

# Document Analysis And Classification Based On Passing Window

Zaher A. Bamasood

Computer Science Department, Hadhramout University

Received (2018-03-04)

Accepted (2019-08-01)

**Abstract:** In this paper we present Document analysis and classification system to segment and classify contents of Arabic document images. This system includes preprocessing, document segmentation, feature extraction and document classification. A document image is enhanced in the preprocessing by removing noise, binarization, and detecting and correcting image skew. In document segmentation, an algorithm is proposed to segment a document image into homogenous regions. In document classification, Neural Network (Multilayer Perceptron- Back propagation) classifier is applied to classify each region to text or non text based on a number of features extracted in feature extraction. These features are collected from different other researchers' works. Experiments were conducted on 398 document images selected randomly from printed Arabic text database (PATDB) which was selected from various printing forms which are advertisements, book chapters, magazines, newspapers, letters and reports documents. As results, the proposed segmentation algorithm achieved only 0.814% as ratio of the overlapping areas of the merged zones to the total size of zones and 1.938% as the ratio of missed areas to total size of zones. The features, that show the best accuracy individually, are Background Vertical Run Length (RL) Mean, and Standard Deviation of foreground.

**Keywords:** Information Retrieval, Document Image Analysis, Segmentation, Feature Extraction, Data mining.

**How to cite this article:**

Zaher A. Bamasood. Document Analysis And Classification Based On Passing Window. J. ADV COMP ENG TECHNOL, 6(1) Winter 2020 : 39-46

## I. INTRODUCTION

Document analysis and classification play an important role in document image processing and its applications. Document analysis and classification usually select the main directions of the whole process of information conversion to digital form. For documents containing text and image regions, document analysis and classification groups the contents of the document into text and image regions. Then other systems may be applied such as an optical character recognition (OCR) system to recognize text, and an image processing system to process the images.

Document analysis and classification

systems usually include document segmentation, feature extraction and document classification. Document segmentation divides a document into homogeneous regions. Because each region has its own characteristics, a number of features are extracted for each region. Finally, document classification uses one or more classifiers to classify each region based on the extracted features into text, mathematical formulas, tables, figures, etc.

To segment document images into homogenous regions, several algorithms have been proposed. Wong et. al. presented the run-length smoothing algorithm (RLSA) which is a bottom-up approach Error! Reference source not found.. This algorithm consists of three steps. In the first step, horizontal white



This work is licensed under the Creative Commons Attribution 4.0 International Licence.

To view a copy of this licence, visit <https://creativecommons.org/licenses/by/4.0/>

runs that are smaller or equal to a threshold are changed to black runs. The second step is like the first but vertically not horizontally. The last step combines the results from the first and the second steps by the logical “and” operator. X-Y cut algorithm which is a top-down algorithm was presented in [1], [2]. In this algorithm, the page must be unskewed and the homogeneous blocks in the page can be bounded by rectangular regions, separated by white space or by horizontal and vertical lines. The document is cut alternatively horizontally and vertically according to white spaces based on projection profile of black pixels in the vertical and horizontal direction.

[3] proposed a hybrid algorithm. Its main idea is performing splits and merges of the regions of the document at the same time. If a region is inhomogeneous, split it into four rectangular subregions based on the projection profiles. If two adjacent regions are homogeneous And their union is also homogeneous, then merge them. Repeat the steps until no splitting and merging can be made.

To label each region of the documents as text, drawings etc. several features have been proposed previously.

[4] described a feature based supervised zone classifier using the information of the widths and heights of connected-components within a given zone. First of all, the bounding boxes of each connected component are selected. The height and width of the connected components are defined as the height and width of their bounding box. Then the histogram of the connected components' bounding boxes' widths and heights for each zone is computed. They observed that, each different zone usually has different distributions of connected component heights and widths. For example, a text zone has many connected components with the size of individual characters. The distribution of the widths and heights is encoded into  $n*m$  dimensional feature vector, where  $n$  and  $m$  are the number of intervals for the widths and heights respectively. The feature values are the normalized connected components' numbers corresponding to the different height and width intervals. As the authors' experiment, the values of  $n$  and  $m$  are 10 and 11 respectively. This system used a binary decision tree to classify each zone into one of eight classes (viz. text of font size 8-12, text of size 13-18, text of size 19-36,

math, table, line drawing, halftone, and ruling). They reported accuracy for text and non-text distinction greater than 97%.

