Enhancement and Cleaning of Handwritten Data by Using Neural Networks

José Luis Hidalgo[1], Salvador España[1], María José Castro[1], and José Alberto Pérez[2]

[1] Departamento de Sistemas Informáticos y Computación
Universidad Politécnica de Valencia, Valencia, Spain
{jhidalgo,mcastro,sespana}@dsic.upv.es
[2] Departamento de Informática de Sistemas y Computadores
Universidad Politécnica de Valencia, Valencia, Spain
aperez@disca.upv.es

Abstract. In this work, artificial neural networks are used to clean and enhance scanned images for a handwritten recognition task. Multilayer perceptrons are trained in a supervised way using a set of simulated noisy images together with the corresponding clean images for the desired output. The neural network acquires the function of a desired enhancing method. The performance of this method has been evaluated for both noisy artificial and natural images. Objective and subjective methods of evaluation have shown a superior performance of the proposed method over other conventional enhancing and cleaning filters.

Keywords: handwritten recognition, form processing, image enhancement, image denoising, artificial neural networks

1 Introduction

The field of offline handwritten recognition has been a topic of intensive research for many years [1–4]. One of the first steps in the classical architecture of a handwritten text recognizer is preprocessing, where noise reduction and normalization takes place. Preparing clean and clear images for the recognition engines is often taken for granted as a trivial task that requires little attention. However, this step undoubtedly influences the overall performance of the system. Neural networks for cleaning and enhancing scanned handwritten images are proposed in this work. For a review of image processing with neural networks, see [5].

There exist several methods to design forms with fields to be filled in. For instance, fields may be surrounded by bounding boxes, by light rectangles or by guiding rulers. These methods specify where to write and, therefore, minimize the effect of skew and overlapping with other parts of the form. These guides can be located on a separate sheet of paper that is located below the form or they can be printed directly on the form. The use of guides on a separate sheet is much better from the point of view of the quality of the scanned image, but requires giving more instructions and, more importantly, restricts its use to tasks where this type of acquisition is used. Guiding rulers printed on the form are more commonly used for this reason. Light rectangles can be removed more easily with filters than dark lines whenever the handwritten text touches the rulers. Nevertheless, other practical issues must be taken into account:

J.S. Marques et al. (Eds.): IbPRIA 2005, LNCS 3522, pp. 376–383, 2005.

- The best way to print these light rectangles is in a different color (i.e. light yellow); however, this approach is more expensive than printing gray rectangles with black-and-white laser printers.
- A more economical and easier approach is to use gray rectangles printed by a black-and-white laser printer. This produces a pattern of pixels that is more difficult to remove.
- Very different types of handwriting instruments and different colors are used by different users.

The work described here consists of filtering the background noise caused mainly by gray rectangles used as guiding rulers. The proper elimination of these rectangles makes it possible to use this approach in the design of forms to be used by handwritten recognition systems, which is much cheaper than other approaches.

In many handwritten recognition systems, preprocessing does not require a binarization step. For this reason, the images should be maintained in gray-level quality. The enhancement of images should also correct traces with low, non uniform ink level produced by some handwriting instruments (such as some ball pens and pencils), which may be broken or disappear in the preprocessing.

2 The Spartacus Database

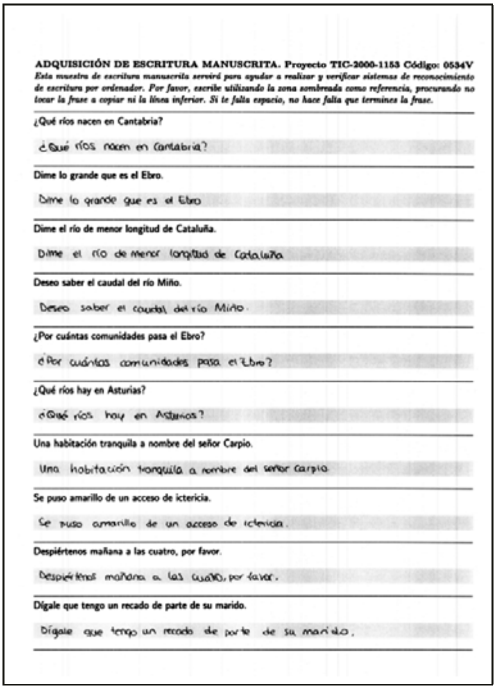
A new offline handwritten database for the Spanish language, which contains full Spanish sentences, has recently been developed: the Spartacus database [6] (which stands for *SPAnish Restricted-domain TAsk of CUrsive Script*). There were two main reasons for creating this corpus. First of all, most databases [7–12] do not contain Spanish sentences, even though Spanish is a widespread major language. Another important reason was to create a corpus from semantic-restricted tasks. These tasks are commonly used in practice and allow the use of linguistic knowledge beyond the lexicon level in the recognition process. The database includes 1 500 forms produced by the same number of writers, scanned at 300 dpi. A total of around 100 000 word instances out of a vocabulary of around 3 300 words occur in the collection.

