EXTRACTION OF OBJECTS AND PAGE SEGMENTATION OF COMPOSITE DOCUMENTS WITH NON-UNIFORM BACKGROUND

Yasser Alginahi, Maher Sid-Ahmed, Majid Ahmadi
Univerisyt of Windsor, 401 Sunset Ave, Windsor, Ontario,Canada

Keywords: Statistical features, Multi-Layer Neural Networks, Non-uniform background.

Abstract: In designing page segmentation systems for documents with complex background and poor illumination, separating the background from the objects (text and images) is very crucial for the success of such system. The new local based neural binarization technique developed by the authors will be used to extract the objects from document images with complex backgrounds. This algorithm uses statistical and textural feature measures to obtain a feature vector for each pixel from a window of size $(2n+1)\times(2n+1)$, where $n \geq 1$. These features provide a local understanding of pixels from their neighbourhoods making it easier to classify each pixel into its proper class. A Multi-Layer Perceptron Neural Network (MLP NN) is then used to classify each pixel in the image. The results of thresholding are then passed to a block segmentation stage. The block segmentation technique developed is a feature-based method that uses a Neural Network classifier to automatically segment and classify the image contents into text and halftone images. The results of page segmentation are then ready to be passed into an OCR system that will convert the text image into a format the can be stored and modified.

1 INTRODUCTION

Document image analysis is an important area of research in image processing, pattern recognition and computer vision. The goal of our research is to process grey level document images with complex backgrounds, bad illumination and poor contrast. The motivation behind most of the applications of off-line text recognition is to convert data from the conventional media into electronic media. In this paper, a document segmentation system is presented to transfer grey level composite images with complex backgrounds and poor illumination into electronic format that is suitable for efficient storage retrieval and interpretation. Such applications are bank cheques, security documents and form processing.

There are many threshold selection schemes published in the literature, and selecting an appropriate one can be a difficult task. Thresholding of documents can be categorized into two main classes: global and local thresholding. Global thresholding techniques use a single threshold value; on the other hand, local thresholding compute a separate threshold based on the neighbourhood of the pixels. (Sahoo et al., 1998) showed that the Otsu's (Otsu, 1979) class separability thresholding method is the best global thresholding method. In (Trier and Jain, 1995), Trier and Jain showed the Niblack's (Niblack, 1986) method to be the best local thresholding method compared to other methods. Few methods used NNs in thresholding grey scale images into two levels. The technique proposed by Koker and Sari (Koker and Sari, 2003), uses NNs to select a global threshold value for an industrial vision system based on the histogram of the image. The method developed by Papamarkos (Papamarkos, 2001) uses Self Organizing Feature Maps (SOFM) to define two bi-level classes. Then, the contents of these classes are used with the fuzzy C-mean algorithm to reduce the character blurring effect. Both methods are not suitable for thresholding composite images with complex backgrounds. In this paper, a new threshold selection algorithm is proposed which handles images with non-uniform and complex backgrounds. The new method uses a MLP NN with statistical and textural feature measures as inputs to the network.

Alginahi Y., Sid-Ahmed M. and Ahmadi M. (2005).
EXTRACTION OF OBJECTS AND PAGE SEGMENTATION OF COMPOSITE DOCUMENTS WITH NON-UNIFORM BACKGROUND.
In *Proceedings of the Second International Conference on Informatics in Control, Automation and Robotics - Signal Processing, Systems Modeling and Control*, pages 344-347
DOI: 10.5220/0001167903440347

2 STATISTICAL FEATURE MEASURE

The new NN local thresholding method, takes advantage of the document texture characteristics by considering the statistical texture descriptors in a neighbourhood of pixels. The statistical texture descriptors used were defined over a window of size $(2n+1)\times(2n+1)$ centered at (i,j), where $n=2$. The features extracted are as follows:

2.1 Actual Pixel Value

The first feature extracted is the center pixel $p(i,j)$ in the window.

2.2 Mean

The mean, μ_{ij}, of the pixel values, in the defined window.

$$\mu_{ij} = \frac{1}{(2n+1)^2}\sum_{x=i-n}^{i+n}\sum_{y=j-n}^{j+n} p(x,y) \qquad (1)$$

2.3 Standard deviation

The Standard deviation, σ_{ij}, is the estimate of the mean square deviation of grey pixel value, $p(x,y)$, from its mean value, μ_{ij}.

$$\sigma = \frac{1}{(2n+1)}\sqrt{\sum_{x=i-n}^{i+n}\sum_{y=j-n}^{j+n}[p(x,y)-\mu_{ij})]^2} \qquad (2)$$

2.4 Skewness

Skewness, S_{ij}, characterizes the degree of asymmetry of a pixel distribution around its mean.

$$S_{ij} = \frac{1}{(2n+1)^2}\sum_{x=i-n}^{i+n}\sum_{y=j-n}^{j+n}\left[\frac{p(x,y)-\mu_{ij}}{\sigma_{ij}}\right]^3 \qquad (3)$$

2.5 Kurtosis

Kurtosis, K_{ij}, measures the relative peakness or flatness of a distribution relative to normal distribution.

$$k_{ij} = \left\{\frac{1}{(2n+1)^2}\sum_{x=i-n}^{i+n}\sum_{y=j-n}^{j+n}\left[\frac{p(x,y)-\mu_{ij}}{\sigma_{ij}}\right]^4\right\} - 3 \qquad (4)$$

The –3 term makes the value zero for a normal distribution.

