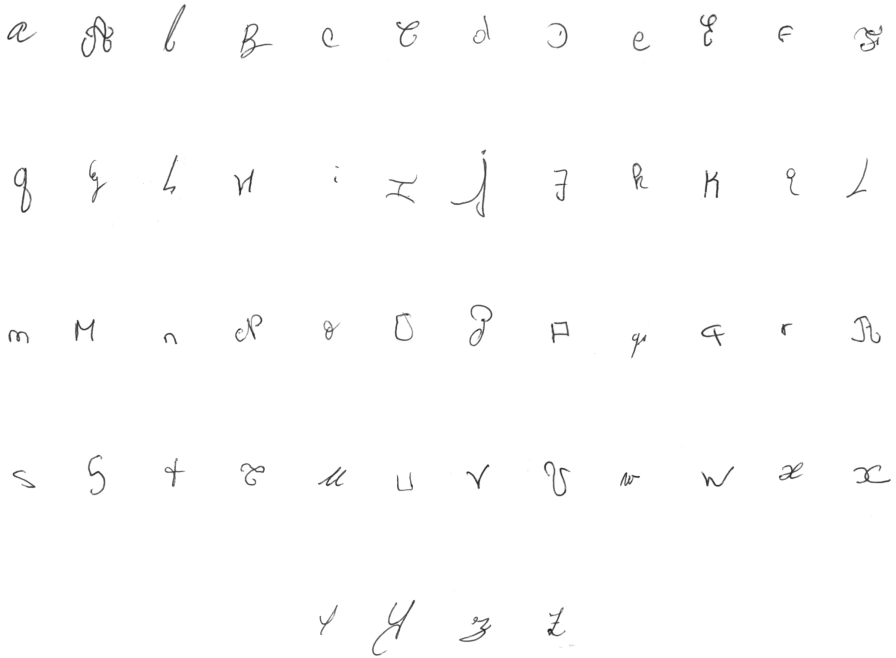


Fig. 4. IRONOFF's samples

4.3 Discussions

Some concluding remarks can be made based on the experimental results reported in the previous section.

- The class-modular network is superior in terms of convergence over the conventional network. This is due to the fact that, for the modular network, the classifier has been decomposed into much smaller subclassifiers, with fewer parameters to estimate.
- The class-modular network has a better performance than the non-modular network, in terms of recognition power.

In addition, training for the class-modular architecture can be easily parallelized, because each sub-network is trained independently of the others, differently from classical architecture in which all the classes are altogether intermingled.

The experimental results for the class-modular classifier are comparable to the best reported in the literature for IRONOFF database (see Table 3). In fact, both the class-modular architecture and the feature extractor proved to be quite appropriate for the handwritten characters recognition problem. The feature extraction method had already produced, in a previous work, one of the best recognition rates for the NIST database [4]. Furthermore, to the best of our knowledge, the combination of these two approaches had not yet been reported by anyone, being therefore an original contribution of this work.

Table 3. Recognition rates

	Letter (%)	Lowercase (%)	Uppercase (%)
Poisson et al. [7]	–	80.5	89.9
Freitas et al. [2]	–	–	85.9
Matos [4]	95.6	–	–
Rifkin and Klatau [9]	96.4	–	–
Conventional	83.1	56.1	63.5
Class-modular	93.6	80.7	89.2

5 Conclusions

In this paper we propose and evaluate a zoning mechanism as well as a class-modular architecture for the recognition of handwritten characters.

Directional histograms applied to the zones were used for feature extraction. This approach was inspired on some aspects of the human visual system, which focus special attention on high spatial features and on the recognition of local features.

The experimental results reveal the superiority of the class-modular network in terms of convergence and recognition rate, over a conventional MLP-NN. The results match the best recognition rates reported in the literature, as shown in Table 3.

The original contribution of this paper is the proposal of a zoning mechanism which combined to a class-modular architecture provides an efficient classifier for the handwritten character recognition problem.

References

1. Cormack, L.K., Bovik, A.I.: Computation models of early human vision in Handbook of image and video processing, p. 891. Academic Press, San Diego (2000)
2. Freitas, C.O.A., Oliveira, L.E.S., de Bortolozzi, F., Aires, S.B.K.: Handwritten character recognition using non-symmetrical perceptual zoning. International Journal of Pattern Recognition and Artificial Intelligence 21, 1–21 (2007)
3. Kapp, M.N., Freitas, C.O.A., Sabourin, R.: Methodology for the design of NN-based month-word recognizers written on brazilian bank checks. International Journal of Image and Vision Computing, Image and Vision Computing 25, 40–49 (2007)
4. Matos, L.N.: Utilização de redes bayesianas como agrupador de classificadores locais e global. PhD thesis, Federal University of Campina Grande, Brazil (2004)
5. Newman, D.J., Hettich, S., Blake, C.L., Merz, C.J.: UCI repository of machine learning databases, Irvine, CA: University of California, Department of Information and Computer Science (1998)
6. Oh, I.-S., Suen, C.Y.: A class-modular feedforward neural network for handwriting recognition. Pattern Recognition 35(1), 229–244 (2002)
7. Poisson, E., Viard-Gaudin, C., Lallican, P.M.: Multi-modular architecture based on convolutional neural networks for online handwritten character recognition. In: IEEE Proceedings of the 9th International Conference on Neural Information Processing, vol. 5, pp. 2444–2448 (2002)

8. Principe, J.C., Euliano, N.R., Lefebvre, W.C.: Neural and adaptive systems: fundamentals through simulations. John Wiley & Sons, New York, USA (1999)
9. Rifkin, R., Klautau, A.: In defense of one-vs-all classification. *The Journal of Machine Learning Research* 5, 101–141 (2004)
10. Shi, M., Fujisawa, Y., Wakabayashi, T., Kimura, F.: Handwritten numeral recognition using gradient and curvature of grayscale image. *Pattern Recognition* 35(10), 2051–2059 (2002)
11. Suen, C.Y., Guo, J., Li, Z.C.: Analysis and recognition of alphanumeric handprints by parts. *IEEE Transactions on Systems, Man and Cybernetics* 24(4), 614–631 (1994)
12. Trier, O., Jain, A., Taxt, T.: Feature extraction methods for character recognition – a survey. *Pattern Recognition* 29(4), 641–662 (1996)
13. Viard-Gaudin, C., Lallican, P.M., Binter, P., Knerr, S., The, I.R.E.S.T.E.: on/off IRONOFF) dual handwriting database. In: *Proceedings of the Fifth International Conference on Document Analysis and Recognition*, pp. 455–458, Bangalore, India (1999)
14. Zhang, G.P.: Neural networks for classification: a survey. *IEEE Transactions on Systems, Man and Cybernetics, Part C: Applications and Reviews* 30(4), 451–462 (2000)