As the Spartacus database consisted mainly of short sentences and did not contain long paragraphs, the writers were asked to copy a set of sentences in fixed places: dedicated one-line fields in the forms. Figure 1 shows one of the forms used in the acquisition process. These forms also contain a brief set of instructions given to the writer.

3 Cleaning and Enhancing Method

There are several classic spatial filters for reducing or eliminating high-frequency noise from images. The mean filter, the median filter and the closing/opening filter are frequently used [13]. The mean filter is a low-pass or smoothing filter that replaces the pixel values with the neighborhood mean. It reduces the image noise but blurs the image edges. The median filter calculates the median of the pixel neighborhood for each pixel, thereby reducing the blurring effect. Finally, the opening/closing filter is a mathematical morphological filter that combines the same number of erosion and dilation morphological operations in order to eliminate small objects from images [14, 15].

Squared neighborhoods of 3×3 pixels with center at the modified pixel were employed in the filter implementations. However, the obtained images were not satisfactory enough (see Figure 2). For this reason, neural network filters [5] were used. Neural networks were used to estimate the gray level of one pixel at a time. The input to the network consists of a square of pixels centered at the pixel to be cleaned (see Figure 3).



ADQUISICIÓN DE ESCRITURA MANUSCRITA. Proyecto TIC-2000-1153 Código: 0534V
Esta muestra de escritura manuscrita servirá para ayudar a realizar y verificar sistemas de reconocimiento de escritura por ordenador. Por favor, escribe utilizando la zona sombreada como referencia, procurando no tocar la frase a copiar ni la línea inferior. Si te falta espacio, no hace falta que termines la frase.

¿Qué ríos nacen en Cantabria?

¿Qué ríos nacen en Cantabria?

Dime lo grande que es el Ebro.

Dime lo grande que es el Ebro

Dime el río de menor longitud de Cataluña.

Dime el río de menor longitud de Cataluña

Deseo saber el caudal del río Miño.

Deseo saber el caudal del río Miño.

¿Por cuántas comunidades pasa el Ebro?

¿Por cuántas comunidades pasa el Ebro?

¿Qué ríos hay en Asturias?

¿Qué ríos hay en Asturias?

Una habitación tranquila a nombre del señor Carpio.

Una habitación tranquila a nombre del señor Carpio

Se puso amarillo de un acceso de ictericia.

Se puso amarillo de un acceso de ictericia.

Despiértenos mañana a las cuatro, por favor.

Despiértenos mañana a las cuatro, por favor.

Dígale que tengo un recado de parte de su marido.

Dígale que tengo un recado de parte de su marido.

"This handwritten sample is intended to help the experimentation and testing of computer handwriting recognition. Please, write using the guiding rectangle as reference, trying not to touch the typographic text nor the bottom horizontal rule. If there is not enough space, the sentence should be left unfinished."

Fig. 1. An example of a filled acquisition form and the translation of the instructions given for filling out the form.

4 Simulated Noisy Image Dataset

The main goal was to train a neural network in a supervised manner to obtain a clean image from a noisy one. In this particular case, it was much easier to obtain a simulated noisy image from a clean one than to clean a subset of noisy images.

The clean image database was obtained by scanning 150 white background handwritten sentences. The handwriting instrument was specially chosen in order to obtain uniform traces. The resolution was set to 300 dpi, which gives $32 \cdot 10^6$ patterns. Pixels are codified as gray-levels in the interval [0,1], where 0 means "black" and 1 means "white".

