Recognition of handwritten digits by image processing and neural network

Gilles Burel, Isabelle Pottier, Jean-Yves Catros

RECOGNITION OF HANDWRITTEN DIGITS BY IMAGE PROCESSING AND NEURAL NETWORK

Gilles BUREL[1] & Isabelle POTTIER & Jean-Yves CATROS
Thomson CSF-LER, Avenue de Belle Fontaine, 35510 CESSON-SEVIGNE, France

ABSTRACT:
Recognition of handwritten digits has been one of the first applications of neural networks. Efficient methods have already been proposed to solve this task. We propose an intermediate approach between classical methods, which are based on extraction of a small set of parameters, and pure neural methods, in which the neural network is fed with raw image data. Complexity and learning time are reduced with still good performances. On a data base of 2589 digits coming from 30 people, we provide experimental results and comparisons of various parameters and classifiers.

KEYWORDS: *Handwritten Digit Recognition, Machine Learning, Neural Networks, Backpropagation, Image Processing.*

1 Introduction

Character recognition has been the subject of many researches over the past thirty years [2][9]. For handwritten digits, the main applications are recognition of handwritten zip codes [4], and cheque processing. Here, we focus on "Off-Line" recognition with unknown scriptor. We consider only isolated digits: the problem of segmentation could be performed by Shridhar's algorithm [7] or it can be discarded by means of prefilled forms.
In previous works, recognition of handwritten digits by classical methods has been studied, among others, by Mitchell [4] (model matching method with morphological features) and by Shridhar [6] (rule based system with parameters on profiles). In the neural networks field, many works have been published over the past three years. The neural network is generally fed with an image of the digit, normalized in size (e.g. [3] [8]).

We propose a method based on the following remarks:

1. On one hand, working directly on raw image data entails the use of a huge neural network, and a very long training because the neural network has to solve a very complex task.

2. On the other hand, working only on a small set of parameters extracted from the digit may create ambiguities in the parameters space. Furthermore, the optimal set of parameters is difficult to select.

The digit profiles [6] bring almost all metric information about the digit, because it is quite possible to rebuild the entire image from them. However, in order to reduce the loss of information and to select a significant set of parameters, our neural net is fed with the complete curves of the profiles and not only with a small set of parameters extracted from them. Furthermore, other features which provide statistical and morphological information about the digit are added to improve the recognition rate. Finally, local connections and shared weights are used, because they reduce the number of free coefficients inside the network, and then they increase its generalization capability.

[1]new address: UBO/LEST, 6 avenue Le Gorgeu, 29285 Brest cedex, France

2 Pre-processing

2.1 The metric features

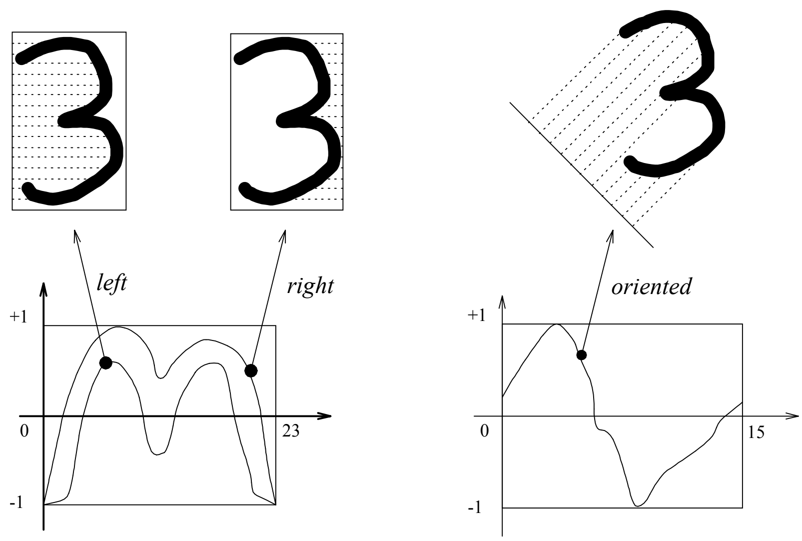

Figure 1: *The left, right, and oriented profiles*

We propose to take advantage of metric features which are normalized digit profiles (fig 1). As suggested by Shridhar [6], a linear interpolation is performed on areas where the digit is broken (e.g. some people write the digit "5" with a top bar unconnected to the rest of the body). We add also an oriented profile (under 45^o) to reduce ambiguities.

