# Document Image Segmentation Using Fuzzy Classifier and the Dual-Tree DWT

*Jamal Saeedi[1], Reza Safabakhsh[2], and Saeed Mozaffari[3]*

[1] Amirkabir University of Technology, Electrical Engineering Department, Tehran, Iran. jamal.saeedi@yahoo.com
[2] Amirkabir University of Technology, Computer Engineering Department, Tehran, Iran. safa@aut.ac.ir
[3] Semnan University, Electrical and Computer Department, Semnan, Iran. mozaffari@semnan.ac.ir

***Abstract* — In this paper, we propose a new method for textual areas extraction of an image using a fuzzy classifier and dual-tree discrete wavelet transform. We have extended our text extraction scheme for classification of document images into text, background, and picture components. Three class fuzzy classifiers and a morphological post-processing operation is used for this purpose. The proposed method shows better results compared to the previous wavelet-based document image segmentation algorithms.**

***Keywords* — document image segmentation, dual-tree discrete wavelet transform, fuzzy classifier, morphological post-processing.**

## I. Introduction

Document image segmentation partitions a document image into distinct regions. In the ideal case, these regions correspond to the image entities such as text and non-text blocks; text, picture, and background blocks. Image and text database retrieval, automated documents reading and processing, and digital documents storing are some applications of the document image segmentation.

Some sates of the art techniques for text–graphics segmentation are reviewed [1]. Top-down and bottom-up approaches have been used frequently for document image segmentation. *Top-down* approaches split the document into blocks, columns paragraphs, text lines, or even words. The most common top-down techniques are run-length smoothing [2], [3] and projection profiles [4]. These methods are not suitable for skewed texts because top-down methods are restricted only to rectangular blocks. On the other hand, *bottom-up* methods tend to group connected components of the same type [5], starting from the pixel level to obtain higher level descriptions of the document such as words, text lines, paragraphs etc. [6].

Recently, wavelet techniques have become powerful tools in document image analysis domain. It is used in many publications for feature extraction and document image segmentation. Li and Gray [7] have used distribution characteristics of wavelet coefficients segment document images. Kundu and Acharya [8] reported another scheme for text segmentation in the document images based on wavelet scale-space features. They used M-band wavelet, which decomposes an image into band-pass channels to easily detect the text regions. Then real text regions were recognized based on k-means clustering using feature vectors obtained from the M-band image. Geum-Boon Lee *et al*. [9] used an algorithm based on local energy estimation in wavelet packet domain and k-means classifier. In another research, a technique for segmentation of an image into printed text, handwritten text and noise has been proposed in [10] which uses Fisher classifier for the classification task. Sunil Kumar *et al*. [11] also used Fisher classifier and matched wavelet for document image segmentation into three classes (text, picture, and background).

In the proposed scheme we used new texture features extracted by the dual-tree discrete wavelet transform (DT-DWT). First, the input image is decomposed by DT-DWT analysis filter bank and then, a gradient image is produced by vanishing scaling sub-bands. The DT-DWT has more directional sub-bands per scale than the DWT, making it appropriate for generating a feature image which contains more directional information of the document image. The key point which distinguishes the proposed feature from other wavelet domain texture features [7]-[11], is the use of coarser level of information. Document images usually include multi-font regions. Using only information of one level of image decomposition in the feature image, text blocks with larger font will be missed in the output of the classification. With the proposed DT-DWT features we tackle multi-font problem.

For document image classification, we proposed a new scheme based on fuzzy logic. The proposed classifier seems to be more efficient and applicable. A post-processing stage is also used for the classification results refinement.

This paper is organized as follows: In Section II we briefly describe dual tree discrete wavelet transform. Section III gives the algorithm for text/non-text classification. In Section IV, the use of fuzzy classifier in locating text/ picture/ background in the document image is explained. Morphology-based post-processing is presented in Section V. Section VI presents various experimental results and performance comparisons. Finally, we conclude with a brief summary in section VII.

