

A Fuzzy Neural Network Approach for Document Region Classification Using Human Visual Perception Features

Red Neuro-Difusa para la Clasificación de Regiones de Documentos Utilizando Características de la Percepción Visual Humana

Mario Ignacio Chacón Murguía

Instituto Tecnológico de Chihuahua, Av. Tecnológico 2909

Chihuahua, Chih., México C.P. 31310

E-mail: mchacon@itcchihuahua.edu.mx

Article received on June 07, 2001; accepted on September 19, 2002

Abstract

This paper describes a fuzzy neural network classifier to perform document region classification using features obtained from human visual perception theories. The foundations of the classifier are derived from human visual perception theories. The theories analyzed are texture discrimination based on textures, and perceptual grouping. Based on these theories, the classification task is stated as a texture discrimination problem and is implemented as a preattentive process. Engineering techniques are then developed to extract features for deciding the class of information contained in the regions. The feature derived from the human visual perception theories is a measurement of periodicity of the blobs of the text regions.

This feature is used to design a fuzzy neural network classifier. The results of this work may be a good support for the assertion that visual perception theories may be incorporated into engineering techniques to produce better results.

Keywords: Document Analysis, Document Segmentation, Visual Perception, Image Segmentation, Image Classification, Fuzzy Neural Networks.

Resumen

Este artículo describe un clasificador neurodifuso para realizar la clasificación de regiones de documentos utilizando características obtenidas de la teoría de percepción visual humana. Los fundamentos del clasificador se derivan de la teoría de percepción visual humana. Las teorías analizadas son discriminación de textura basada en textones, y agrupamiento perceptual. Basada en esas teorías, la tarea de clasificación se establece como un problema de discriminación de texturas y se implementa como un proceso de pre-atención. Después se desarrollan teorías de ingeniería para extraer características para decidir el tipo de información contenida en las regiones. La característica derivada de las teorías de percepción visual humana es una medición de la periodicidad de los núcleos de las regiones de texto. Esta característica se utiliza para diseñar a clasificador neurodifuso. Los resultados de este trabajo se consideran como una buena justificación para afirmar que teorías de percepción visual humana pueden ser incorporadas dentro de las técnicas de ingeniería para producir mejores resultados.

Palabras clave: Procesamiento de Documentos, Análisis de Documentos, Redes Neurodifusas.

1 Introduction

The objective of document image analysis is to extract the information from a document in much the same way as a person would. This means, given a digital document image, perform a segmentation and classification of the blocks of the document (O'Gorman 1995). The document image analysis process has five phases; data capture, pixel-level processing, document segmentation (structural analysis), document classification, and document interpretation (functional layout analysis). These operations are all important because the result of each operation affects the total performance of the system.

However, segmentation can be considered as a major operation in this process. Under the domain of document image analysis, segmentation corresponds to divide the document into possible primitive subsections such as text and picture blocks. Once the document is divided, measurements of specific features are obtained from the blocks. These features are used to classify the blocks into text or pictures. The segmentation and classification process described above is related to structural layout analysis or physical and geometric analysis. Functional layout analysis deals with the classification of the blocks according to the function of the block in the document, like title, byline, abstract, etc., to make a semantic representation of the document. Figure 1 illustrates the document image analysis process, and the main stage addressed in this work is shown in gray. Figure 2 shows an example of structural and functional analysis of a document.

Correct segmentation and classification of blocks can have a much larger effect on the final performance of the system than the other operations at the pixel-level phase because these processes involve decisions concerning large amounts of information in the document. If a document is incorrectly segmented, correct classification may not be possible. If the system reports miss-classification or not classification, the document cannot be further processed, or the system will require human assistance to solve the problem. Human interaction is not desirable in a process

that is supposed to be automatic. Therefore, there is a specific need to develop effective and precise algorithms to segment and classify blocks. Classification of the document components into blocks of text and pictures increases the performance of the document analysis system since it allows application of appropriate processing to the different parts of the document.

Several methods for document segmentation and classification have been developed (Bloomberg, 1992; Etemad et al., 1994; Fletcher, and Kasturi 1998; Haralick et al., 1973; Jain, and Bhattacharjee, 1984). However, most of them are based only on an engineering approach. This approach makes the methods sensitive to many parameters related to the format of the document. The classification

method proposed here is based on human visual perception features that are classified by a fuzzy neural network (Bezdek and Pal, 1992; Boskovitz, and Guterman, 2002; Chung-Hoon F., and L. Changsu, 2001; Vaucher, and 2000). The main objective to use human visual perception theories is to incorporate characteristics of the human visual perception system to obtain a robust system.

The classifier was designed to work for free format document images to test its robustness and the process is over low resolution images, 40 and 50 dpi to prove the power of the method since a human is able to classify a document even if the document is not readable.

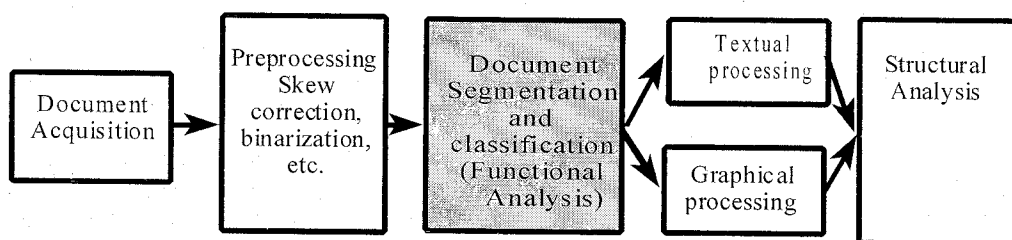


Figure 1: Document Image Analysis Process.