The process for obtaining simulated noisy images follows the scheme presented in Figure 4. This process requires images of the background (gray rectangles) of the acquisition forms, which were obtained by printing and scanning the same background

Fig. 2. An example of an original scanned image (a) and the clean images obtained with the filters: Mean filter (b), Median filter (c), and Opening/Closing filter (d).

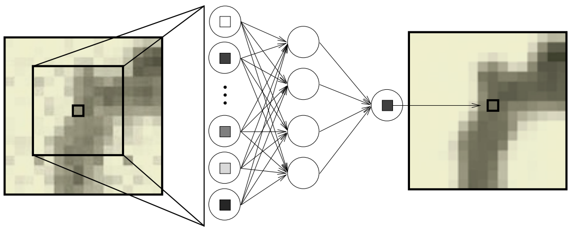


Fig. 3. Architecture of the artificial neural network to enhance images. The entire image is cleaned by scanning with the neural network.

of the original forms. First, in order to simulate the variability of the traces produced by some handwriting instruments (pencils, some ball pens, etc.), a trace noise was obtained by generating a white noise and applying an "oil" effect [16]. This trace noise was applied to the clean-trace image using the maximum operation, which only affects the ink and not the white background. Secondly, the noisy-trace image was combined with the scanned background noise to obtain the simulated noisy image. An example of a simulated noisy image is shown in Figure 5.

5 Enhancement and Cleaning with Neural Networks

5.1 Architecture

Multilayer perceptrons (MLPs) were used for the enhancement and cleaning of images. Only one output unit was needed to estimate the energy level (gray level) of the clean pixel. The activation function of the units of the hidden layer(s) was the sigmoid function, while the activation function of the output unit was the identity function. Due to the linear activation function, the output may be out of range, but, in practice, values were in the interval [0,1].

We employed the identity function at the output layer instead of the more commonly used sigmoid function because the characteristics of an MLP were improved significantly with the identity function when applied to regression problems such as image processing (see, for example, [17]). It should be noted that using a sigmoid activation function at the output layer is useful for applications where the output is in the form of binary values such as binarization image processing.

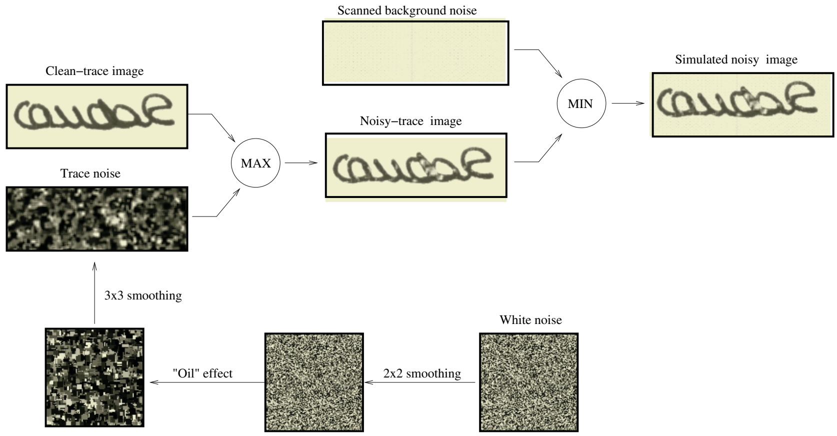


Fig. 4. Simulated noisy process.

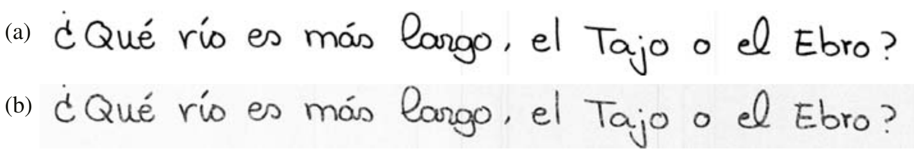


Fig. 5. (a) Clean-trace image, and (b) Simulated noisy image.

The input units consisted of a squared window of pixels centered at the pixel to be cleaned. Neighborhoods from 2 to 5 were tested, where a neighborhood of n pixels means a squared $(2n+1)$-sided input window to the MLP.

