A Text Detection Method Using Wavelet Packet Analysis and Unsupervised Classifier

Geum-Boon Lee, Wilfred O. Odoyo, Kuk-Se Kim and Beom-Joon Cho

Abstract— In this paper we present a text detection method inspired by wavelet packet analysis and improved fuzzy clustering algorithm(IAFC).This approach assumes that the text and non-text regions are considered as two different texture regions. The text detection is achieved by using wavelet packet analysis as a feature analysis. The wavelet packet analysis is a method of wavelet decomposition that offers a richer range of possibilities for document image. From these multiscale features, we adapt the improved fuzzy clustering algorithm based on the unsupervised learning rule. The results show that our text detection method is effective for document images scanned from newspapers and journals.

Index Termss— document image segmentation, wavelet packet analysis, improved fuzzy clustering algorithm.

I. INTRODUCTION

The conversion of document images into electronic form is a significant and growing task in information and technology. Document image analysis is concerned with the segmentation of the document image into text / non-text regions and the classification of the regions according to their textures. A number of techniques have been proposed in the literature aiming to facilitate automatic segmentation [1].

The earlier approaches to document image segmentation have assumed that all printed areas in the document images are rectangular. This limits not only in the case where these areas are of different shapes, but also where skew has been introduced during the scanning of the document images [2]. Several other approaches use the contours of the white space to represent the text and non-text regions [3], [4], [5]. These methods assumed that all objects are separated by a white background and objects do not touch each other. Other common approaches are texture-based methods which use the different texture properties of text regions and non-text regions. Texture is an important characteristic for the analysis of many types of images including natural scene, remotely sensed data and biomedical data. The perception of texture is believed to play an important role in document image segmentation and classification [5].

In this paper, we introduce a new technique for texture discrimination based on wavelet packet analysis. We suggest that such analysis can provide a powerful method for accomplishing robust texture classification, compared to traditional single resolution techniques. In section 2, we discuss the wavelet packet decomposition and review the local energy analysis for texture discrimination. Also, we show an unsupervised clustering algorithm and discuss the several parameters considered in our investigation for segmentation. Section 3 presents segmentation experiments of document images and the results of wavelet packet analysis. Finally, section 4 summarizes the results and conclusions of our study.

II. TEXT DETECTION USING PROPOSED METHOD

The wavelet packet method is a generalization of wavelet decomposition that offers a richer range of possibilities for signal analysis. The wavelet packet is a family of scaling functions and wavelets constructed by following a binary tree of dilations and translations. In wavelet analysis, a signal is split into approximation and detail. The approximation is then itself split into a second-level approximation and detail, and the process is repeated. For n-level decomposition, there are n+1 possible ways to decompose or encode the signal. The wavelet analysis and the wavelet packet analysis are similar but in the corresponding wavelet packet situation, each detail coefficient is also decomposed into two parts using the same approach as in approximation splitting. The purpose of this decomposition is to start from a scale-oriented decomposition and then analyze the obtained signals on frequency subbands. In the wavelet packet analysis, a suitable wavelet packet that best represents the feature of the clutter is used to perform the wavelet packet transform of the input signal [6].

2.1 Wavelet Packet Analysis

Wavelet packets may be described by the collection of functions $W_j = (W_j(x),\ j \in Z^+)$ obtained from [4].

Manuscript received November 6, 2006.

Geum-Boon Lee is a ph.D. course student in Dept. of Computer Engineering, Graduate School, Chosun University, Seosuk-dong, Dong-gu, Gwangju, 501-759, Korea(email : goldylee@empal.com)

$$2^{\frac{p-1}{2}} W_{2n}(2^{p-1}x - l) = \sum_m h_{m-2l} 2^{\frac{p}{2}} W_n(2^p x - m) \tag{1}$$

$$2^{\frac{p-1}{2}} W_{2n+1}(2^{p-1}x - l) = \sum_m g_{m-2l} 2^{\frac{p}{2}} W_m(2^p x - m) \tag{2}$$

Where p is a scale index, l is a translation index. $W_0(x) = \phi(x)$, $W_1(x) = \psi(x)$, $\phi(x)$ is a scaling function and $\psi(x)$ is a wavelet. The discrete h_k and g_k are quadratic mirror filters [7].