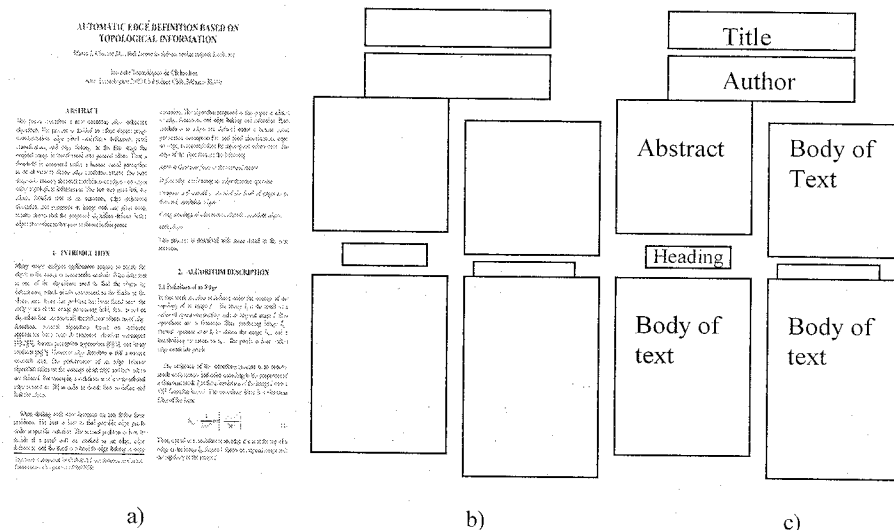


Figure 2: a) Document, b) Structural, c) Functional analysis.

2 Visual Perception Concepts

In this work it is assumed that the document regions have been segmented by a texture variance algorithm (Chacon 1998; Jain et al., 1995; Zhong et al., 1995). Thus, the next step is to decide what kind of information is in each region, text or non-text.

As mentioned before one of the purposes of this research is to incorporate concepts related to human visual perception

theories to perform document segmentation and classification. Thus, under this framework the first step is to discover what type of features a person may extract from the images to perform the classification. These features must be robust since the systems must work for free format documents relative to, size, skew, region shape, font, etc.

2.1 Textons and Preattentive Process

There are some theories from the human visual perception field that may be tested to perform document classification. Julesz has investigated how persons can discriminate different textures (Julesz 1975; Julesz et al., 1983). An important concept from Julesz's theories is the concept of textons. Julesz proposes that texture discrimination can be done in a preattentive process through detection of textons. Textons are defined as rectangular, line segments, or elliptical blobs that have specific characteristics like, color, angular orientation, width, length, and movement disparity. The preattentive process can be defined as a parallel process that detects textons differences without a conscious effort making texture discrimination possible. Figure 3 shows an example of how a texture can be discriminated without conscious effort.

Scialfa and Joffe (Scialfa et al., 1995) provide another example of how different textures can be discriminated by detecting general differences present in texture regions.

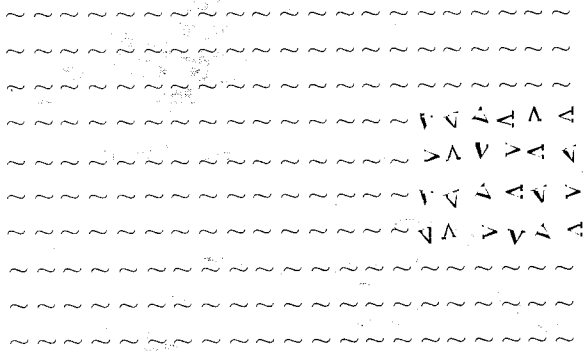


Figure 3: Preattentive texture discrimination.

This process can also be classified as a preattentive process. They note the characteristic that it is possible to define "where" the different regions are without knowing "what" they are. This experiment can be related to the text and non-text classification problem since a person can identify text versus non-text regions by looking at the difference in texture of both regions using very low resolution images in which the letters are not visible in the text regions. Figure 4 shows an example of how a target, different texture, is easily detected because of the gradient between the two regions.

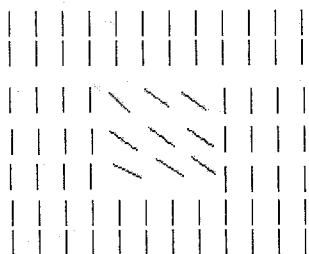


Figure 4: Texture target detection.

2.2 Perceptual Organization

The previous works describe how regions with different textures can be discriminated based on texton differences. McCafferty presents another important aspect of the human visual system, perceptual organization (McCafferty et al., 1990). Perceptual organization can be defined as the human visual process of finding relevant structures and groupings in a scene without prior knowledge. An important result of McCafferty's work is that his system developed to implement perceptual organization is based on Gestalt principles. The Gestalt psychology states that shapes involve specific features used by the human visual system to generate grouping, find structure, and generate interrelationships among the shapes. Some of the Gestalt principles are, proximity, similarity, good continuation, closure, etc. These are illustrated in Figure 5. In Figure 5a three vertical lines are perceived because the vertical distance is smaller than the horizontal distance. This is an example of the proximity principle. In Figure 5b four horizontal lines emerge from the grouping because of similarity of the elements. In Figure 5c, two curved lines are perceived, one from top to bottom and other from bottom to top due to good continuation. Good continuation says that the position of the next element is determined by a set of previous elements. Figure 5d illustrates closure which includes good continuation with filling. McCafferty tries to find structures and organization based on the minimization of an energy equation that includes some of the Gestalt principles. The structures are found by obtaining groupings such that the energy represented in equation 1 is minimized.

$$Total\ energy = \zeta_1 |E_{prox}(g)|^2 + \zeta_2 |E_{sim}(g)|^2 + \zeta_3 |E_{clos}(g)|^2 + \zeta_4 |E_{cont}(g)|^2 \quad (1)$$

Where each element of the equation represents a measurement of a Gestalt principle, and the parameters ζ are regularization parameters to weight the contribution of each principle to the grouping.