## II. The Dual-Tree DWT

The DT-DWT is a modified version of the DWT [12], [13]. The DT-DWT was proposed to overcome shift variance and directionality limitations of the DWT while maintaining the perfect reconstruction property with limited redundancy. The DT-DWT is basically two parallel DWT filter bank trees. The wavelet and scaling functions used in one tree can be defined as approximate Hilbert transforms of the functions in the other tree [14]. The filters used in both trees are real, but the combined filters are referred as analytic. This combination led to complex extension of real signals. Figure 1 shows a 1-D DT-DWT 3-Level Decomposition Filter Bank Tree.

As complex wavelets can distinguish between positive and negative frequencies, the diagonal sub-bands can be discriminated from horizontal and vertical sub-bands. Later on, horizontal and vertical sub-bands are divided giving six distinct sub-bands at each scale (at orientation $\pm 15^{\circ}, \pm 30^{\circ}$, and $\pm 75^{\circ}$).

These orientated and scale dependant sub-bands are shown spatially in Figure 2, demonstrates the improved directional selectivity of the DT-CWT.

## III. TEXT/NON-TEXT CLASSIFICATION

### A. Generation of gradient image

The first step in all other texture-based document image segmentation method is the texture extraction. In this paper we used gradient image as texture feature, which is obtained by image decomposition using DT-CWT. Figure 3 shows DT-CWT image decomposition using at one level.

There are two methods to generate a gradient image from the wavelet sub-bands. The first method uses the sum of squares of the corresponding coefficients in (high-frequency) detail sub-bands (e.g. in the DWT, LH and HL) to generate the gradient image [15]. The second

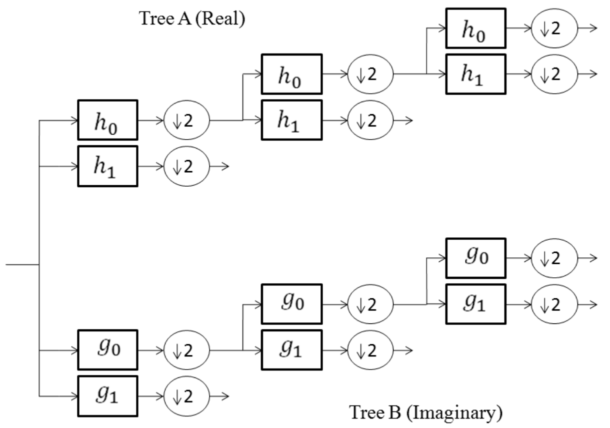



Figure 1: 1-D DT- DWT 3-Level Decomposition Filter Bank Tree

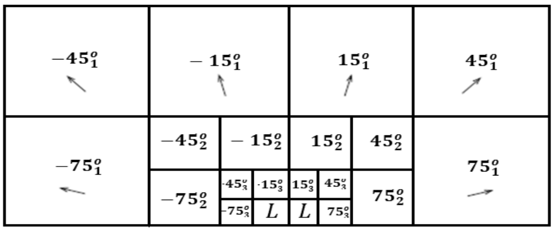



Figure 2: 2D DT- DWT, Scale and orientation labeled sub-bands.


Figure 3: An example of 2D DT-CWT.


Figure 4: The gradient image using vanishing scaling sub-bands.

method takes the inverse wavelet transform of wavelet sub-bands to generate a gradient image while the (low-frequency) scaling sub-bands are vanished [16]. The gradient image generated by the first method is shrunk (i.e. down sample by factor 2) but the second method is generating gradient image in the actual image size. Here, we use the second approach to generate the gradient image. Figure 4 shows the gradient image using the second method.

### B. Feature extraction

In order to enhance the gradient image information, we use two texture features. Texture analysis attempts to quantify intuitive qualities described by terms such as rough, smooth, silky or bumpy as a function of spatial variation in pixel intensities. The first feature calculates local entropy and the second one calculates local range of the gradient image. We calculate the two features using:

$$\text{Entropy: } F_1 = -\sum_{x,y \in W} |G(x,y)| \, . \, log(|G(x,y)|) \quad (1)$$

$$\text{Range: } F_2 = max_{x,y \in W}(|G(x,y)|) - min_{x,y \in W}(|G(x,y)|) \quad (2)$$

where G is the gradient image which can be negative in DT-DWT domain and w is the size of local window.