2.2 The statistical features

Some global information is also used. It consists in characterizing the distribution of the digit pixels inside the including frame. This kind of feature has been proposed by Beauville [1]. The including frame is divided in regions by 6 different ways (fig 2). There is a total of 20 regions, which produce 20 features $R_j = n_j/n$, where n_j is the number of digit pixels inside region j, and n is the total number of digit pixels.

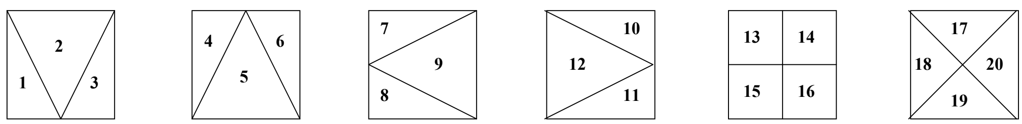

Figure 2: *The 20 regions defined with 6 masks*

2.3 The morphological features

Cavities of different types and holes are the last characteristics used, by the means of their relative surfaces. We define 5 cavity types: West, East, North, South and Center. Adding the hole, we obtain 6 morphological features (fig 3). These features were introduced by Mitchell [4] to specify a symbolic model of the digit.

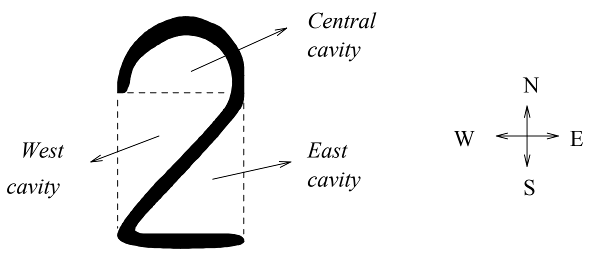

Figure 3: *Some morphological features*

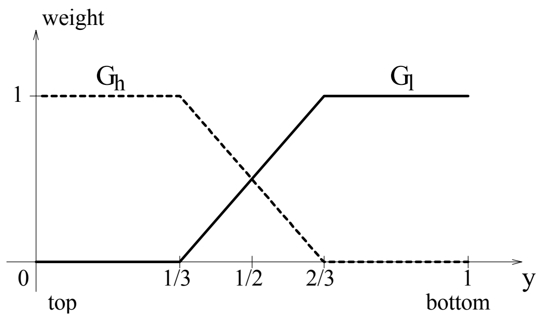

Figure 4: *Weighting of morphological features*

Each morphological feature could be given to the neural net input as the surface of the feature divided by the total surface of the whole set of morphological features. Furthermore, to consider information about spatial arrangement, a feature is represented in fact by 2 values which are surfaces weighted by functions G_h and G_l (fig 4). For instance, the West cavities are represented by the 2 following values where M_{CW} (resp. M_{CA}) is the matrix of the West cavities (resp. all cavities and holes). $M_{CW}(r, c)$ is set to 1 if the pixel at row r and column c belongs to a West cavity, else it is set to 0.

$$CW_h = \frac{\sum_{rows} \left\{ \sum_{columns} M_{CW}(r,c) G_h(r) \right\}}{\sum_{rows} \left\{ \sum_{columns} M_{CA}(r,c) G_h(r) \right\}} \qquad CW_l = \frac{\sum_{rows} \left\{ \sum_{columns} M_{CW}(r,c) G_l(r) \right\}}{\sum_{rows} \left\{ \sum_{columns} M_{CA}(r,c) G_l(r) \right\}}$$

3 The neural network

The neural network is a Multi-layer Perceptron. The neuron model is a weighted adder followed by a non-linear function (hyperbolic tangent). It includes an adaptative bias. The neural network is trained with the backpropagation algorithm [5]. The proposed network consists of 5 layers (fig 5). Instead of a full-connected structure, we take advantage of the neighborhood notion between points of the profile curve, by introducing local connections and shared weights. Thus, each neuron of the second layer becomes responsible for a particular area of a profile.