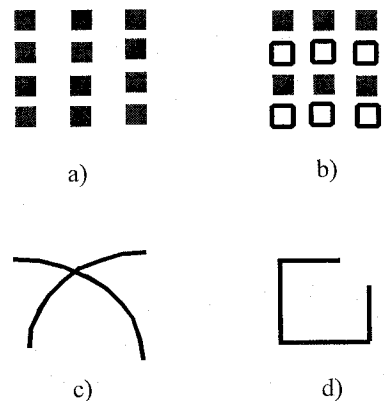


Figure 5: Gestalt principles. a) Proximity, b) Similarity, c) Good Continuation, d) Closure.

Concerning object discrimination, McCafferty (McCafferty et al., 1990) quotes Triesman who says that preattentive discrimination is possible if there exists a single property difference in the human visual mapping of properties. One possible property used by people to discriminate could be line orientation. This can be related to the concepts stated by Julesz and mentioned before. Julesz says that texture discrimination can be accomplished through the textons and one class of texton is line segments that have specific characteristics like, angular orientation. Figure 6 presents an example of how a target can be discriminated because of line orientation.

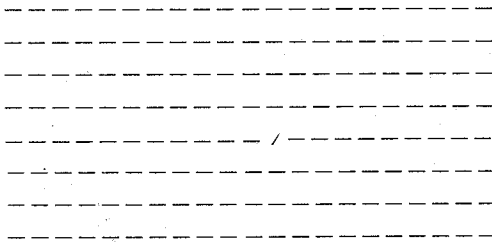


Figure 6: Preattentive discrimination, line orientation.

Based on these concepts and considerations, the classifier of a document system can be stated as a preattentive texture classification problem. A texture classification problem because the regions that will be classified, text and non-text, show different texture in a global sight, as shown in Figure 7. A preattentive process, since no available a priori knowledge is assumed

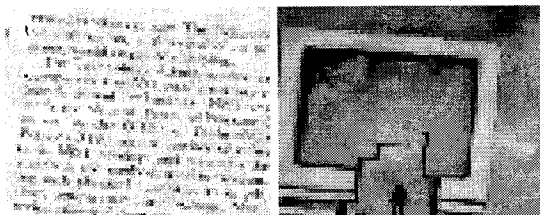


Figure 7: Text and picture regions.

3 Engineering Techniques to Obtain Features

The problem of document region classification has been stated on human visual perception characteristics. Now the next step is to obtain a measurement of this characteristics, features, that can be used for region classification.

Visual inspection of the images in Figure 8 reveals an important characteristic, the text section shows a set of blobs with specific characteristics. The blobs present similarity, proximity, continuity, and orientation. The blobs in the picture region look more like large uniform regions. These blobs can be related to the textons described by Julesz. Also, the structure generated by the blobs in the text region can be described under the perception organization domain with the principles of the Gestalt theory.

Considering Julesz's theories, Gestalt principles and perceptual organization concepts, the structures generated by the blobs in the text region may be used to discriminate between the texture of a text region and the texture of a picture region.

From these images, it is easy to detect that text regions are different from picture regions due to the organizational structure of the blobs. Text regions present well defined structures of lines with specific orientation, and periodicity. Picture regions present more uniform sections without any specific structure.

The organization of the blobs into line structures can be justified by the Gestalt principles of proximity, similarity and good continuation. From the point of McCafferty's work, line structures result by energy minimization by tracing a path from blob to blob. As shown in figure 9.

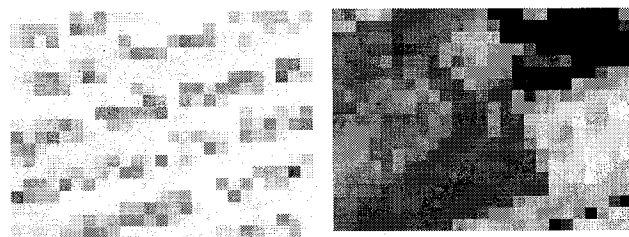


Figure 8: Augmented sections of text and picture regions.

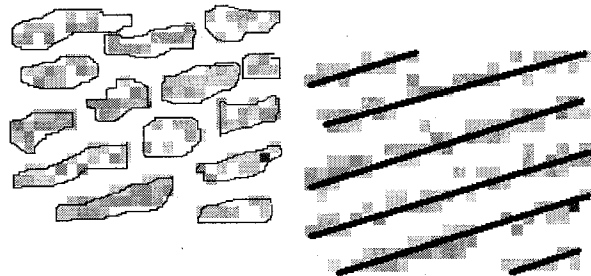


Figure 9: Blob organization in a text region.

3.1 Detection of Texton Organization

One way to detect the structure of the blobs (textons) lines with orientation, is applying the Hough transform. According to the definition of textons by Julesz, they can be rectangular, line segments, or ellipses. These shapes can be extracted by various forms of the Hough transform.

The Hough transform is a mapping from the (x,y) plane to the parameter space (Gloger 1992). For the purpose of this research, the Hough transform for line detection is sufficient. Equation 2 defines the Hough transform for line detection.

$$x \cos \theta + y \sin \theta = \rho \tag{2}$$

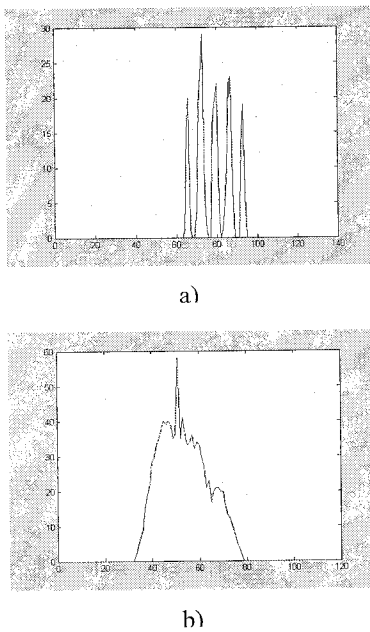


Figure 10: Typical θc of a) text and b) picture regions.