In the gradient image obtained in the previous section, text regions have large local variations, and the two texture features capture and magnify these variations making the input image appropriate for the classification step. In addition for improving the features a nonlinear averaging filter is used to reduce noise and taking into accounts neighbor dependency. This operation is implemented as follows:

$$NF_i = \frac{\sum_{x,y \in W(k,l), k \neq x, l \neq y} \mu(x,y) \times F_i(x,y)}{\sum_{x,y \in W(k,l), k \neq x, l \neq y} \mu(x,y)} \quad (3)$$

where $i = 1{,}2$ for two texture features, and $\mu(x, y)$ is calculated using:

$$\mu(x,y) = exp\left[-\frac{(x-k)^2 + (y-l)^2}{NW}\right] \quad (4)$$

where NW is the length of local window. For example $\mu(x, y)$ for a $5 \times 5$ window is:

$$\mu(x,y) = \begin{pmatrix} 0.73 & 0.82 & 0.85 & 0.82 & 0.73 \\ 0.82 & 0.92 & 0.96 & 0.92 & 0.82 \\ 0.85 & 0.96 & 0 & 0.96 & 0.85 \\ 0.82 & 0.92 & 0.92 & 0.92 & 0.82 \\ 0.73 & 0.82 & 0.85 & 0.82 & 0.73 \end{pmatrix} \quad (5)$$

The central weight is set to 0 to remove isolated noise effectively. Figure 5 shows two texture features for the given image.

### C. Classification

Having the texture features, we used fuzzy classifier to classify pixels as either text or non-text. The simplest fuzzy rule-based classifier is a fuzzy if-then system, similar to that used in fuzzy control [17]. This classifier can be constructed by specifying classification rules as linguistic rules:

1. IF $\mathrm{NF}_1$ is *large* AND $\mathrm{NF}_2$ is *large* THEN class is **text**
2. IF $\mathrm{NF}_1$ is *large* AND $\mathrm{NF}_2$ is *small* THEN class is **non-text**
3. IF $\mathrm{NF}_1$ is *medium* AND $\mathrm{NF}_2$ is *large* THEN class is **text**
4. IF $\mathrm{NF}_1$ is *medium* AND $\mathrm{NF}_2$ is *small* THEN class is **non-text**
5. IF $\mathrm{NF}_1$ is *small* THEN class is **non-text**

Each linguistic value is represented by a membership function. Figure 6 shows triangular membership functions for $\mathrm{NF}_1$, which is normalized and $T_1$ is a constant value. For the pair of values $(\mathrm{NF}_1, \mathrm{NF}_2)$, the degree of satisfaction of the antecedent part of the rule determines the *firing strength* of the rule. For example, the firing strength of rule 1 is calculated as:

$$\tau_1 = \mu^1_{large}(\mathrm{NF}_1) \text{ AND } \mu^2_{large}(\mathrm{NF}_2) \quad (6)$$

The AND operation is typically implemented as *minimum*, but any other t-norm may be used. Some well-known triangular norms are shown in Table 1. We have chosen algebraic product for the AND operation. Therefore *firing strengths* of the fuzzy rules are as follows:

(a) (b)

Figure 5: The two texture features. (a) Local range. (b) Local entropy.

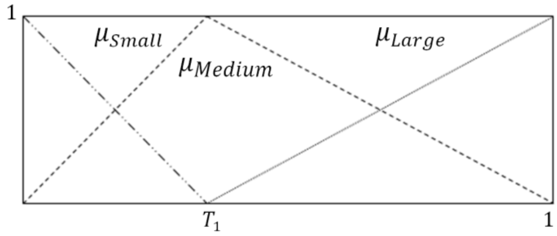



Figure 6: Fuzzy membership function for the linguistic terms of $\mathbf{NF_1}$

$$\tau_1 = \mu^1_{large}(NF_1) \times \mu^2_{large}(NF_2)$$
$$\tau_2 = \mu^1_{large}(NF_1) \times \mu^2_{small}(NF_2)$$
$$\tau_3 = \mu^1_{medium}(NF_1) \times \mu^2_{large}(NF_2)$$
$$\tau_4 = \mu^1_{medium}(NF_1) \times \mu^2_{small}(NF_2)$$
$$\tau_5 = \mu^1_{small}(NF_1)$$