The network has 127 neurons. Among the 2043 connections, only 1319 are independant because many weights are shared.
We use share weights and local connections for the second and third layers. These layers are in charge of profiles. On the second layer, 5 lines of 10 neurons are in charge of the right and left profiles. Each neuron receives 12 inputs (6 by profile). The inputs are shifted by step of two when we move from a neuron to its neighbor inside a line. Neurons on the same line share their weights. On the third layer, 5 lines of 5 neurons are in charge of the right and left profiles. Each neuron receives 2 inputs from the 2 neurons which are assigned to it in the previous layer. We share the two weights to create an average between the inputs in order to reduce the effect of slight local distorsions of the profiles. A similar

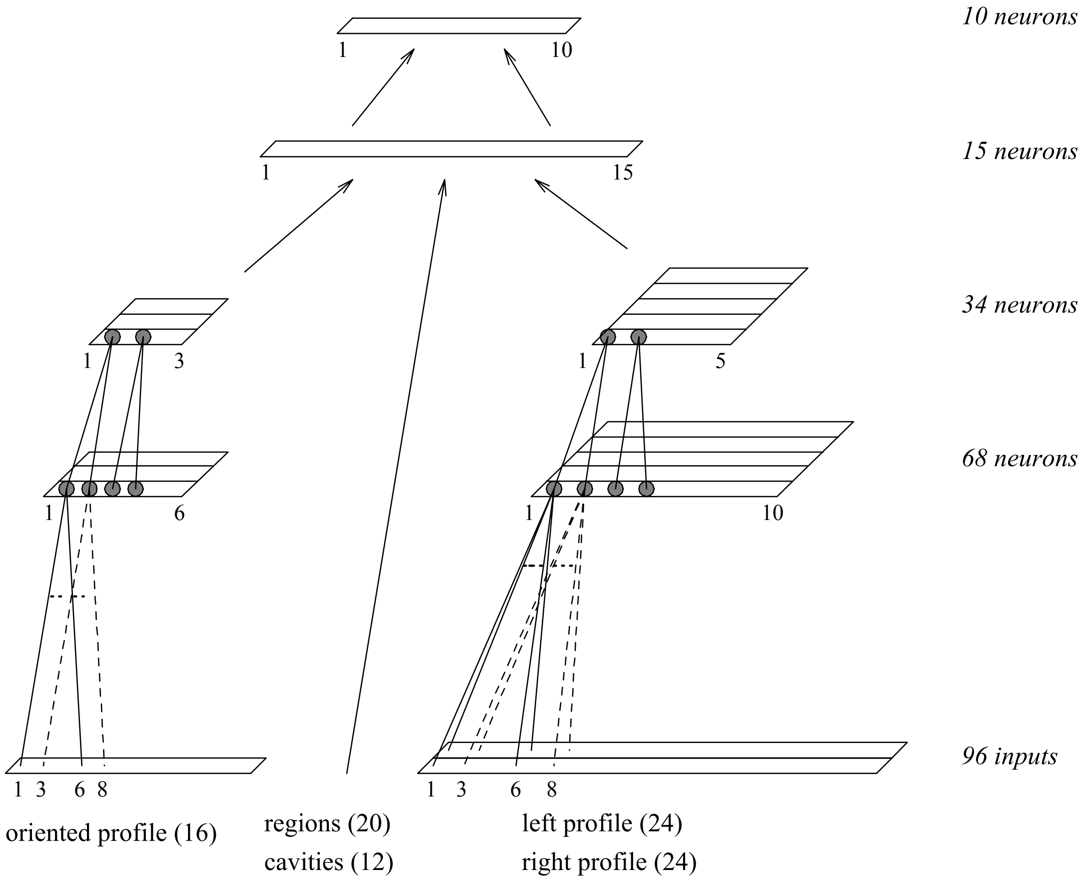

Figure 5: *Structure of the multi-layer perceptron*

strategy is used for the oriented profile.

4 Experiments

The neural network is trained on a set of 1414 digits coming from 16 scriptors, and evaluation is performed on 1175 digits coming from 14 other scriptors.