Where x and y are the coordinates of the point to map into the parameters θ and ρ . The parameter ρ is the magnitude of the perpendicular vector from the origin of the plane (x,y) to a line in the plane (x,y) . The parameter θ is the angle between the x axis and the perpendicular vector. The Hough transform is generated as follows. Equation 2 is solved for each point in the plane (x,y) assigning values to θ and solving for ρ . Each point in the (x,y) plane generates a set of values in the plane (ρ, θ) according to equation 2. Each of these points in the plane (ρ, θ) generates a vote in an accumulator. The accumulator is an $N \times M$ matrix. Where the rows correspond to values of ρ and the columns to values of θ . Each element of the matrix corresponds to a point in the plane (ρ, θ) . If the image under analysis contains one line, the accumulator of the Hough transform will contain an element with the maximum number of votes. The coordinates of this element in the plane (ρ, θ) correspond to the distance of the line with respect the origin in the plane (x,y) and the orientation of the line.

For the text region, the accumulator of the Hough transform presents a significant feature, the maximum values, in the accumulator of the Hough transform, are located at a specific angle, but even more important, they include the periodicity present in the organizational structure of the blobs. This means that it is possible to reduce the dimension of the feature from two dimensions to one dimension by selecting the column of the accumulator with the maximum value, θc , as a feature.

The accumulator corresponding to the picture region does not present this characteristic. This is an expected result since in the text region different lines can be traced in

different directions but the blobs are strongly organized into lines of text, thus most of the lines present a specific orientation, meanwhile in the picture region a large number of lines can be traced in different directions without any specific tendency. Figure 10 shows a typical pattern of θc for text and picture regions.

3.2 Feature Dimension Reduction

From Figure 10, it is obvious that θc of the text region is different from the θc of the non-text region. θc of text regions exhibits a quasi-periodic signal. The frequency depends on the interline spaces and the number of peaks is a function of the number of lines in the region selected. Since θc can characterize the two classes of regions, the following step is to compute a feature from θc that allows the discrimination of the regions. At first sight, periodicity is the most obvious characteristic in θc . Several signal processing techniques were used to extract a measurement of periodicity present in θc ; Fourier transform, Cepstrum, Autocorrelation, and Power Spectral Density (PSD). An analysis of the features obtained from the previous transformation concluded that the transformation that drew the best results turned out to be the PSD. For large text regions the PSD shows a significant peak as a consequence of the frequency content in the region and for non-text the spectrum is almost flat, revealing absence of frequency content. Thus, the features considered to classify large text region and picture regions are the peak magnitude of the PSD of θc and its index.

Based on the distributions of the samples and that the classifier is based on a preattentive process, the classification problem was stated as the classification of two types of regions: large text regions, LTR, more than one line of text as class 1, C1, and other regions, pictures, one line of text, one word, and one line, non-LTR class 2, C2. This decision is strongly supported by perception theories since one line of text and one word do not present the features presented in a large text region, line periodicity and orientation.

3.3 Biasing the Hough Transform

One of the disadvantages of the Hough transform is that it is computationally expensive. In order to reduce the computational time of the Hough transform, it is computed only over a section of the region that will be classified.

Therefore, a sampling procedure was chosen. The sample of the Hough transform to obtain the features of the region is computed only in a section of the region. In order to obtain more information about possible horizontal lines in the region, the Hough transform is computed over a rectangular section of size 32×64 pixels. This window biases the Hough transform reducing the amount of traces of vertical lines in text regions and reinforcing the principle of proximity as shown in Figure 11.

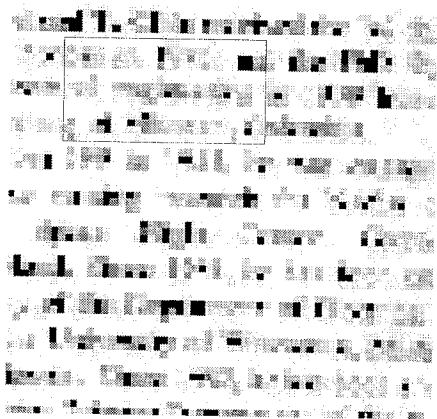


Figure 11: Reinforcement of proximity using a rectangular window.

Using the rectangular window, presence of horizontal lines is reinforced. At this point a new problem arises, what section of the region would be used to compute the Hough transform? The easiest solution is to select the section starting from the position of the seed. The seed is the first point inside of the region used to extract the information from the region. However, this draws some problems because of the shapes of the regions. For example the seed can be in a position such that just a few points of the region are used to compute the Hough transform as illustrated in Figure 12. Another solution is to use several sections. The problem of this solution is that it increases the computational time.

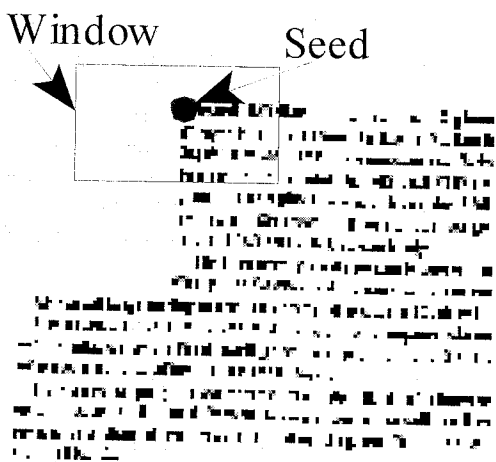


Figure 12: Problem if the position of the seed for the window is used.

The solution chosen was to make a visual classification test using information of three sections. The section from the seed, the section that consists of the quadrant of the region that contains the most black pixels, and the quadrant that

contains second largest number of black pixels. The classification test is performed using only the magnitude of the PSD of θ_c since this feature seems stronger than the index. This can be verified by looking the distributions of magnitude and index, Figure 13. The thresholds considered to make the classification are 0.2 and 0.8. These thresholds were selected from the class distributions by visual inspection of Figure 13a y 13 b. The results of this classification experiments are shown in table 1.