We can show that such wavelet packets are orthonormal in $L^2(R)$ and well localized in both time and frequency. The inverse relationship between wavelet packets of different scales can be shown by,

$$2^{\frac{p}{2}} W_n(2^p x - k) = \sum_l h_{k-2l} 2^{\frac{p-1}{2}} W_{2n}(2^{p-1}x - l) + \sum_l g_{k-2l} 2^{\frac{p-1}{2}} W_{2n+1}(2^{p-1}x - l) \tag{3}$$

The function $f(x) \in L^2(R)$ can be decomposed onto a wavelet packet basis. The coefficients of this decomposition are simply the inner products of $f(x)$ with distinct wavelet packets. For example, coefficients from the inner product $\langle f(x),\ W_n(2^p x - k)\rangle$ indicate the intensity of this component in $f(x)$ using wavelet W_n at a scale 2^p can be written as

$$A_n^{2p} f(x) = \sum_k S_{n,k}^p 2^{\frac{p}{2}} W_n(2^p x - k) \tag{4}$$

Where

$$S_{n.k}^p = 2^{\frac{p}{2}} \int_{-\infty}^{\infty} f(x) W_n(2^p x - k) dx \tag{5}$$

We show how wavelet packets may be computed efficiently. From 3, we may write

$$S_{n,k}^p = \sum_l h_{k-2l} S_{2n,l}^{p-1} + \sum_l g_{k-2l} S_{2n+1,l}^{p-1} \tag{6}$$

Using (1) and (2), coefficients at coarser scales are calculated by

$$S_{2n,l}^{p-1} = \sum_m h_{m-2l} S_{n,m}^p \tag{7}$$

$$S_{2n+1,l}^{p-1} = \sum_m g_{m-2l} S_{n,m}^p \tag{8}$$

The basis functions are obtained by translation and dilation. They remain well localized in both spatial and frequency domains and thus represent scale and spatial information. Thus, a complete tree presents the distribution of a signal within a scale space. The total number of coefficients in complete tree decomposition is exactly equal to the number of points in an original signal. Energy distributions within transform spaces have been applied in Fourier analysis [8]. Since wavelet packets form orthogonal bases, their decompositions preserve energy. It is easy to show that

$$\sum_k (S_{n,k}^p)^2 = \sum_l (S_{2n,l}^{p-1})^2 + \sum_l (S_{2n+1,l}^{p-1})^2 \tag{9}$$

2.2 Local Energy Analysis

After the above step, the original image is characterized by its local energy analysis. The local energy analysis is utilized for the aim of identifying areas in each subband. 2D Haar wavelet packet analysis is used to characterize the local energy values of pixels. Experiment shows that Haar wavelet packet analysis has not only rather good performance to characterize texture features but also its computation is efficient. In (8), we consider orientation selectivity in the tensor product of low-pass filter h and high-pass filter g. The combination of h and g are substituted another notations that are explicated in [9]. In each scale level, the image is decomposed into four subbands: LL, LH, HL and HH. LL means the horizontal low-frequency and vertical low-frequency of the image, LH the horizontal low-frequency and vertical high-frequency, HL the horizontal high-frequency and vertical low-frequency and HH the horizontal high-frequency and vertical high-frequency. The filter with particular direction covers certain region in the 2D special-frequency domain. Several wavelet decomposition filters are designed by summations, where combination of filters denotes the frequency sector of a certain direction and scale. We here adopt three high frequent subbands to extract local text / non-text features. Let $c_k = LH_k + HL_k + HH_k$ be the addition of high-frequent coefficients, where k = horizontal, vertical, diagonal, horizontal-diagonal, vertical-diagonal filters. These filter outputs make a measure of signal energies at different direction.