4.1 Comparison between classifiers

Method	learning rate	generalization rate	multiplication number
network LC5	**98.5%**	**93.6%**	**2439**
network FC4	99.0%	89.3%	2642
1-nn	100.0%	88.5%	158368
3-nn	97.6%	88.2%	158368

We obtain a generalization rate of 93.6% with the network of figure 5 (network LC5). Some recognition results are shown on figure 6. Under each digit, the system has written the class it has found, the real class, and the confidence (difference between the two highest network outputs). If we accept a reject, it is possible to reduce the error rate. When the confidence is lower than a threshold, no decision is taken (reject). An error rate of 1% may be obtained, at the cost of a reject rate of 15%.
In the previous chart, some comparisons with other classifiers are summarized in decreasing order for

the generalization rate. The network FC4 stands for a full-connected network comprising 4 layers of size 96, 20, 15, and 10. "k-nn" designs the k-nearest-neighbors classifier.

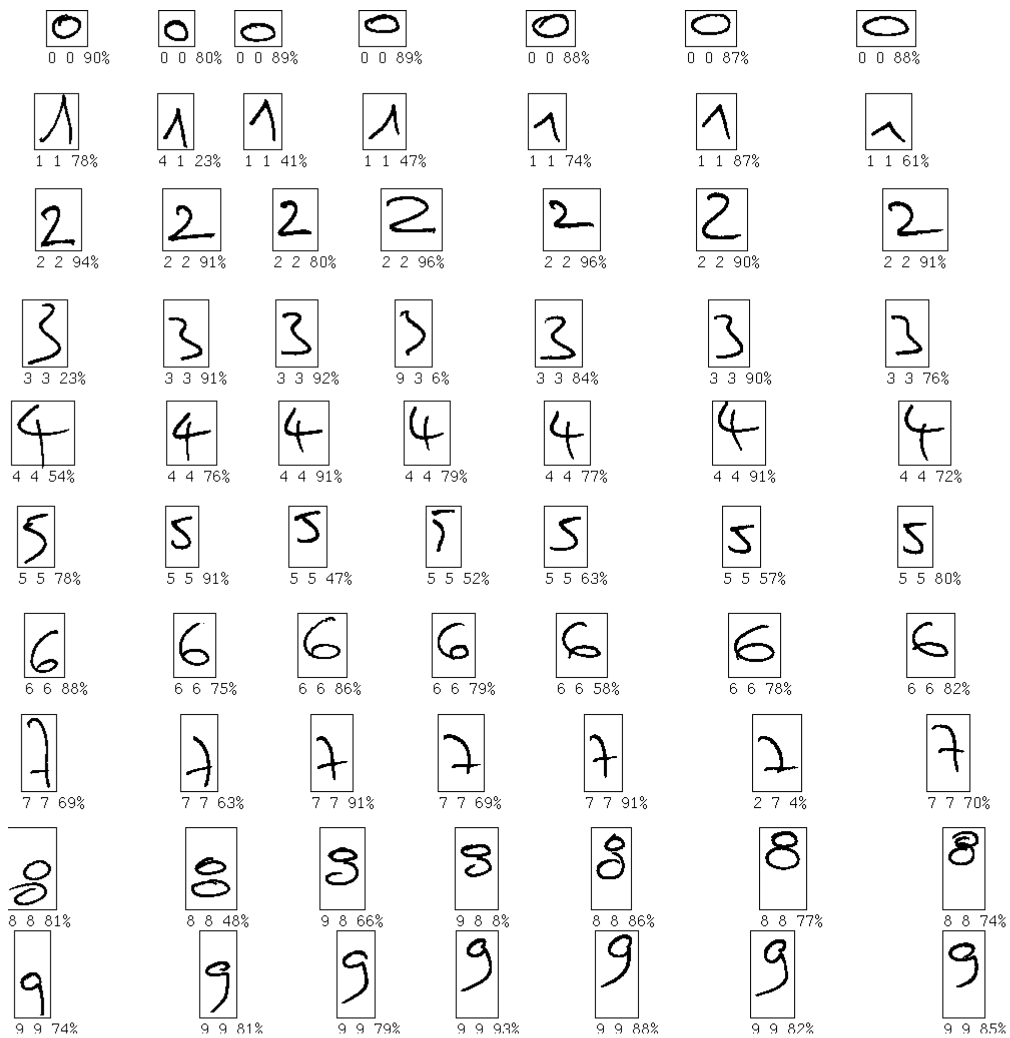

Figure 6: *Some results*

4.2 Evaluation of the features

METHODS / FEATURES	network FC2	network FC3	3-nn	1-nn	network LC5
region densities	70.7%	71.3%	83.5%	83.3%	-
cavities and holes	62.3%	69.8%	65.8%	66.3%	-
left, right profiles	85.5%	87.1%	86.1%	86.0%	90.1%
left, right, oriented profiles	86.0%	90.0%	86.5%	86.4%	92.3%
all profiles + regions + cavities	87.4%	90.1%	88.2%	88.5%	**93.6**%

To focus on the interest of the various features, some experiments with subsets of characteristics have been performed and are given in the previous chart. The networks called FC are full-connected networks

with 2 or 3 layers. For FC3, the results correspond to the optimal number of hidden neurons.
It is interesting to note that, in the case of single use of statistical features, the best method is the nearest-neighbor one owing to the nature of these features: each one provides significant information independently of the others, whereas the global shape is more meaningful for a profile curve. Despite their small number, the morphological features provide correct results. However, they are not so robust: for instance, a non well closed zero provides various cavities instead of a meaningful hole. So, we couldn't build a system based only on morphological features.
We also note that, by adding the oriented profile to the left and right profiles, we add 2.2% on the performances of LC5. The generalization rate finally increases to 93.6% by using the whole set of features.