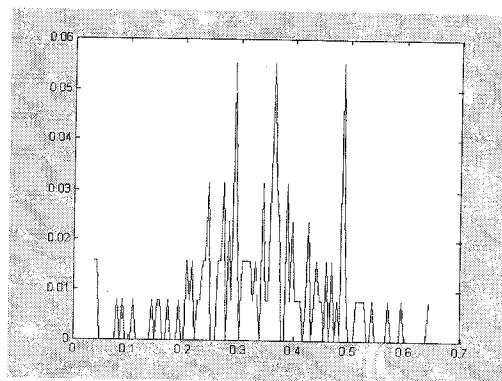


Figure 13a: Scattered graphs of the new class, Peak magnitude of Class 1.

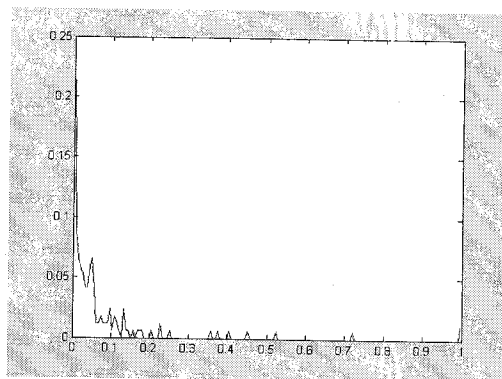


Figure 13b: Scattered graphs of the new class, Peak magnitude of Class 2.

Table 1: Classification by Section.

Classification by section			
Section Used	%Total Incorrect Classification	% LTR Wrong	%Non-LTR Wrong
1 Seed Section	9.5	15	6
2 More Black Pixel Section	9	15.9	5
3 Second Section With More Black Pixels	8	11.4	6
4 Using 1 & 2	9.5	15	5
5 Using 1 & 3	8.6	11	6.9
6 Using 2 & 3	9	12.65	6.9
7 Using 1 & 2 & 3	7.6	11.4	5.4

The total number of samples was 209, 79 were of large text regions and 130 of non-text regions. The best results considering total correct classification correspond to the experiment 3 and 7. Experiment 3 was selected since experiment 7 requires the computation of three sections increasing the computation time.

The complete process to obtain the features of θc , magnitude and index of maximum peak of θc , is illustrated in Figure 14. The first step is to binarize the region. The second step is to locate the second section of the region with more black pixels. To find this section the region is divided in four sections, and the number of black pixels are counted in each section. The third step is to compute the Hough transform. The fourth step is to find θc and multiply θc by a Hamming window defined by equation 3, centered at the maximum value of θc .

$$W_{ham} = \begin{cases} 0.54 - 0.46 \cos(2\pi n / (N-1)) & 0 \leq n \leq N-1 \\ 0 & \text{elsewhere} \end{cases} \quad (3)$$

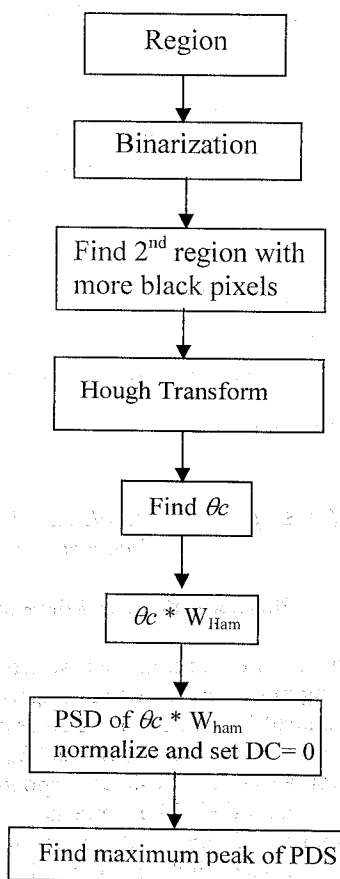


Figure 14: Flow diagram to obtain magnitude and index features.

The windowing process objective is to reduce the presence of artifacts in the Fourier transform. Then power spectral density is computed and normalized, and the DC value is set to zero. The PSD is differentiated to find the peaks by detecting changes of slope from positive to negative. From these peaks the maximum peak is selected. The approximation used to compute the derivative is:

$$x(n)' = [x(n) - x(n-1)]/T \quad (4)$$

Now that the features and the procedure to obtain the features have been defined, the next step is to design the classifier.

4 Classifier Design

The classifier designed is an artificial fuzzy neural network (AFNN). AFNN has been used with good results in recognition and classification problems (Mohamed, and Nyongesa, 2002; Xiaou; Xinbo, Jianzhuang, and Hongjiang, 2002). The basic idea of using a AFNN is that in some cases it is better to train a neural network to produce intermediate results, preserve generalization, instead of overtraining the network to produce final results (Masters 1993). Final results are computed by a postprocessing module. This module is design to resolve the uncertainty present in the outputs of the network. There are two main advantages by proceeding in this way. One is that the training stage will be easier since the network will have to approximate just the result, consequently the training is faster. The other advantage is versatility.

Versatility in the sense that if different but related results are desired, it is not necessary to train the network again. Only the postprocessing module needs to be updated. One disadvantage of this process, as with some other issues related to artificial neural networks, is that there does not exist analytical foundations to define where to stop the training of the neural network. Therefore, several networks were trained for different sum of squared errors (SSE). Only the networks with better performance are described.

The architecture of the AFNN is as follows:

Input vector $\mathbf{X} = [\text{peak magnitude, index}]$

Layers: 5:15:2

Activation function: $f(\text{net}) = 2/(\exp\{-\text{net}\}) - 1$

Outputs: two

The architecture 5:15:2 was defined by experimental analysis of performance of different architectures.