The entire image was cleaned by scanning all the pixels with the MLP. The MLP, therefore, functions like a nonlinear convolution kernel. The universal approximation property of a MLP guarantees the capability of the neural network to approximate any continuous mapping [18].

5.2 Training the Neural Networks

The obtained simulated noisy image corpus was divided into a training set, a validation set and a test set. The trained neural networks differed in the number of neighbor pixels (from 2 to 5), the number of hidden layers (one or two hidden layers) and the number of hidden neurons in each layer (from 2 to 16 hidden units). In every case, the online version of the backpropagation learning algorithm with momentum was used. For the same topology, several trainings were performed varying the learning rate, the momentum term and using different initializations of the weights. The stopping criteria was the mean squared error in the validation set.

6 Evaluation of the Cleaning and Enhancing Method

The proposed approach was objectively evaluated by using the simulated noisy image dataset. We measured the “closeness” of the original image (clean) and the cleaned image (the simulated "noisy" image after being cleaned by each of the MLPs). This measure was obtained by calculating the mean squared error (MSE) between each pair of images in the test set. Figure 6 plots the MSE of all the trained MLPs. As can be observed, the best results were achieved with many different MLPs, demonstrating the robustness of the methodology. The best MLP (the one that obtained the lowest MSE in test set) used 5 neighbors at the input and two hidden layers of 16 and 8 units, respectively.

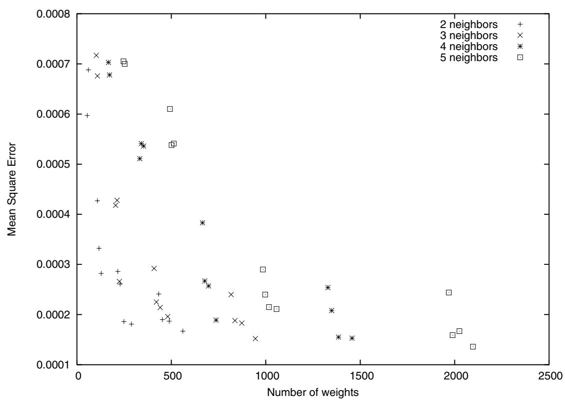


Fig. 6. Mean Squared Error of the test set for the trained MLPs. The number of neighbors and the complexity (number of weights) of the MLPs are displayed.

In order to perform a subjective evaluation of the cleaned Spartacus database, we visually inspected a subset of the cleaned images. An example of the performance of the proposed neural method, along with the result of the best used filter (opening/closing filter), is shown in Figure 7. As can be seen from the examples, the result clearly improves the image quality.

7 Summary and Conclusions

In this paper, we have described a generic cleaning and enhancing system for automatic form processing using neural networks. It takes clean and simulated noisy images to train and select the best neural network. Subjective and objective evaluations of the cleaning method show excellent results to clean forms with printed gray-areas to indicate where to fill in the information. The same idea could be used to clean and restore other types of images, such as noisy backgrounds in scanned documents, folded documents, stained paper of historical documents, vehicle license recognition, etc.

The proposed approach should also be evaluated objectively in a goal-directed manner [19], which means testing an image recognition system based on the results of our enhancing and cleaning method. We are planning to use both a standard HMM-based recognition system that has been developed in our research group and a commercial product. The purpose of using more than one recognizer in the evaluation is to prove that the improvement of performance brought about by the cleaning and enhancing procedure is independent of the features or methods that are used in the recognizers.