The rules vote for the class of the consequent part (i.e. the fuzzy rules in resultant part vote for a class). The weight of this vote is $\tau_i$.To find the output of the classifier, the votes of all rules are aggregated. Among the variety of methods that can be applied for this aggregation, we considered the maximum aggregation method. Let k be the class labels (text, non-text), $j$ denote the number of rules, and $i \rightarrow k$ denote that rule $i$ votes for class $k$.Then:

$$\text{IF } \tau_i = \max_{j=1,\dots,5} \tau_j \quad \text{AND } i \rightarrow k \text{ THEN class } is\ k. \quad (7)$$

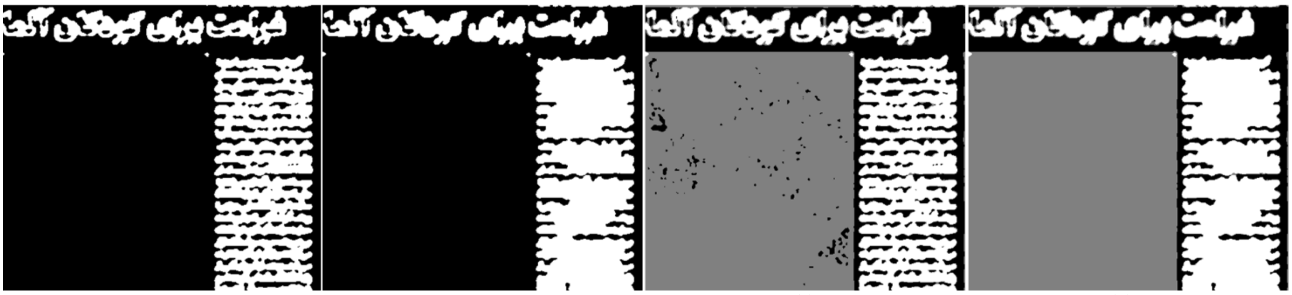
Also for building fuzzy membership function $T_1$ must be defined. We obtained $0.1 \le T_1 \le 0.2$ using test images and try and error.

## IV. TEXT/PICTURE/BACKGROUND CLASSIFICATION

In the previous section (section III), we explained the fuzzy classifier for text/non-text classification. In the document image, also, we consider three components: *text*, *picture*, and *background*. Backgrounds are continuous tone low frequency regions with monotonous features although mixed with noise. Images are continuous tone regions falling in between text and background. Thus, for document images we have extended our work described in the previous section (text/non-text classification) to segmentation of document images into three classes, *i.e.* text, picture, and background. We have used the same features and classified them into three classes. For classification, we have used a fuzzy classifier with three classes. This classifier can be constructed by specifying classification rules as linguistic rules similar to the two class problem mentioned before:

1. IF $\mathrm{NF}_1$ is *large* AND $\mathrm{NF}_2$ is *large* THEN class is **text**
2. IF $\mathrm{NF}_1$ is *large* AND $\mathrm{NF}_2$ is *small* THEN class is **picture**
3. IF $\mathrm{NF}_1$ is *medium* AND $\mathrm{NF}_2$ is *large* THEN class is **text**
4. IF $\mathrm{NF}_1$ is *medium* AND $\mathrm{NF}_2$ is *small* THEN class is **picture**
5. IF $\mathrm{NF}_1$ is small THEN class is **background**

After finding the *firing strength* of each rule, the output of classifier is obtained by equation (7).



(a) (b) (c) (d)

Figure 7: Classification output. (a), (c) before post-processing. (b), (d) after post-processing.

## V. POST-PROCESSING

After completion of classification, there may existed some misclassified pixels. In order to solve this problem, we have used a morphological reconstruction, namely *region filling* [18].

In the two-class problem, we set zero (black) to background and one (white) to text area. In result map when black regions surrounding by white regions, they change to white. For three-class problem, we set 0 to background, 0.5 to picture and 1 to the text areas. In this case we change those pixels which surrounding by lager value. For example if some pixels with 0 values surrounding by pixels with 0.5 values, they change to 0.5.

Figure 7 explore important refinement using post-processing of the classifier outputs.