The magnitude was already normalized, so it was decided to normalize the index element to avoid traversing long valleys in the index dimension, since the neural networks are trained with the backpropagation algorithm (a gradient

descendent method). The normalization of the index values is performed by a Min-Max approach

$$q' = [(q - \min1) / (\max1 - \min1)](\max2 - \min2) + \min2 \quad (5)$$

where q represents the original value to normalize, q' is the new value, $\min1$ and $\max1$ are the minimum and maximum values of q , and $\min2$ and $\max2$ are the new minimum and maximum limits.

The uncertainty of the neural network outputs is managed through fuzzy logic operations related to fuzzy sets. Fuzzy sets and fuzzy logic are extensions of crisp set and two-valued logic. A fuzzy set is a generalization of a crisp set by generalizing the range of the membership function, or characteristic function defined as $\mu_D(x)$ (Dubois, and Prade, 1988) which is a generalization of the indicator function $I_C(x)$. The membership function maps the range of the crisp set, $\{0,1\}$ to the unit interval $(0,1)$.

$$I_C(x) = \begin{cases} 1 & \text{if } x \in C \\ 0 & \text{if } x \notin C \end{cases} \quad (6)$$

$$\mu_D(x) : U \rightarrow (0,1) \quad (7)$$

Thus a fuzzy set D in the universe of discourse U is defined as a set of ordered pairs

$$D = \{ (x, \mu_D(x)) \mid x \in U \} \quad (8)$$

$\mu_D(x)$ defines the degree of membership of the element x to the fuzzy set D .

The design of this fuzzy neural network starts by training the network to keep generalization. Several fuzzy neural networks were designed and tested but just the most important experiments are described.

Generalization is achieved by setting a high sum of squared error SSE. A fuzzy processor is then used to implement the decision rule. The first experiment, neural network 8a, was trained for an SSE of 18 corresponding to 6 samples incorrectly classified. The fuzzy postprocessing decision rule is implemented as follows.

Assume that the target vectors for the neural network are

$$c1 = [-.9 \quad .9]^T \quad \text{and} \quad c2 = [.9 \quad -.9]^T$$

Define the fuzzy set for I with the membership function $\mu_I(x)$, by an S -curve

$$S(x : \alpha, b, \gamma) = \begin{cases} 0 & x \leq \alpha \\ 2((x - \alpha) / (\gamma - \alpha))^2 & \alpha \leq x \leq \beta \\ 1 - 2((x - \alpha) / (\gamma - \alpha))^2 & b \leq x \leq \alpha \\ 1 & x \geq \gamma \end{cases} \quad (9)$$

and fuzzy set $-I$ with membership function $\mu_{-I}(x)$ by

$$1 - S(x : \alpha, \beta, \gamma) \quad (10)$$

Figure 15 illustrate the fuzzy sets $\mu_I(x)$ and $\mu_{-I}(x)$. Given the outputs of the network, O_1 and O_2 , obtain the membership of O_1 to the fuzzy set I , $\mu_I(O_1)$, the membership of O_1 to the fuzzy set $-I$, $\mu_{-I}(O_1)$, the membership of O_2 to the fuzzy set I , $\mu_I(O_2)$, and the membership of O_2 to the fuzzy set $-I$, $\mu_{-I}(O_2)$.

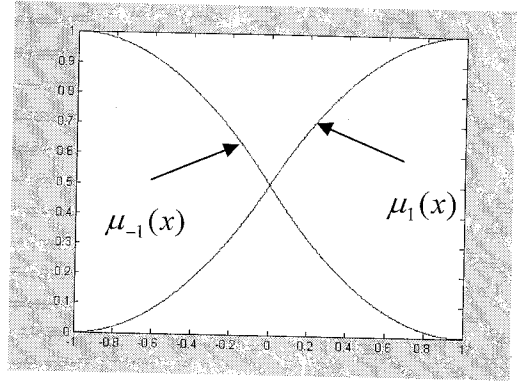


Figure 15: Fuzzy sets $\mu_I(x)$, and $\mu_{-I}(x)$.

Now compute the membership value of the feature vector X to class one $\mu_{c1}(x)$, and to class two, $\mu_{c2}(x)$. The feature vector X has a high membership value to class 1 as O_1 tends to $-I$ and O_2 tends to I . In the same way, the feature vector X has a high value membership to class 2 if O_1 tends to I and O_2 tends to $-I$, according to the target vectors. These membership values are obtained by fuzzy addition as follow.

$$\mu_{c1}(x) = \mu_{-I}(O_1) + \mu_I(O_2) - \mu_{-I}(O_1) \mu_I(O_2) \quad (11)$$

and

$$\mu_{c2}(x) = \mu_I(O_1) + \mu_{-I}(O_2) - \mu_I(O_1) \mu_{-I}(O_2) \quad (12)$$

Finally the decision rule is

$$\begin{array}{ll} \text{If} & \mu_{c1}(x) \leq \mu_{c2}(x) \quad \text{decide class1} \\ \text{Else} & \text{decide class2} \end{array}$$

The performance of this network was 94.08% correct classification.

The next experiments were to train neural networks for a larger SSEs allowing more generalization. Experiment 8b is a neural network trained for an SSE of 72 corresponding to 21 samples incorrectly classified. Experiment 8c is a neural network trained for an SSE of 140 corresponding to 42 samples incorrectly classified. The percentage of correct classification is 95.3 and 95.06 for the experiment 8b and 8c respectively. These findings now indicate that subtraining the neural network up to some SEE draw better generalization and using a fuzzy process to solve the network uncertainty produces a better classifier.