We have hired local standard deviation to estimate the local energy. It calculates overlapping windows around each pixel.

$$\sigma(x,y) = \sqrt{\frac{1}{W}\sum_{m=1}^{w}\sum_{n=1}^{w}| c_k(m,n)^2 - \bar{c}_k(x,y)^2)|} \tag{10}$$

Where the local energy is $\sigma(x,y)$ around the x,y th pixel and w is small window size and $W = w \times w$, $\bar{c}_k(x,y)$ is the mean around the x,y th pixel.

$$G(x,y) = \frac{1}{2\pi\sqrt{\sigma}} e^{-\frac{1}{2\sigma^2}}(x^2+y^2) \tag{11}$$

Where σ determines the band width of averaging window. Finally, we obtain the feature images corresponding to $c_k = LH_k + HL_k + HH_k$ through Gaussian low-pass filtering.

$$F(x,y) = \frac{1}{W}\sum_{(m,n)\in W} | G(c_k(m,n))| \tag{12}$$

We adapt the W size from 9×9 to 15×15 to be appropriate in most of our experiment images $(256\times 256, 512\times 512)$.

2.3 Unsupervised Classifier

The output of wavelet packet analysis is applied to the IAFC (Integrated Adaptive Fuzzy Clustering) [10, 11] based on unsupervised learning rules. This paper is two class segmentation problem. A supervised version of the classifier for document image segmentation is dependent on the knowledge of scale, scanning resolution, rotation, skewness, font size, type of layout, etc. of the document [12]. Whereas our aim has been to make the segmentation independent of all the aforesaid issues, so we need an unsupervised classifier. Also, to avoid the vagueness of text / non-text regions, we have hired fuzzy clustering algorithm. After feature vector is given as input, winner-take-all competition occurs among outputs.

$$I = {}_{i}^{\min} \left\| X - V_j \right\| \tag{13}$$

Where X is the feature vector, V_j is weight vector related to the $i-th$ output node. In next step, the vigilance test is performed according to the vigilance criterion.

$$e^{-\gamma u_i}\left\| X - V_i \right\| \le T \tag{14}$$

In (14), γ is the constant which controls the boundaries of classes. T means vigilance parameter.

$$u_i = \frac{\left[\frac{1}{\left\| X - V_i \right\|^2}\right]^{\frac{1}{m-1}}}{\sum_{j=1}^{c}\left[\frac{1}{\left\| X - V_j \right\|^2}\right]^{\frac{1}{m-1}}} \tag{15}$$

where c shows the number of clusters, m is constant $(m = 2)$. If the winning cluster satisfies the vigilance criterion, the centroids of all clusters are updated.

$$V_i(t+1) = V_i(t) + f(t)\times\lambda(X - V_i(t)) \quad \text{for all } i, \tag{16}$$

Where

$$f(t) = \frac{1}{k(t-1)+1}, \quad \lambda = u_i^2 \cdot \prod(X, V_i(t), T) \tag{17}$$

The updating equation is in (16), where k is constant and t is iteration. $\prod(X, V_i(t), T)$ is defined as follows.

$$\prod(X,V_i(t),T) = \begin{cases} 0 & \left\|X-V_i(t)\right\| \ge T \\ 1-2\left(\frac{\left\|X-V_i(t)\right\|}{T}\right)^2 & 0 \le \left\|X-V_i(t)\right\| < \frac{T}{2} \\ 2\left(1-\frac{\left\|X-V_i(t)\right\|}{T}\right)^2 & \frac{T}{2} \le \left\|X-V_i(t)\right\| < T \end{cases} \tag{18}$$

In this unsupervised form, the vigilance parameter T is important of controlling the number of classes. We consider only two classes corresponding to the text/non-text.

III. EXPERIMENTAL RESULTS

Several documents images are analyzed using our method, so as to demonstrate the performance of our algorithm. These document images are scanned from Korean newspapers and journals. In these experiments, we have taken into account those subbands which have the highest values of energies, meaning that these subbands would supply more information of text than the others.