In this paper we present Document analysis and classification system which is divided into four phases:

- 1) preprocessing,
- 2) document segmentation,
- 3) feature extraction, and

4) document classification. Preprocessing is used to enhance the images by removing noise, correcting any image skew, etc. Document segmentation phase segments document image into homogeneous regions. In feature extraction, features are extracted for each region. Based on these features, we classify the regions in the document into text and non-text regions The remainder of the paper is organized as follows: Section (2) focuses on the proposed system Section (3) emphasizes on experimental results

## II. PROPOSED SYSTEM

The document analysis and classification system has four components: preprocessing, document segmentation, text feature extraction, and classification. A document image is enhanced in the preprocessing phase by removing noise, binarization, and detecting and correcting image skew. The image is segmented into homogeneous regions in the document segmentation phase. Then each region is classified into text or non text in the classification phase based on a number of features that are extracted from each region in the feature extraction phase.

In some literature works, regions are classified into text, math, table, halftone, map/drawing, ruling, logo, etc. In our research, we represent the text and math classes as text classes, while we represent the other classes as non-text classes

### 1. Preprocessing

Preprocessing is necessary to improve document images before the segmentation phase. The preprocessing includes binarization, noise removal, and skew detection and correction.

The Otsu algorithm [5] is used to identify the threshold, used to convert a gray-level image into binary image. So, each pixel on the image is

transformed into white (foreground) if its gray level is smaller than the threshold, otherwise it set to black (background).

For noise removal, the Statistical Based Smoothing algorithm [6] is employed.

### 2. Document Segmentation

We proposed an algorithm to partition a document into homogeneous regions. The algorithm consists of three steps:

**Step1:** Scan the image by an  $n \times n$  window. Create a binary matrix 'C' depending in the following rules (where each cell represents the  $n \times n$  window in the image as shown in Figure 1).

1. If all the pixels of the window are black (foreground), assign 0 to the corresponding cell of the matrix
2. If all the pixels of the window are white (background), assign 1 to the corresponding cell of the matrix
3. If the window has mixed (white and black) pixels, then
  - a. If the black pixels in the window are less than Threshold, assign 1 to the corresponding cell of the matrix
  - b. Otherwise assign 0.
- The matrix 'C' represents the scaled image

**Step2:** Find boundaries of the connected components in the scaled image. Change each black pixels into white pixels except the border pixels and store the result in a new image called "perimeters". Figure 2.(c) shows the result of this step.

**Step3:** Each connected component on "perimeters" is assigned as a region if it is not internal component. If a connected component is an internal component, merge it with its container component.

Figure 2 shows some samples, which are fed to the algorithm

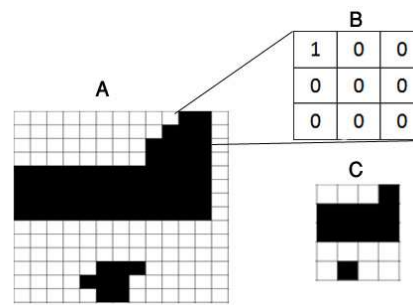


Fig.1: an example of rescaling the images with 3\*3 window

### 3. Feature Extraction

The following features are extracted from each extracted region.

#### 3.1. Standard deviation of the foreground pixels

$$\sigma = \sqrt{\left( \frac{1}{(n-1)} \sum_{i=1}^n (x_i - \bar{x})^2 \right)} \quad (1)$$

#### 3.2. The number of connected components

This feature is the number of connected components in the zone divided by the zone's area. Where zone's area is the number of foreground pixels. The feature is taken from [7].

#### 1.1. The Foreground / Background Means:

The ratio of the foreground/ background pixels to the total number of pixels in each region. This feature is divided into two features which are foreground mean and background mean

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \quad (2)$$

Where  $n$  are the number of elements. If foreground pixel represents '1',  $\bar{x}$  is the mean of the foreground pixels otherwise it is the mean of the background pixels.

The text regions, which have multiple text lines and multiple words, has fairly comparable number of foreground or background pixels. While in non-text region, the number of foreground pixels are always more than the number of background pixels or vice versa



Fig.2:the proposed algorithm for segmentation; (a) a sample image; (b), and (c)are the result images after applying the first and second steps, and (d) shows the boundaries of the resulting regions

1.2. Aspect Ratio

The ratio of height to width of each connected components is calculated in a zone. The average of the aspect ratio, which is used in [7], is taken as a feature.

The height and width of the connected components in some non-text regions have big difference, while they are possible to be close in most text regions.

1.3. Circularity

The circularity, which is taken from [7], is the square of the perimeter of the zone divided by the average of the areas of all connected components.

1.4. The Means of The Horizontal and Vertical Projections:

This feature is divided into two features which are Means of The Horizontal and Means Vertical Projections.