The last experiment performed with the fuzzy neural network, experiment 8m, incorporates a weight to the risk

for incorrect classification of class 1 samples as class 2, (c2/c1). The risk cost was incorporated in the fuzzy process by means of a hedge. A hedge is basically a membership function modifier. The objective of the hedge is to transform a membership function such that the new membership function better represents a linguistic characteristic. The hedge used in this experiment is of the form

$$\mu_A^\eta(x) \tag{13}$$

where η has values greater than 1. This type of hedge will decrease the membership function as shown in Figure 16. In order to decrease the error (c2/c1), penalize this error, the hedge is applied to the membership function $\mu_1(x)$ to compute the value of $\mu_1(O_1)$ and to $\mu_{-1}(x)$ to compute $\mu_{-1}(O_2)$. The function $\mu_{-1}(x)$ to compute $\mu_{-1}(O_1)$ and $\mu_1(x)$ to compute $\mu_1(O_2)$ stay unchanged. The value of η , in equation 13, tested was 1.5, 2, and 3. The fuzzy membership functions with the hedge operator used to make the decision are shown in Figure 17.

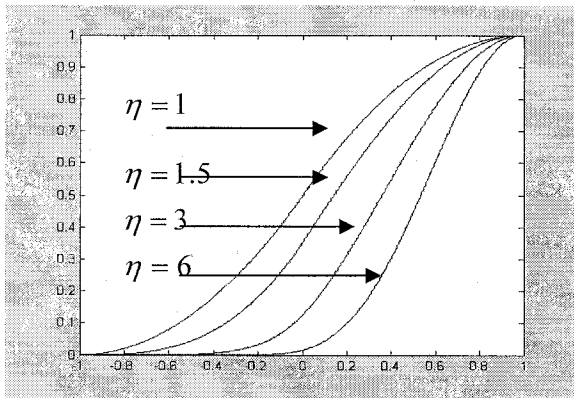


Figure 16: Original membership function and the function after the hedge operator.

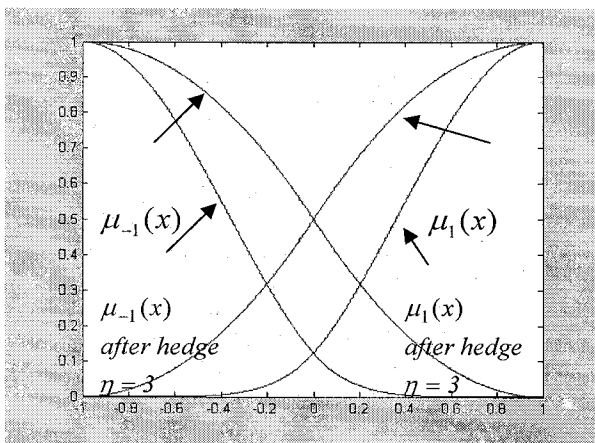


Figure 17. The fuzzy membership functions with the hedge operator used to make the decision

The results of the experiment using a hedge were similar to the experiments 8b. It can be observed that adding a risk cost to the classifier does not improve its performance.

Figure 18 2 summarizes the best fuzzy neural network. The results are based on 512 samples.

Figure 19 illustrate some examples of document images after classification with the fuzzy neural network classifier. LTR regions are marked with gray color and non-LTR with black.

5 Conclusions

The method developed in this research is a novel method for document region classification. The main breakthrough is that the method is based on human visual perception theories that makes this method very different from another methods. The method simulates in some way a human behavior for document region classification. It works with low resolution document images (40 and 50 dpi), and free format documents, yielding a robust system for document region classification.

This work has proved that more robust document region classification systems may be developed if human visual perception theories are incorporated in the design. The system described here may be faster than systems derived from traditional methods since this method works with low resolution images. Working with low resolution images reduces the amount of information that is processed. Also, since the system is based on human visual perception theories, it incorporates some of the robustness of the human system.

The fuzzy neural network classifier was tested on 40 different document images yielding a 95.7% of correct classification. An important conclusion can be drawn from the analysis of the results of the classifier that penalizes the error of classification. Since a cost for error classification was incorporated and the results do not change significantly, it may indicate that the classification errors are directly related to the quality of the images. Therefore, it can be concluded that the classifiers are operating in their performance boundary. In other words, optimal classification for low resolution images was achieved. One possible way to improve the classification rate is to try to improve the quality of the images or increase the resolution to reduce segmentation errors and obtain better information inside the regions. This, of course, will increase computational time.

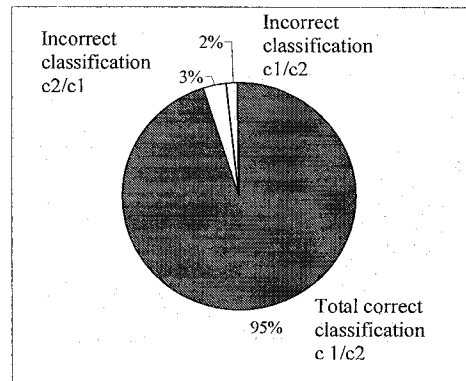


Figure 18. Best fuzzy ANN classifier final results

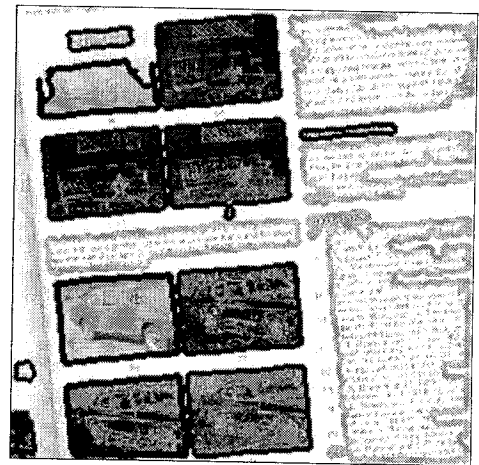
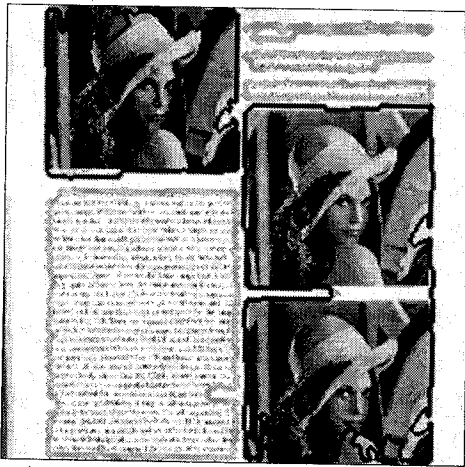


