A neural network hybrid model for an optical braille recognitor

GIOVANNA MORGAVI, MAURO MORANDO
Institute for Electronic Circuits
National Research Council
Via De Marini 6, 16149 Genova
ITALY

Abstract: - Braille is the most widely used system for written communication using tactile means. Automation of the Braille reading process will effectively enable the transcription and duplication of existing documents as well as their preservation. By converting a Braille document into electronic form, all the benefits of developments for the electronic form of printed documents can be enjoyed: easier transmission, efficient storage and manipulation etc.. The proposed procedures use approaches commonly applied in standard OCR systems. It uses a commercially available flatbed scanner to acquire a grayscale image of a Braille document from which the characters are recognized. This prototype system has been developed by using methods from image processing and analysis to AI. These methods achieve high speed and accuracy by using hybrid neural network approaches. The system dynamically adapts to factors such as the quality of the input pattern, its intrinsic similarities and differences between patterns.

Key-Words: - hybrid model, neural network application, character recognition

1 Introduction

Braille is a tactile code that is used by unsighted people instead of print as a reading and writing medium. Each character or "cell" is made up of 6 dots, and there are several different codes that are in general use. Sighted people, who are not experienced tactile Braille readers, are able to recognize the Braille coded text from its visual appearance. Recognition is slow and tiring, but possible. Although the Braille dots have the same colors as the background, they cast soft shadows when scanned with standard flatbed scanner. These shadows are used to locate the dots on the page. The process of recognition is based on image processing, artificial intelligence and computer vision methods. The recognition engine is able to process both single sided and double-sided documents with various paper color. Scanned image of a Braille sheet consists of a "grid" of dot patterns. Image of a dot formed towards the scanner window differs from the image of dot formed backwards. Thus, the picture of embossed Braille sheet captured by a standard flatbed scanner contains information about the Braille coded text.

The whole optical Braille recognition process can be divided into several steps: dot matrix location, Braille character identification and image to Braille code translation. Braille dot geometry objective measurement shows a complex nature: the dot is a small, three-dimensional damageable and delicate object. Reliability of classic mechanical measurement methods is very limited due to easy deformations of cellulose fibers forming the dot by the measuring device touch. Readability of Braille after repeated reading as well as its readability after its storage with some damaging force applied to it is often demanded. The standard usual solution is the usage of thicker Braille paper with higher grammage – however that can be an expensive solution: different embossing technologies and different Braille paper materials result in very different dot "stability". Braille embossed on single sheets brings a lot of translation advantages, but, quite often, their cost are too high. The "double feeds" give a lot of more difficulties in automatic translation processing. Dot overlapping are often sources of trouble. Recognition of the dot matrix is the core of the optical Braille recognition system. Based upon advanced methods of mathematics, statistics and computer vision we built a hybrid intelligent model.

2 The hybrid intelligent model

In recent years there has been an explosive growth in the successful use of hybrid intelligent systems in many diverse areas. The main contributing factor for the development of hybrid systems has been the increased use of neural networks for pattern recognition, classification and optimization tasks [1]. The ability of neural networks to perform tasks that would otherwise prove difficult or intractable to symbolic computing systems is now recognized and they are often used as modules within intelligent hybrid systems. The activity of module integration has initiated the field of hybrid system development, which is one of the most promising research areas for building intelligent systems.

Generally the reasons given for coupling neural and symbolic components involve such issues as reducing the brittleness of the rule-based component by incorporating the robustness of a neural network. Most expert systems experience a sharp drop of operational capability when confronted with novel situations for which no specific rules exist. Other reasons include the lack of rule-based adaptability to changing external conditions that can only be addressed through manual modification of the knowledge base. However, by fusing the two techniques together in tighter configurations a number of interesting opportunities arise for increasing the power of neural networks. These new opportunities come at the cost of increased operational complexity and computational overheads but empower neural systems with symbolic processing abilities and vice versa. Our Optical Braille Character Recognition (OBCR) scheme integrates separate symbolic and neural elements. The main feature of our sequential configuration is the serial processing of data as it is passed from one module to the next. One module acts as a pre-processor of data. A rule-based module pre-processes data for a neural network by identifying the relevant parameters for the appropriate input units. The architecture of our OBCR system represents the simplest of the modular hybrid architectures. A great deal of pre-processing must be carried out since the image has a high dimensionality and only a small fraction can be used to train the neural networks. A number of different neural network models were trained and tested on data generated by a different scanner and scanning. The system performed well using multi-layer perceptrons.