We have tested 250 images, in which Korean texts appear in them. Most of texts are aligned horizontally but there are some aligned vertically. Some characters are laid over simple background. However, there are still a lot of characters laid over complex background. By experiments, we found that the contrast of characters with their local background extends a wide range in some images.

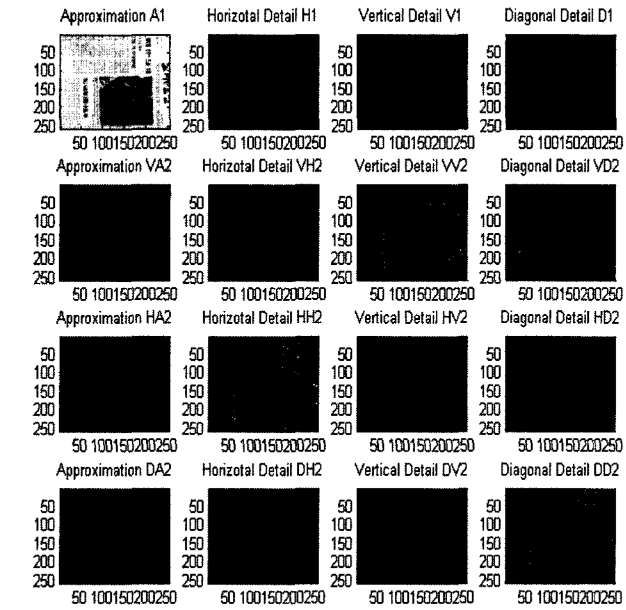
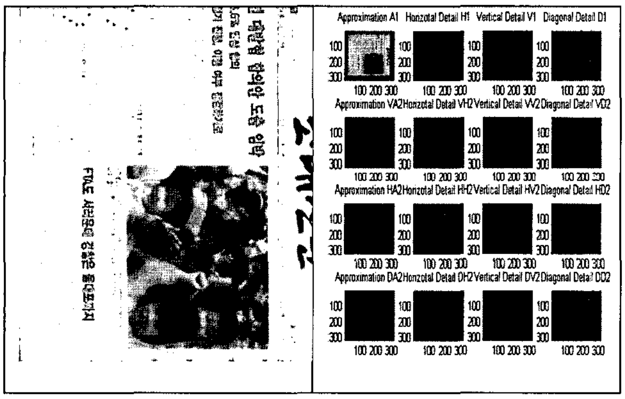
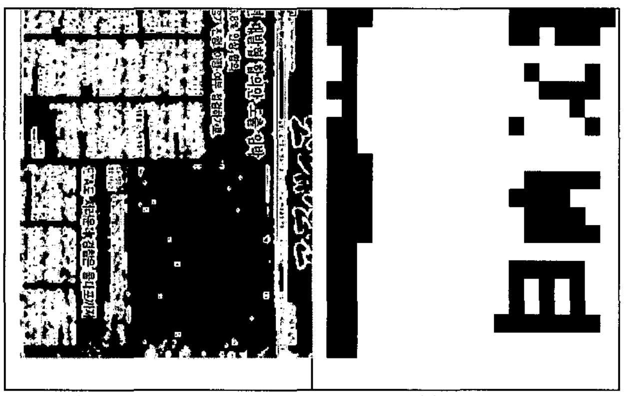


Fig. 1 Newspaper image in wavelet packet analysis.