Figure 19 cont. Results of classification

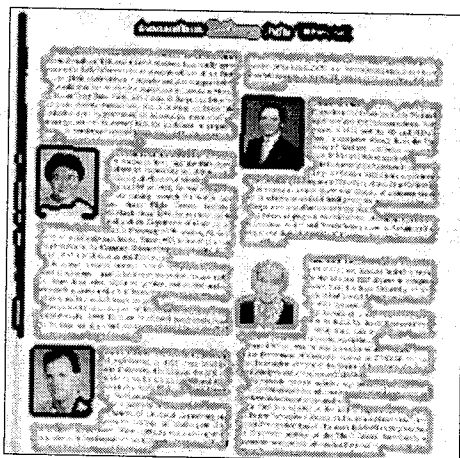


Figure 19. Results of classification

References

Bezdek, J., and S. Pal. "Fuzzy Models for Pattern recognition Methods that Search for structures in Data", 1992, IEEE Press.

Bloomberg, D. "Multiresolution Morphological Analysis of Documents Images". SPIE Visual Communication and Image Processing, Vol. 1818, 1992, pp 648-662.

Boskovitz, V., and H. Guterman, "An Adaptive Neuro-Fuzzy System for Automatic Image Segmentation and Edge Detection", Fuzzy Systems, IEEE Transactions on , Volume: 10 Issue: 2 , April 2002, pp 247 –262.

Chacon M. "Document Segmentation Using Texture Variance and Low Resolution Images", in IEEE International Southwest Symposium on Image Analysis and Interpretation., Tucson, Arizona, April, 1998 pp 164-167.

Dubois, D., and H. Prade, "Fuzzy Sets and Systems Theory and applications". 1980. Academic Press.

Etemad K., D. Doermán, and R. Chellapa , "Page Segmentation Using Decision Integration and Wavelete Packets", in IEEE International Conference on Pattern Recognition, 1994, pp 345-349.

Fletcher, L., and R. Kasturi. " A Robust Algorithm for Text String Segmentation from Mixed Text/Graphics Images". IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 10, No 6, 1988, pp 910-918.

Chung-Hoon F., and L. Changsu, "Region Based Fuzzy Neural Networks for Face Detection", IFSA World

Congress and 20th NAFIPS International Conference, 2002. Joint 9th, Volume: 2, 2001, pp 1156-1160.

Gloger, J. " Use of Hough Transform to Separate Merged Text/Graphics in Forms". IEEE International Conference on Pattern Recognition, 1992., pp 268-271.

Haralick R., K. Shanmigan, and I. Dinstein , "Textural Features for Image Classification", in IEEE Transactions on System, Man and Cybernetic", Vol SCM-3 No 6, 1973, pp 610-621.

Jain, A., and S. Bhattacharjee, " Text Segmentation Using Gabor Filters for Automatic Document Processing", in Machine Vision and Application. Spring-Verlag. Vol 5, 1984,pp 169-184.

Jain, A., and Y. Zhong, "Page Segmentation Using Texture Discrimination Masks", in IEEE International Conference on Image Processing, 1995,pp 308-311

Julesz, B. ,"Experiments in the Visual Perception of Texture", Scientific America, 1975, pp 34-43.

Julesz, B., and R. Bergen, " Textons, the Fundamentals Elements in Preattentive Vision and Perception of Textures", The Bell System Technical Journal, Vol 62 No 6., 1983

Masters, T., "Practical Neural Networks", Academic Press. 1993.

McCafferty J, "Human and Machine Vision Computing Perceptual Organization", Ellis Horwood. 1990.

Mohamed S., and H. Nyongesa "Automatic Fingerprint Classification System Using Fuzzy Neural Techniques", in Fuzzy Systems, 2002. FUZZ-IEEE'02. Proceedings of the 2002 IEEE International Conference on , Volume: 1 , 2002, pp 358 –362.

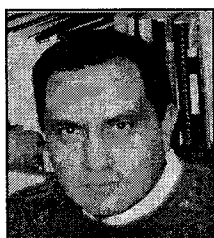
O'Gorman, L., and R. Kasturi , "Document Image Analysis", IEEE Computer Society Press, 1995.

Vaucher G.,A. R. Baig, and R. Séguier, "A Set of Neural Tools for Human-Computer Interactions: Application to the Handwritten Character Recognition, and Visual Speech Recognition Problems", Neural Computing & Applications, Volume 9 Issue 4 , 2000,pp 297-305.

Scialfe, C., and K. Joffe , "Preferential Processing of Target Features in Texture Segmentation". Perception and Psychophysics. Vol 57 No 8, 1995, pp 1201-1208.

Xiaou T., G. Xinbo, L. Jianzhuang, and Z. Hongjiang , "A spatial-temporal approach for video caption detection and recognition", Neural Networks, IEEE Transactions on, Vol. 13 Issue: 4 , July 2002 , pp 961 –971.

Zhong, Y., K. Karu, and A. Jain, " Locating Text in Complex Color Images", in Pattern Recognition, Vol. 28, No. 10, 1995, pp 1523-1535.



Mario Ignacio Chacón Murguía, received the BS in Electrical Engineering and the MS in Computer Engineering degrees from the Instituto Tecnológico de Chihuahua, Chihuahua México in 1982 and 1985 respectively and his Ph.D. in Electrical Engineering from New Mexico State University, USA, in 1998. He has made research for the industry and other universities. Dr. Chacón has several scientific and technical international publications. He is a member of the IEEE and the Pattern Recognition Society. His current research interests include Computer Vision, Neural Networks and Fuzzy Logic theory applied to DSP problems, Digital Signal Processing (DSP), Pattern Recognition, and Digital Systems design.