3 OBCR basic principles

The main principles used in the optical Braille recognition rules based system are the following:

- The Braille dot grid is arranged in a "regular grid".
- The Braille dots embossed on the same side of paper perform pictures similar each other.
- The image parameters are distinguished by their statistical distributions.
- The image to Braille copy function is implemented namely for the cases, where the ASCII format decoding is not required or has no principal sense – like in the Braille music notation copying.

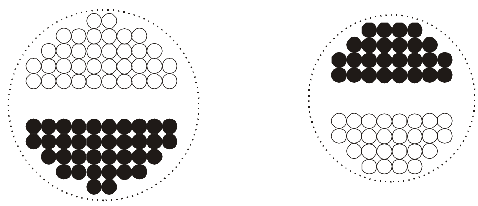
Although the Braille dots have the same color as the background, they cast soft shadows when scanned with a standard flatbed scanner. These shadows are used to locate the dots on the page. A Braille text is recognizable through optical scanning thanks to the acquisition technique of scanners, which, by inclined light, illuminate the sheet in question. Since each punching produces shadows when they get illuminated, it is possible to discriminate punching which do not have any different color compared to the color of the remaining paper.

Most practical pattern recognition systems are composed of multiple modules. There is a field locator that extracts regions of interest, a field segmenters that cuts the input image into images of candidate characters, a recognizer which classifies and scores each candidate character. Typically each module is optimized for the special purpose and sometimes it is trained outside its contest.

So many Braille pages are larger than the scanning area of most scanners, the OBCR incorporates a powerful method to overcome this problem. The Braille page can be scanned per parts and the OBCR will automatically find the correct matching of these parts and will merge them together to form the whole page.

Input files were standardized as tiff standard files. Tiff format has been chosen for many reasons. It may use only lossless compression. It allows the access to all the scanning bits. It is well distributed and flexible. Its usage frees this procedure from the used scanner. The only requirement for the ordinary flatbed scanner is to gives back light and shadow of scanned dots Braille documents. Undergoing scanning procedures are often printed double-faced. The two prints are displaced in such a way that punching of both faces will not interfere between each other for tactile sensor. Unlucky, also inverted punching leave an image during optical scanning though leaving an image that is nearly the negative of the precedent one. Punching of this kind are cavities rather than prominences. Moreover, such an image has slight lower dimensions compared to that of positive dots. When dots are close to the light shadow image can be confused and its translation in character can give some trouble. Each scanner has its proper standard procedures and its range function values. It is necessary to proceed with a normalization step. By this operation you pass from a matrix with fixed dots identified through a gray level, that may account for any number between 0 and 255, corresponding to a shade of gray ranging from black to white, to a matrix with normalized dots, which may assume only one of the possible three values: black, white or paper.

Within this paper, "dot" will mean the "result of the optical scanning through punching"



(a) (b)

Fig.1 - Ideal normalized dots: (a) direct punching, (b) inverted punching

Talking about reality, the scanning process suffers from imperfections, which produce dots with much more irregular forms, both lacking of white and black dot part and provided with further dots outside its ideal outline

3.1 Field location

Most practical pattern recognition systems are composed of multiple modules. There is a field locator that extracts regions of interest, a field segmenters that cuts the input image into images of candidate characters, a recognizer, which classifies and scores each candidate character. Typically each module is optimized for the special purpose and sometimes it is trained outside its contest. Analysis of various Braille documents showed irregularities within character printing, that brings to non-constant distances (both vertical and horizontal) among characters in the sheet. This means that, as well as differences among values, which we can acknowledge in two different documents, we have to take into account the possibility that vertical and horizontal distances among characters in the same document vary between one line and the following one or among the columns. Punching irregularities related to heavy paper feed makes the Braille character recognition similar to the hand written black character processing. The first important problem to be solved is the location of a field.

The creation of the location of a field implies the construction of an optimal grid to locate characters. This operation can be done through several steps like trusty dot research, tentative grid creation, mean distances optimization, grid-positioning optimization

Cross-correlation function has been used to identify the single dots within the normalized matrix and to discriminate the positive from the negative dots.

The identification of a 'possible' dot will be carried out by computing the cross-correlation of the rectangular matrix that might contain the 'possible' dot with the matrix containing the ideal dot. If the function gets back to a value sufficiently close to 1, then the scanning position is assumed as a dot position from which the matrix has been extracted.