## VI. EXPRIMENTAL RESULT

Several images with different properties have been used in this section to demonstrate the performance of the proposed algorithm. The scanned images are taken from different websites. Figure 9.1 and 2 are English pages while 3 and 4 are chosen from Farsi newspapers with sizes $(772 \times 842)$, $(944 \times 590)$, $(908 \times 798)$ and $(1910 \times 1378)$ respectively. Figure 9.1 is a single font, double column page with approximately the same text/picture ratio. Figure 9.2 is a multi-font, single column image with wide background. Figure 9.3 is also a multi-font, single column image with one large, smooth picture. Figure 9.4 has the same characteristics; however, its picture has multi objects.

First, we have discussed the experimental results of text/non-text classification. Then, results of text/picture/background document image segmentation are given. Finally, the results of post-processing algorithm are presented for both two and three class decomposition.

### A. Text/non-text classification results

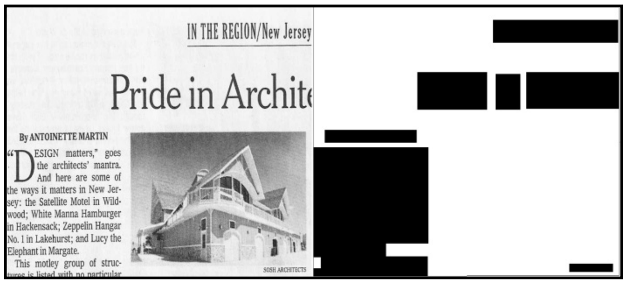
In this section, we have shown the results obtained by the three methods: K-means clustering [8], [9], fisher classifier using matched wavelet [11], and our proposed algorithm based on the fuzzy classifier. In the results, regions painted **black** indicate the text area and regions painted **white** are non-text areas. Figure 9 explore the results of three different classification methods.

We use some further criteria to evaluate these methods objectively. Text extraction is a one class classification problem [11]. The precision and recall accuracy of this classification is computed as:

$$\text{Recall rate} = \frac{\#\text{of text pixels correctly identified}}{\#\text{of text pixels in the ground truth}} \quad (8)$$

$$\text{Precision rate} = \frac{\#\text{of text pixels correctly identified}}{\#\text{of text pixels detected}} \quad (9)$$

Tables 2 and 3 tabulate the precision and recall rates of K-means, Fisher, and proposed methods respectively. This quantity has been evaluated against hand-picked ground-truth images. Figure 8 shows a sample image and its corresponding ground-truth image made by us.




(a) (b)