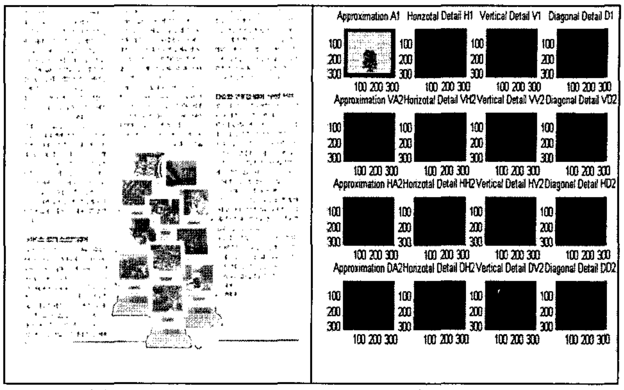


(a) (b)

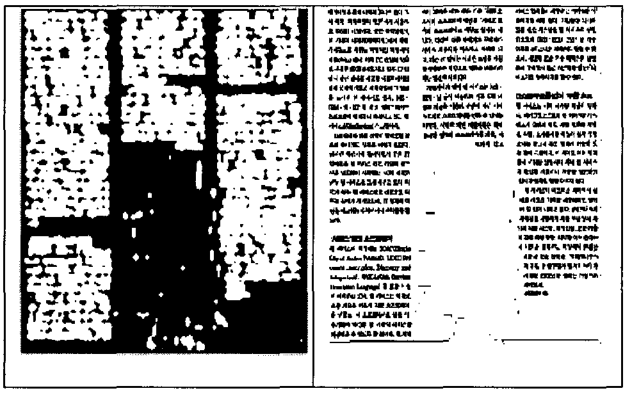


(c) (d)

Fig. 2 experiments on newspaper images.

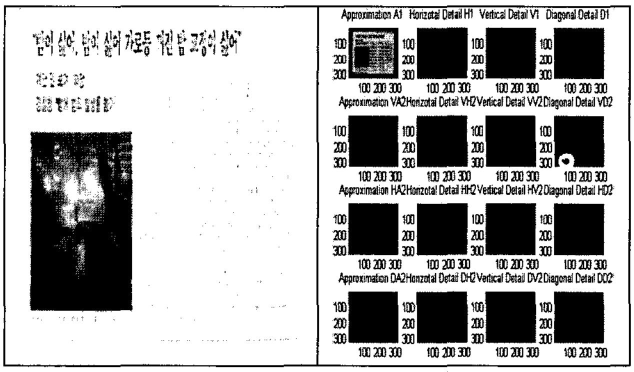


(a) (b)

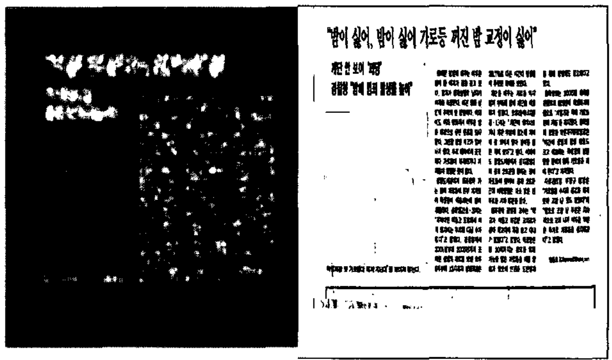


(c) (d)

Fig. 3 experiment on journal image.



(a) (b)



(c) (d)

Fig. 4 experiments on newspaper images.

Figure 1 demonstrates the set of equivalent frequency subbands associated with a wavelet packet transform for levels one and two. These filters output basically give a measure of signal energies at different directions and scales. In Figure 2 (a) is the original image from newspaper, (b) shows wavelet packet transform for level two. Gaussian smoothing filter image after local energy analysis in (c), finally (d) shows the result image through IAFC. In Figure 3, the journal image was tested. Figure 4 shows another newspaper image and the experimental result image.

IV. CONCLUSIONS

A document image segmentation method has been proposed in this paper, which used textural properties of document image to separate text /non-text regions. The segmentation method has used wavelet packet analysis and the IAFC. First we use wavelet packet which decompose an image into N by N subbands. These subbands represent the image at different scales and orientations in the frequency domain. We measure the local energy around each pixel at different scales. It creates the feature vector which is input of the improved IAFC algorithm. We achieve segmentation by simple wavelet decomposition with unsupervised clustering algorithm, IAFC. Through several experiments, the proposed method was useful for discriminating between text and non-text. The method can be adapted in the specific classification problem based on textural properties.