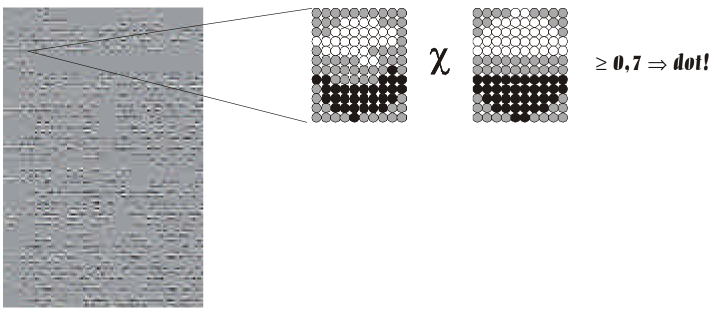


Fig.2 - Cross correlation process

The first operation to carry out on the normalized matrix is the identification of "trusty dots". Trusty dots are dots yielding a high value of cross-correlation, (greater then 0,85). During the first elaboration step, these dots are considered reference dots. In this step the procedure will find geometrical parameters of the grid to be used for the final identification of all Braille characters punched on the sheet.

We recorded obtained trusty dots in a matrix structure, in which all dots found in a strip are inserted into one line; when the horizontal matrix reaches the right border of the sheet, a new line of the matrix containing the found trusty dots is created. Indeed, at the end of the process, the information about horizontal distribution of trusty dots is available: each line of the matrix will contain a certain (variable) number of trusty dots, each one with its “homogeneous” vertical co-ordinate (usually, it will not be equal due to imprecision of punching, of processing digitalization and of positioning of the sheet on the scanner). The trusty dots research is performed both horizontally and vertically generating two position matrixes to allow the identification of mean distances of dots.

The analysis of many Braille documents showed irregularities on dots distribution both regarding distance among dots and distance among characters. Therefore, it is not possible to define a priori the four values: horizontal and vertical distance among dots of a character, horizontal and vertical distance among characters

The final activity of identification of dots and, as a consequence, of Braille characters requires a precise positioning for research points of characters, that is the definition of a grid providing the areas in which dots are expected, areas that will be the only ones taken into examination through the mechanism of discrimination of dots, since the presence of negative dots printed on the opposite side of the sheet makes the identification of positive dots, that are contiguous to negative dots, problematic. Determination of the best research grid is carried out by exploiting once again trusty dots and placing an artificial sheet with ideal dots in position where trusty dots have been found. For each grid positioning an alignment value of the grid is computed and at the end of the test session the research grid with the best alignment value is chosen.

3.2 Braille Dot detection through neural network

The Braille characters are translated as a series of symbols ranging between 0 and 63.

Analysis of various Braille documents has found out irregularities within character printing, that brings to non-constant distances (both vertical and horizontal) among characters in the sheet. This means that, as well as differences among values, which we can acknowledge in two different documents, we have to take into account the possibility that vertical and horizontal distances among characters in the same document vary between one line and the following one or among the columns. This makes the use of a grid with constant step less reliable and, therefore, introduces the necessity to carry out a continuous identification activity of the document geometries in the character recognition phase too.

The pre-processing module reduced the dimensionality of the raw input data by selecting the most important parameters, which are easily calculated using heuristics. The transformed data is, at this step, passed onto the neural network module, which is designed to detect the dot presence. A neural network is required for this task since several dots exhibit the activation of different pixel. The output of the neural network is simply a Boolean value. A number of different neural network models were trained and tested on data generated by a different scanner and scanning. The system performed well using multi-layer perceptrons. It has been implemented with the Back Propagation algorithm modified to reduce the convergence time. We scanned 100 sheets both sides and we used half a sheet as training set and half a sheet as test set. Then a Multi-Layer Perceptron was trained by epoch. The Vogl (Vogl et al., 1988) acceleration algorithm has been introduced.

The problem of the definition of the architecture of the MLP is hardly discussed in literature. It is well known that a feed forward network with two layers is sufficient to store any number of patterns if a sufficient number of hidden neuron is used (Hertz et al. 1991) .The number of neurons in the hidden layer is a concern in the application of neural networks to signal classification. A rule of thumb (Baum et al., 1989), known as the Baum-Haussler rule, has been used.