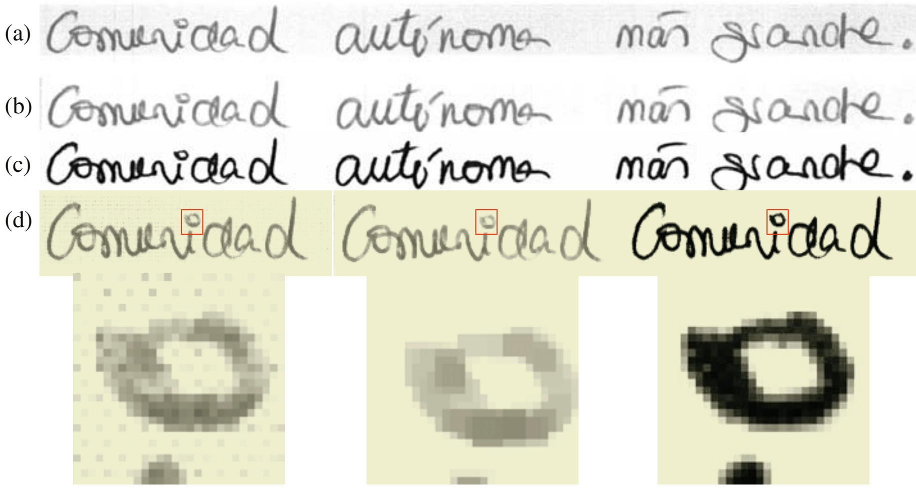


Fig. 7. (a) Original image, (b) result of applying the opening/closing filter, (c) result of applying the best MLP, and (d) Detail of the former images: Original image (left), result of applying the opening/closing filter (middle), and the result of applying the best MLP (right).

Acknowledgments

Thanks to the Generalitat Valenciana under contract 20040479 for funding.

References

1. Plamondon, R., Srihari, S.N.: On-line and off-line handwriting recognition: A comprehensive survey. IEEE Trans. on PAMI **22** (2000) 63–84
2. Bozinovic, R.M., Srihari, S.N.: Off-Line Cursive Script Word Recognition. IEEE Trans. on PAMI **11** (1989) 68–83
3. Bunke, H.: Recognition of Cursive Roman Handwriting – Past, Present and Future. In: 7th International Conference on Document Analysis and Recognition (ICDAR 2003), Edinburgh, Scotland (2003)
4. Toselli, A.H., et al.: Integrated Handwriting Recognition and Interpretation using Finite-State Models. International Journal of Pattern Recognition and Artificial Intelligence **18** (2004) 519–539
5. Egmont-Petersen, M., de Ridder, D., Handels, H.: Image processing with neural networks – a review. Pattern Recognition **35** (2002) 2279–2301
6. España, S., Castro, M.J., Hidalgo, J.L.: The SPARTACUS-Database: a Spanish Sentence Database for Offline Handwriting Recognition. In: IV International conference on Language Resources and Evaluation (LREC 04), Lisbon, Portugal (2004) 227–230
7. Guyon, I., Haralick, R.M., Hull, J.J., Philips, I.T.: Data Sets for OCR and Document Image Understanding Research. In Bunke, H., Wang, P.S.P., eds.: Handbook of Character Recognition and Document Image Analysis. World Scientific, Singapore (1997) 779–799
8. Hull, J.J.: A database for handwritten text recognition research. IEEE Trans. on PAMI **16** (1994) 550–554

9. Marti, U.V., Bunke, H.: The IAM-database: an English sentence adatabase for offline handwriting recognition. International Journal on Document Analysis and Recognition **5** (2002) 39–46
10. Suen, C.Y., et al.: Computer recognition of unconstrained handwritten numerals. Special Issue of Proc IEEE **7** (1992) 1162–1180
11. Viard-Gaudin, C., et al.: The IRESTE On/Off (IRONOFF) Dual Handwriting Database. In: Fifth International Conference on Document Analysis and Recognition, Bangalore, India (1999) 455–458
12. Wilkinson, R., et al.: The first census optical character recognition systems conference. In: #NISTIR 4912, The U.S. Bureau of Census and the National Institute of Standards and Technology, Gaithersburg, MD (1992)
13. Gonzalez, R., Woods, R.: Digital Image Processing. Addison-Wesley Publishing Company (1993)
14. Serra, J.: Image Analysis and Mathematical Morphology. Academic Press. New York (1982)
15. Serra, J.: Image Analysis and Mathematical Morphology. Vol: 2. Academic Press. New York (1988)
16. Holzmann, G.: Beyond Photography. Prentice Hall Professional Technical Reference (1988)
17. Suzuki, K., Horiba, I., Sugie, N.: Neural Edge Enhancer for Supervised Edge Enhancement from Noisy Images. IEEE Trans. on PAMI **25** (2003) 1582–1596
18. Bishop, C.M.: Neural Networks for Pattern Recognition. Oxford University Press (1995)
19. Øivind Due Trier, Jain, A.K.: Goal-Directed Evaluation of Binarization Methods. IEEE Trans. on PAMI **17** (1995) 1191–1201