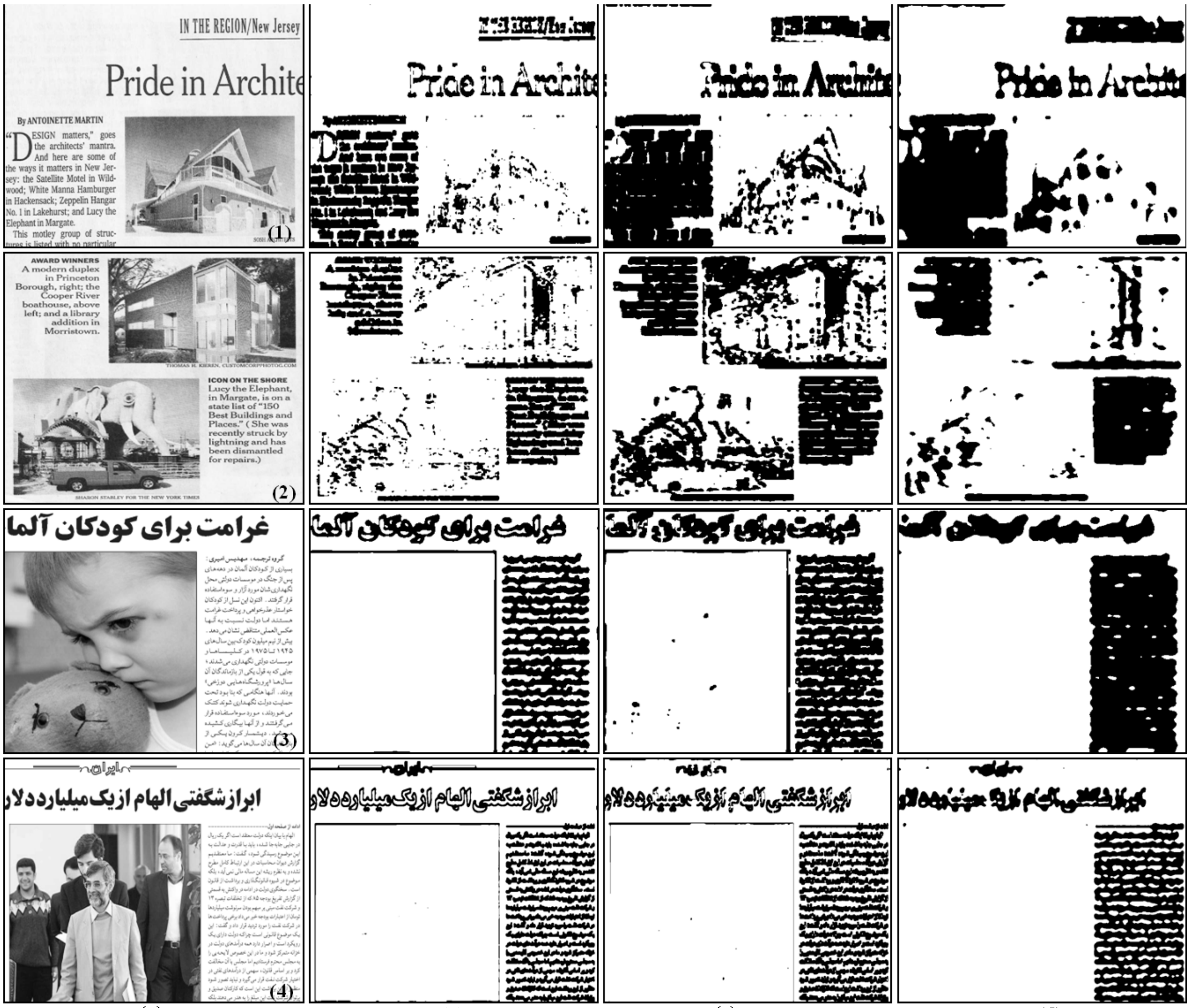
Figure 8: (a) Sample image; (b) its corresponding ground-truth.

Table 1: Some well-known triangular norms

| T-norms | |
|---|---|
| minimum | $min(x, y)$ |
| algebraic product | $x.y$ |
| weak | $\begin{cases} \min(x, y) & if \max(x, y) = 1 \\ 0 & otherwise \end{cases}$ |
| Bounded sum | $Max(0, x + y - 1)$ |

Table 2: Precision Rate obtained using three classification methods

| Image NO. | K-means | Fisher | Proposed |
|---|---|---|---|
| **Figure 9.1** | 64% | 42% | **70%** |
| **Figure 9.2** | **86%** | 79% | 79% |
| **Figure 9.3** | 92% | 88% | **98%** |
| **Figure 9.4** | 88% | 93% | **98%** |
| **Average** | 82.5% | 75.5% | **86.25%** |




(a) (b) (c) (d)

Figure 9: (a) The test document images in the experiment; (b)-(d) results using K-means clustering, Fisher classifier and the proposed fuzzy classifier.

Table 3: Recall Rate obtained using three classification methods

| Image NO. | K-means | Fisher | Proposed |
|---|---|---|---|
| **Figure 9.1** | 71% | 91% | **96%** |
| **Figure 9.2** | 57% | 81% | **82%** |
| **Figure 9.3** | 59% | 70% | **83%** |
| **Figure 9.4** | 46% | 52% | **64%** |
| **Average** | 58.25% | 73.5% | **81.25** |

## B. Text/picture/background classification results

In this section, we have shown the results of three-class document image segmentation. Since K-means is a two-class classifier, unlike the previous section, it is not used here. In the obtained results, regions in **black** indicate the text area; the regions in **gray** show the picture area and the regions in **white** are background areas. Figure 11 shows the results of Fisher classifier and the proposed fuzzy classification method.