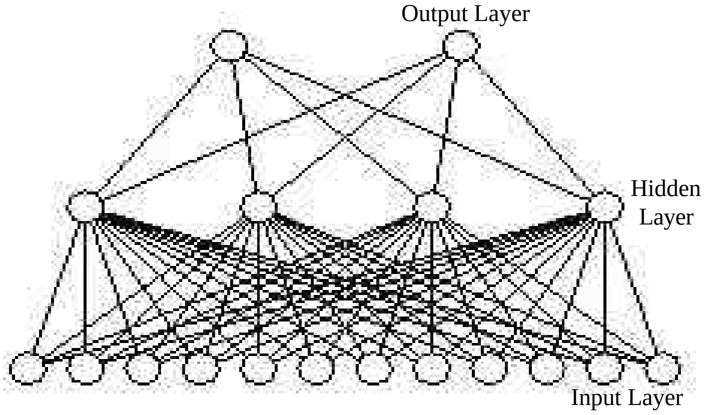


Fig.3 - Multilayer Perceptron Neural Network

This rule generally ensures that neural networks generalize, rather than memorize. We tested many values for hidden neuron. Final results are satisfactory.

4 Results

In fig.4 one output window produced by the Braille recognizer is shown. The blue boxes indicate the position where the grid suggests the possible existence of a Braille character. The green circles indicate the location of trusty dots. The red rings indicate recognized dots. The user can check the interpretation correctness and can interact with this picture with his mouse. He can switch on or off any dot within the grid system: in such a way if a wrong dot was assumed or an existing dot wasn't recognized the user can correct it, even if he doesn't know anything about Braille characters. Depending on the goodness of the scanning tiff file the OBCR reach in mean about the 98% of automatic correctness. By using this feature the 100% can be easily reached.

Fig.4 - Recognition step: output window for correction

4.1 Result Discussion

What metrics are good for evaluating OBCR results? In this section we describe a few metrics that we consider important and give advantages and disadvantages of using them.

Let A represent the number of characters in the original sheet, B the number of dots in the original sheet, C the number of characters in the OBCR-generated text, D the number of deleted dots, E the number of inserted dots.

Accuracy: The number of characters correctly recognized on a page normalized by the total number of characters in the original sheet. Thus accuracy is C/A:

Precision: This is the number of dots correctly recognized on a page normalized by the number of dots in the OCR-generated text. Thus precision is (B-(D+E))/B

Insertion: The number of dots inserted normalized by the number of total dots on the original page: Thus insertion is E/B

Deletion: This is the number of dots deleted normalized by the number of total dots on the original page: Thus deletion is D/B

Substitution: The number of characters substituted normalized by the number of characters on the original page. Thus substitution is (A-C)/A.

We scanned images relatively clean from books, magazines and computer documents. The data set we used for test consists of 300 generated tiff images. The original sheets were chosen with different characteristics respect to the level of overlapping between front and back cover, the darkness of the original paper, the usage and consumption level of the sheets.

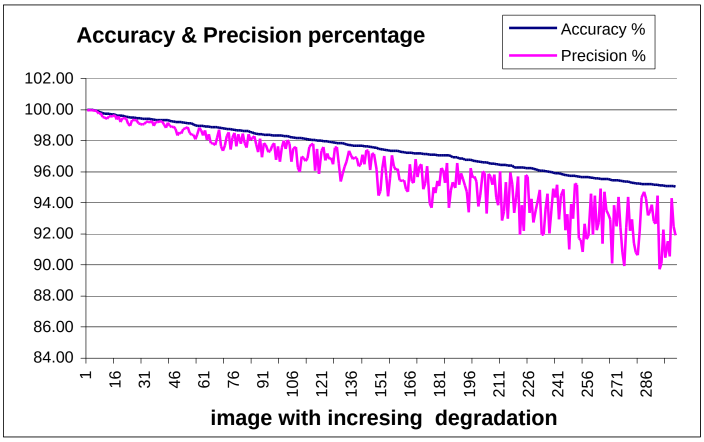


Fig.5 - Accuracy and precision distribution percentage

In fig.5 the accuracy distribution percentage and precision distribution percentage are plotted as function of the increasing paper degradation. The input papers on X-axis are ordered following the precision percentage.