2.6 Entropy

Entropy, h_{ij}, describes the distribution variation in a region. Entropy of pixel $P(i,j)$ can be calculated as

$$h_{ij} = -\sum_{k=0}^{L-1} p_k \log p_k \qquad (5)$$

Where p_k is the probability of the k^th grey level, which can be calculated as $z_k/(2n+1)^2$, z_k is the total number of pixels with the k^th grey level and L is the total number of grey levels in the window.

2.7 Relative Smoothness

Relative Smoothness, R_{ij}, is a measure of grey-level contrast.

$$R_{ij} = 1 - \frac{1}{1+\sigma_{ij}^2} \qquad (6)$$

2.8 Uniformity

Uniformity, U_{ij}, is a texture measure based on histogram and is defined as:

$$U_{ij} = \sum_{k=0}^{L-1} P_k^2 \qquad (7)$$

Before computing any of the descriptive texture features above, the pixel values of the image were normalized by dividing each pixel by 255 in order to achieve computation consistency.

3 NN LOCAL THRESHOLDING TECHNIQUE USING MLP

The new local NN thresholding technique uses a MLP neural network (Sid-Ahmed, 1995) to classify document images into background and foreground. Random representative sample pixels from several images with different complex backgrounds and degraded images were selected as training sets for the NN.

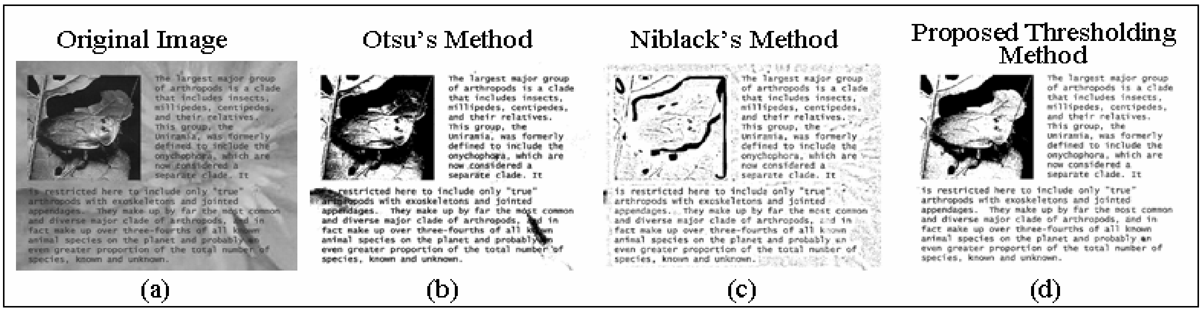

Figure 1: Example results showing the thresholding of the Otsu's, Niblack's and proposed methods

The features related to the neighborhood pixels in the neighborhood window, were calculated and the actual output value for the pixel was recorded, 1 for background and 0 for foreground. These feature values were then fed into the MLP NN to train the network. The MLP NN is trained by supervised learning using the iterative back-propagation algorithm, which minimizes the mean square error between the network's output and the desired output for all input patterns. Once the NN has been trained, the weights are used in the classification phase. During classification, image data feature vectors extracted from each pixel in the image are fed into the network that performs classification by assigning a class number, either 1 or 0, for each pixel. Figure 1 shows the results of the Otsu's, Niblack's and the proposed NN local thresholding techniques for an example image.

4 NEURAL SEGMENTATION

4.1 Introduction

The RLSA (Wong and Casey, 1982) technique is one of the most widely used top-down segmentation algorithms. It is used on binary images to classify images into text and halftone images. By linking together the neighbouring black pixels that are within a certain threshold to form blocks of text and/or images, this method is applied row-by-row (horizontal smearing) and column-by-column (vertical smearing), then both results are combined in a logical AND operation. From the RLSA results, single blocks of text lines and images are produced. Wahl et al. in (Wong and Casey, 1982) used a statistical classifier to classify these blocks, but in this work a NN classifier will be used to automatically segment the contents of the images. Figure 2 shows the thresholding, and segmentation process.

4.2 Block Labelling

Labels have to be assigned to different blocks to identify each block separately to be used in the feature extraction step. All connected pixels must have the same label. A Local Neighbourhood Algorithm (Rosenfeld and Kak, 1976) is used which scans the image horizontally until it hits the first pixel then a fire is set at this point that propagates to all 8-neigborhood of the current pixels giving them the same label.

4.3 Feature Extraction

After block labeling, the coordinates of each block are known. The next step is to extract features. Geometrical and statistical features were extracted from each block of text and images. The following features were extracted: the height (H_i), mean pixel value (μ_i), standard deviation of pixels (σ_i) and black pixel count (BC_i) calculated from the binary image where i is the label corresponding to each block.