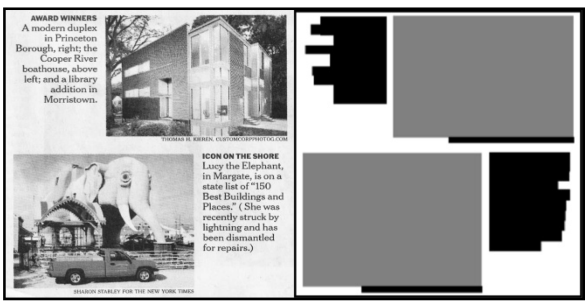
We performed other experiments to objectively evaluate these methods. This is a three class classification problem. Accuracy measurement is defined as follow:

$$\text{Accuracy} = \frac{\#\text{of pixels correctly identified}}{\#\text{of pixels in the ground truth}} \tag{10}$$

Table 4 shows the accuracy rates of the two methods. This quantity has been evaluated against hand-picked ground-truth images. Figure 10 shows a sample image and its corresponding ground-truth image generated manually.

## C. Post-processing results

Morphological post-processing step can reduce the noise and enhance the classification result. Figure 7 shows the results of morphological post-processing of a sample document image.

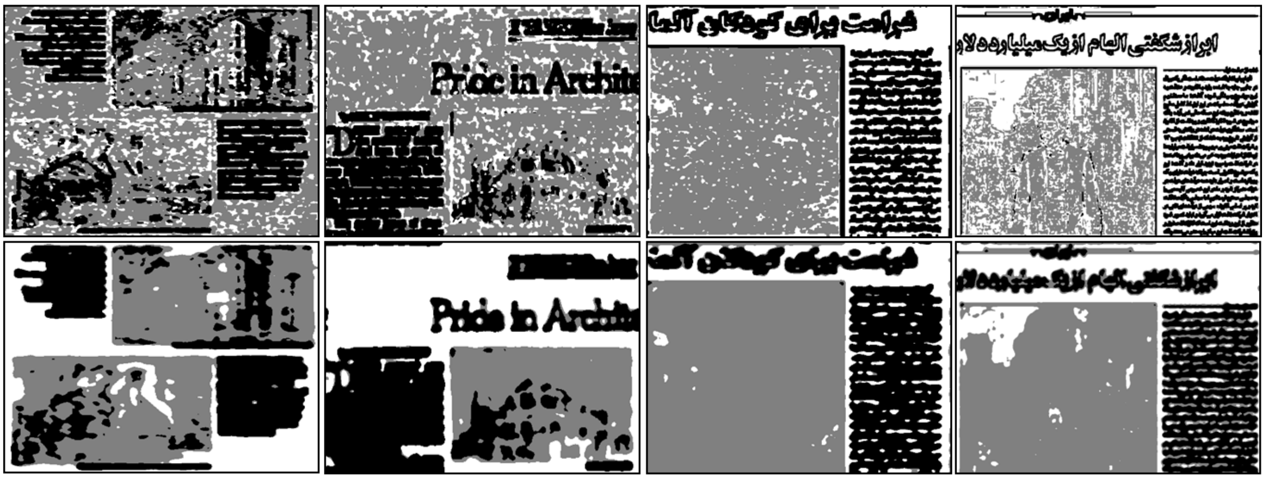


(a) (b)

Figure 10: (a) Sample image; (b) its corresponding ground-truth.

(a)

(b)



(1) (2) (3) (4)

Figure 11: (a) the results using fisher classifier; (b) the results using proposed method.

Also Table 5 explores the accuracy rates of the results before and after post-processing.

Table 4: accuracy Rate obtained using two classification methods

| Image NO. | Fisher | Proposed |
|---|---|---|
| **Figure 11.1** | 52% | **81%** |
| **Figure 11.2** | 51% | **84**% |
| **Figure 11.3** | 78% | **89%** |
| **Figure 11.4** | 60% | **80%** |
| **Average** | 60.25% | **83.5%** |

Table 5: accuracy Rate obtained before and after post-processing

| Image NO. | Before P-Proc | After P-Proc |
|---|---|---|
| **Figure 11.1** | 81% | **85%** |
| **Figure 11.2** | 84% | **85%** |
| **Figure 11.3** | 89% | **91%** |
| **Figure 11.4** | 80% | **86%** |
| **Average** | 83.5% | **86.75%** |

## VII. CONCLUSION

In this paper, we have presented a new algorithm for locating the text part based on textural attributes using the dual-tree DWT and fuzzy classifier. The proposed feature extraction method is simple, effective with low computational requirements. Compared to other existing methods [8], [9] and [11], feature dimensionality, and the computation of the feature space, is considerably reduced. We have applied our algorithm on several document images with different characteristics and complex backgrounds. The obtained results are promising. Considering the input image as a three-class set and classifying it into text/picture/background components is another paper's novelty.