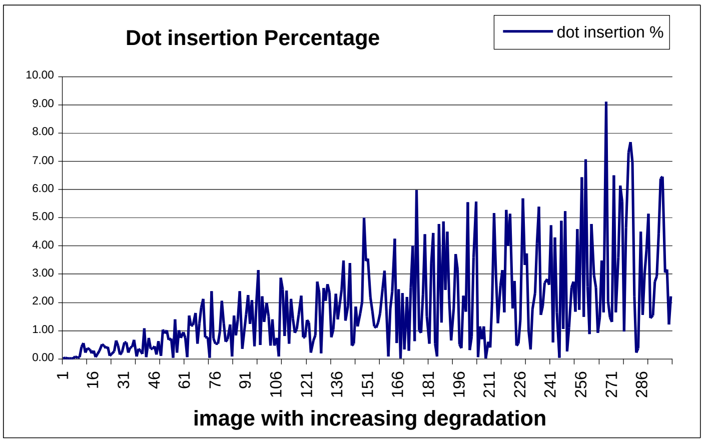


Fig.6 - Wrong dot insertion error percentage

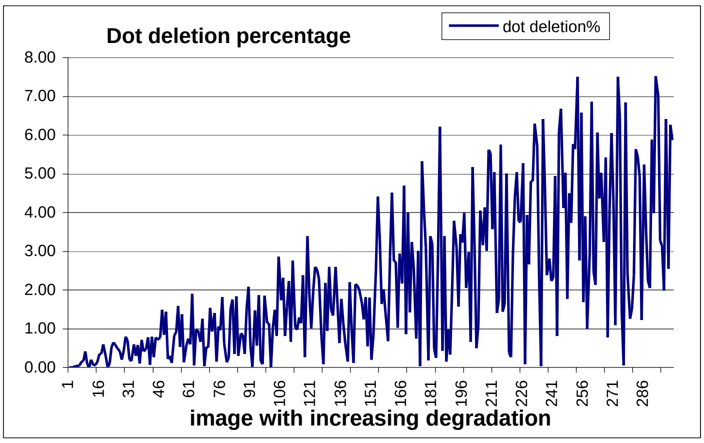


Fig.7 - Wrong dot deletion error percentage

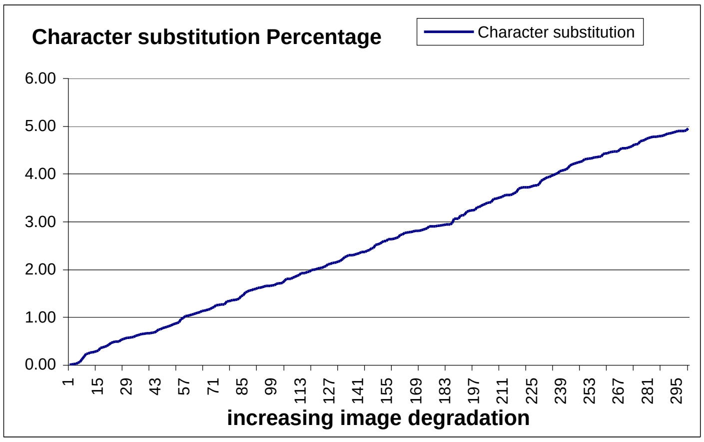


Fig.8 - Character substitution percentage

In fig.6, 7, 8 the wrong dot insertion distribution percentage, the wrong dot deletion distribution percentage and the character substitution are plotted respectively as function of the increasing paper degradation.

	Average %	Stand. Dev.
Accuracy	97.46	1.46
Precision	96.04	2.49
Inserted dots	1.84	1.71
Deleted dots	2.12	1.92
Substituted Char.	2.54	1.45

Table 1 - Automatic OBCR Average performances

In Table 1 the average percentage performances of the Automatic OBCR on the data set we tested is shown. The error percentage is very small and the performances can be improved through the interactive step.

5 Conclusion

The Prototype product is efficacy and user friendly. Results in Braille Character Recognition look good for this application field. The possibility of interactive check and correction is very useful especially when input images are degraded.

The performance analysis is adequate to literature on OCR comparison. Even if it is very difficult to quantify the level of degradation of input paper: unluckily, in this field there are not such reference standard databases like in the black character recognition.

This work has been carried on within the European Community Project IST-11908, PLAY2 (software for storage of various music input, conversion of different Braille syntax into sound, graphic, Braille output and telematics access).

References:

[1] D. E. Rumelhart, G. E. Hinton and R. J. Williams, (1986) "Learning Representations by Backpropagating Errors", Nature, vol. 32), pp. 533-536.

[2] Vogl, T.P., J.K. Mangis, A.K. Rigler, W.K. Zink, and D.L. Alkon, (1988) Accelerating the convergence of back-propagation method, Biol. Cybern., 59, 257-263.

[3] Hertz, J., A. Krogh, R. G. Palmer, (1991) “Introduction to the theory of neural computation”, Lecture notes Vol. 1, Santa Fe Institute Studies in the sciences of complexity, Addison-Wesley, Redwood City, CA 94065.

[4] Baum, E. Haussler, D. (1989). “What size net gives valid generalization?” Neural Computation, 1, 151-160