5 Conclusion

We have proposed an intermediate approach between classical handwritten digit recognition methods, and pure neural methods. Compared with classical methods, this approach has the advantage to avoid the difficult task of selecting a right parameter set and decision rules. Compared with pure neural methods, it has the advantage to have a lower complexity. For instance, our network contains only 127 neurons, and 2043 connections, compared with 3850 neurons and 98442 connections in [3].
We have compared the performances of neural networks with k-nearest neighbors classifiers. The results show that the performances of the neural networks are better, and that they need less run-time. Local connections and shared weights are also a good point for better generalization.

References

[1] J.P. ASSELIN de BEAUVILLE et al., "Reconnaissance de chiffres manuscrits saisis sans contrainte" BIGRE n^o68 (Reconnaissance automatique de l'écrit), Le Havre - France, may 18th, 1990, pp 89-101

[2] V.K. GOVINDAN, A.P. SHIVAPRASAD, "Character Recognition - A Review", Pattern Recognition, vol 23, n^o7, pp 671-683, 1990

[3] Y. LE CUN et al., "Optical Character Recognition and Neural-Net Chips", International Conference on Neural Networks, Paris, july 9-13th, 1990, pp 651-655

[4] Brian T. MITCHELL, Andrew M. GILLIES, "A model based computer vision system for recognizing handwritten ZIP codes", Machine Vision and Applications, vol 2, n^o4, Fall 1989, pp 231-243

[5] D.E. RUMELHART, G.E. HINTON, R.J. WILLIAMS, "Learning internal representations by error backpropagation", Parallel Distributed Processing, D.E. RUMELHART and J.L. Mc CLELLAND, Chap8, Bradford book - MIT Press - 1986

[6] M. SHRIDHAR, A. BADRELDIN, "Recognition of isolated and simply connected handwritten numerals", Pattern Recognition, vol 19, n^o1, pp 1-12, 1986

[7] M. SHRIDHAR, A. BADRELDIN, "Context-directed segmentation algorithm for handwritten numeral strings", Image and Vision Computing, vol 5, n^o1, February 1987

[8] L. STRINGA, "Efficient classification of totally unconstrained handwritten numerals, with a trainable neural network", Pattern Recognition Letters 10, pp 273-280, October 1989

[9] Charles C. TAPPERT, Ching Y. SUEN, Toru WAKAHARA, "The state of the art in On-Line handwriting recognition", IEEE-PAMI, vol 12, n^o8, August 1990