## REFERENCES


[1] S. N. Srihari, "Document image understanding," *Proc. IEEE Computer Society Fall Joint Computer Conf.*, pp. 87–96, 1986.

[2] P. Chauvet, J. Lopez-Krahe, E. Taflin, and H. Maitre, "System for an intelligent office document analysis, recognition and description," *Signal Processing*, vol. 32, no. 1–2, pp. 161–190, 1993.

[3] F. Shih, S.-S. Chen, D. Hung, and P. Ng, "A document image segmentation, classification and recognition system," in *Proc. Int. Conf. Systems Integration*, pp. 258–267, 1992.

[4] M. Krishnamoorthy, G. Nagy, S. Seth, and M. Viswanathan, "Syntactic segmentation and labeling of digitized pages from technical journals," *IEEE Trans. Pattern Anal. Machine Intell*, vol. 15, no. 7, pp. 737–747, 1993.

[5] F. Lebourgeois, Z. Bublinski, and H. Emptoz, "A fast and efficient method for extracting text paragraphs and graphics from unconstrained documents," in Proc. Int. Conf. Pattern Recognition, pp. 272–276, 1992.

[6] C. L. Tan, B. Yuan, W. Huang, and Z. Zang, "Text/graphics separation using pyramid operations," in Proc. Int. Conf. Document Analysis and *Recognition*, pp. 169–172. 1999

[7] J. Li and R. M. Gray, "Context based multi-scale classification of document images using wavelet coefficient distribution," *IEEE Trans. Image Process.*, vol. 9, no. 9, pp. 1604–1616. 2000.

[8] Mausumi Acharyya and Malay K. Kundu, "Document Image Segmentation Using Wavelet Scale–Space Features," *IEEE trans. on Circuits and systems for video technology*, VOL. 12, NO. 12, pp. 1117-1127. 2002

[9] Geum-Boon Lee, Wilfred 0. Odoyo, Jae-Hoon Lee, "Two Texture Segmentation of Document Image Using Wavelet Packet Analysis," *ICACT2007*, pp 395-398. 2007

[10] Y. Zheng, H. Li, and D. Doermann, "Machine printed text and handwriting identification in noisy document images," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, no. 3, pp. 337–353, 2004.

[11] Sunil Kumar, Rajat Gupta, Nitin Khanna, , Santanu Chaudhury, and Shiv Dutt Joshi, "Text Extraction and Document Image Segmentation Using Matched Wavelets and MRF Model," *IEEE Trans. On Image Processing*, VOL. 16, NO. 8, , pp 2117-2128 AUGUST 2007

[12] Kingsbury, N.G., "A Dual-Tree Complex Wavelet Transform with Improved Orthogonality and Symmetry Properties," Proc. IE*EE Conf. on Image Processing*, 2000.

[13] Kingsbury, N.G., "Complex Wavelets for Shift Invariant Analysis and Filtering of Signals," *Journal of Applied and Computational Harmonic Analysis*, 10(3), pp. 234-253, 2001.

[14] Selesnick, I.W., "The Design of Approximate Hilbert Transform Pairs of Wavelet Bases," *IEEE Trans*. on Signal Processing, 50(5), pp.1144-1152, 2002.

[15] L. Feng, C.Y. Suen, Y.Y. Tang, L.H. Yang, Edge extraction of images by reconstruction using wavelet decomposition details at different resolution levels, *Int. J. Pattern Recognit. Artif. Intell*, pp. 779–793. 2000

[16] S. Mallat, S. Zhong, Characterization of signals from multiscale edges, *IEEE Trans. Pattern Anal Machine Intell*, pp. 710–732, 1992

[17] Kuncheva L.I., "Fuzzy Classifier Design," *Springer-Verlag*, Heidelberg, 2000.

[18] Rafael C. Gonzalez, Richard E. Woods, *digital image processing*, Tom Robbins, Upper Saddle River, New Jersey, 2002.