V. ACKNOWLEDGMENT

This study was supported in part by research funds from Chosun University, 2004.

REFERENCES

[1] S. N. Srihari, "Document Image Understanding," *Pro. IEEE Computer Society,* Fall Joint Computer Conf. pp. 87-96, 1986.

[2] S. Sural and P. K. Das, "A two step algorithm and its parallelization for the generation of minimum containing rectangles for document image segmentation," *in Proc. Int. Conf. Document Analysis and Recognition,* pp. 173-176, 1999.

[3] T. Pavlidis and J.Zhou, "Page segmentation using the description of the background," *Comput. Vis. Image Understanding,* Vol. 70, No. 3, pp.350-369, 1998.

[4] A. Antaonacopulos, "Page Segmentation by white streams," *in Proc. 10th. Int. Conf. Pattern Recognition,* pp. 945-953, 1991.

[5] A. Laine and J. Fan, "Texture Classification by Wavelet Packet Signatures," *IEEE Tran. on pattern Analysis and machine Intelligence,* pp. 1186-1191, 1993.

[6] H. Kim and P. Liang, "Target Extraction from Clutter Images Using Wavelet Packet Analysis," *IEEE National Radar Conference,* pp. 195-200, 1998.

[7] S. Mallat, "A Theory for Multiresolution signal decomposition: The Wavelet Representation," *IEEE Trans. Pattern Analysis, Machine Intell.,* Vol. 11, pp. 674-693, 1989.

[8] R. Bajcsy, "Computer description of textured surfaces," *in Proc. 3rd Int. joint Conf. Artificial Intell.,* Aug., pp. 572-579, 1973.

[9] Rafael C. Gonzalez and Richard E. Woods, Digital *Image Processing 2nd Edition,* Prentice Hall, 2002.

[10] G. B. Lee and Y. S. Kim, "A Fuzzy Contrast Enhancement Technique using the improved IAFC model," *Korea Fuzzy Logic and Intelligent Systems Society,* Vol. 11, NO. 9, pp. 777-781, 2001.

[11] Y. S. Kim, "An Unsupervised Neural Network Using Fuzzy Learning Rule," *IEEE International Fuzzy Systems Conference Proceedings,* Vol. 1, pp. 349-343, 1999.

[12] M. Acharyya and M. K. Kundu, " Document Image Segmentation Using Wavelet Scale-Space Features," *IEEE Trans. Circuits and Systems for Video Technology,* Vol. 12, No. 12, pp. 1117-1127, 2002.

Geum-Boon Lee
Member of KIMICS. Received M.S. degree in Computer Engineering from Daejeon University in 2002. She is currently working toward the ph.D. degree in department of computer engineering at Chosun University. Her research interests include the Pattern Recognition, Wavelet and Computer Vision.

Wilfred O. Odoyo
Member of KIMICS. Received B.S. degree in Computer Engineering from Chosun University in 2005. He is currently working toward the M.S. degree in department of computer engineering at Chosun University. His research interests are in the areas of Wavelet, Image Segmentation and Pattern Recognition.

Kuk-Se Kim
Received B.S. and M.S. degrees in Computer Engineering from Chosun University in 1999 and 2001, respectively. He is currently working toward the ph.D. degree in department of computer engineering at Chosun University. His research

interests include the Embedded System, Home Networking and Ubiquitous.

Beom-Joon Cho
Received B.S. and M.S. degrees in Electrical Engineering from Chosun University in 1980 and 1982, respectively. And his ph.D. degrees in Electrical Engineering from Hanyang University in 1988 and Electrical Engineering & Computer Science from KAIST in 2004. In 1980 he joined the faculty of Chosun University where he is currently a professor in the Department of Computer Engineering. His research interests include the Pattern Recognition, Neural Network, Character Recognition and Computer Vision.