4.4 Block Classification

The final step in page segmentation is the classification of each block into its proper class. In Whal (Wong and Casey, 1982), a statistical classifier was used with parameters that were not optimized and are application dependent. In this paper, a NN classifier using MLP is used with the features extracted fed into the NN to produce the proper classification between text and images. The NN contained three layers, with 4 nodes at the input layer, 7 nodes in the hidden layer and one node at the output layer. The results of block classification are shown in the example of Figure 2.

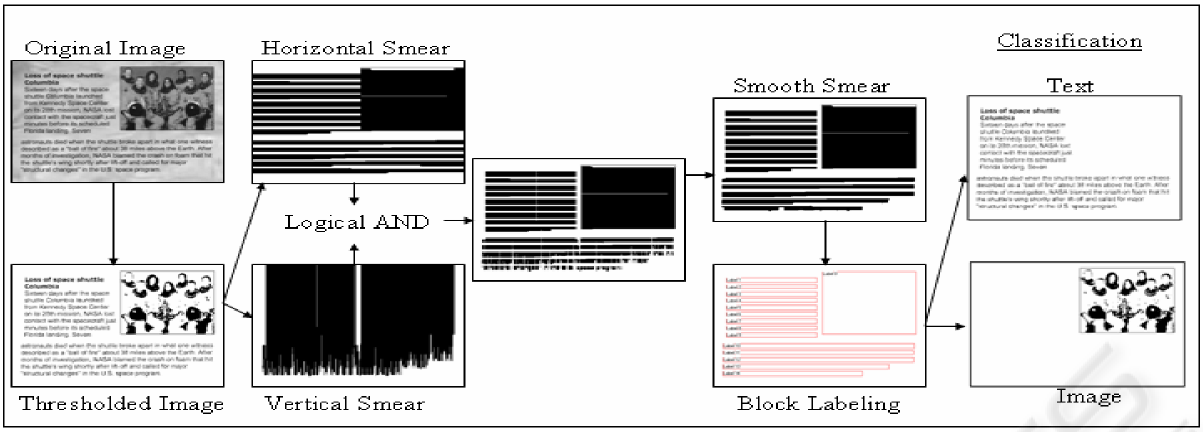

Figure 2: Example showing the process of thresholding and page segmentation**.**

5 RESULTS

The new local NN thresholding method was tested on over hundred images and the results obtained outperformed other thresholding selection schemes as seen from Figure 1. This new neural based technique calculates an optimal threshold value for each pixel when passed through the NN making it a suitable method to be used in character recognition applications. The proposed neural based page segmentation method which uses the RLSA to produce blocks of text and images provided good separation between different contents of a document image as seen in Figure 2. These results in most cases do not need any post processing since the thresholding produced binary images that are free of noise. The results are then ready to be fed into an OCR system to extract the text from the image. Testing this system on hundred images provided a 98.6% rate in separating different contents of an image into text and halftone images.

6 CONCLUSION

The proposed thresholding technique outperformed other techniques and the results produced are free of noise making it a good choice to use in extracting objects from image documents with non-uniform background. The Neural based page segmentation provided a good separation between text/halftone image contents of an image document
and can be extended to include other contents in image documents. These results requires no or minimal post processing to be used in further stages of document analysis to guarantee the success of the character recognition system in providing higher recognition rate.

REFERENCES

PK. Sahoo, S. Soltani, AKC Wong, 1988. "A Survey of thresholding techniques", *Computer Vision, Graphics and Image Procs*, Vol. 41, pp. 233-260.

N. Otsu, 1979. "A Threshold Selection Method from Grey-Level Histograms", *IEEE Trans. On System Man and Cybernetics*, vol. 9, no. 1, pp. 62-69.

O. D. Trier and A. K. Jain, 1995. "Goal-Directed Evaluation of Binarization Methods", *IEEE Trans Pattern Analysis and Machine Intelligence*, Vol. 17, No. 12, pp. 1191-1201.

W. Niblack, 1986. "*An Introduction to Digital Image Processing*" Englewood Cliffs, Prentice Hall, NJ. pp 115-116.

R. Koker and Y. Sari, 2003. "Neural Network Based Automatic Threshold Selection For an Industrial Vision System," *Proc. Int. Conf. On Signal Processing*. pp. 523-525.

N. Papamarkos, 2001. "A Technique for Fuzzy Document Binarization," *Procs. Of the ACM Symposium on Document Engineering*, pp.152-156.

M. A. Sid-Ahmed, 1995. "*Image Processing: theory, algorithms, and architectures*", McGraw-Hill, pp. 313-375.

Wahl, K. Wong and R. Casey, 1982. "Block Segmentation and Text Extraction in Mixed Text/Image Documents," *Computer Vision, Graphics and Image Processing,* Vol. 20, pp. 375-390.

A. Rosenfeld and A. C. Kak, 1976. "*Digital Picture Processing,*" Academic Press, NY, pp. 347-348.