1.5. Background features:

The background features are based on background pixels. These features, which are taken from [8], follow:

1.5.1. Total area of large horizontal blank blocks.

A horizontal blank block is a large horizontal blank block if it satisfies the following rules:

1) Its number columns are large enough compared with the current zone. Specifically

$$\frac{b_c}{Col} > \theta \tag{3}$$

where  $b_c$  is the number of columns of the horizontal blank blocks,  $Col$  is the number of columns in the current zone, and  $\theta$  is 0.1 based on [8]’s experiments.

2) It does not touch left or right sides of the zone bounding box.

1.5.2. Total area of large vertical blank blocks.

A vertical blank block is a large vertical blank block if it satisfies the following rules:

1) Its number of rows are large enough compared with the current zone. Specifically

$$\frac{b_r}{r_w} > \theta \tag{4}$$

where  $b_r$  is the number of rows of the vertical blank blocks,  $rw$  is the number of rows in the current zone, and  $\theta$  is 0.1.

2) It does not touch the upper or bottom sides of the zone bounding box.

### 1.6. Run Length (RL) Features

16 run length features [8] include foreground/background run length mean and variance in four directions (viz. the horizontal, vertical, left-diagonal, and right diagonal directions) as shown in Figure 3.

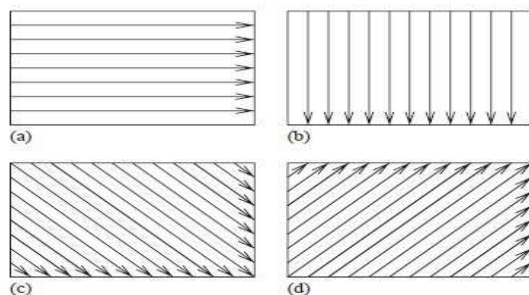


Fig.3: Illustrates the four directions. (a) Horizontal; (b) vertical; (c) left-diagonal; (d) right-diagonal

## III. EXPERIMENTAL RESULTS

### 1. Data and Tools

MATLAB is used to implement and test the prototype of this research. In the document analysis part, we used 398 document images selected randomly from printed Arabic text database (PATDB) which was presented in [9], [10], and [11]. PATDB database consists of 6954 document images selected from various printing forms (viz. advertisements, book chapters, magazines, newspapers, letters and reports). These images are stored in three different formats:

1. Black & white (binary) format with color depth of 1-bit per pixel;
2. Grayscale format with color depth of 8-bit (1-byte) per pixel (0 to 255 gray levels); and
3. Color (or RGB) format with color depth of 24-bit (3-byte) per pixel.

Each format was scanned with 200,300, and 600 dpi resolutions.

The database of 398 images was partitioned randomly into 120 images for testing, and 278 images for training. For validation, 55 images are selected randomly from the training set.

### 2. Evaluation Criteria

The evaluation criteria of page segmentation phase is different from the evaluation criteria of zone classification. For that, this research has the following types of evaluation criteria:

#### 2.1. Page Segmentation Evaluation Criteria

To evaluate a page segmentation algorithm, there are a number of measures that can be used [12]. In this work, two types of region-based error measures are defined.

##### 2.1.1. Merged Zones:

A merged zone error is a segmented zone that includes two or more zones with different ground truth values. Figure 4 shows an example of a merged zones' error. We find the overlapping areas of the merged zones (OAMZ) in our estimation. The overlapping areas of the merged zones (OAMZ) is the ratio of the overlapping areas of the merged zones (OAMZ) to the total zones.

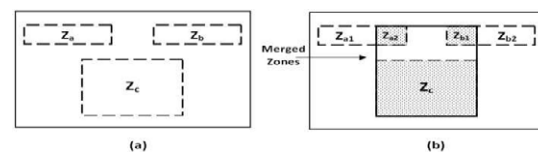


Fig. 4: Merged Zones Error Measure; (a) the ground truth zones where the ground truth  $Z_a$  and  $Z_b$  are text zones and  $Z_c$  is non text. (b) a merged zone error; the shaded rectangle denotes segmented merged zone.

##### 2.1.2. Missed Zones

The missed zones are the zones that did not match any foreground zones in the hypothesized segmentation. Some zones are partially missed. So, only the missed area of those zones is counted. The missed zones is the ratio of missed areas on the zones to total size of zones. Figure 5 shows missed zone error.



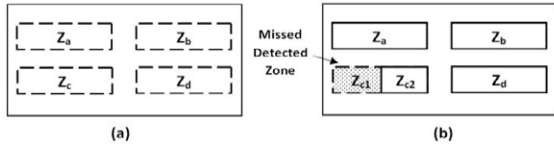


Fig.5: Missed Zones Error Measure (a) the ground truth. (b) a missed zone error; solid-line rectangles show the segmented zones while the shaded area represents the missed zone

### 2.2. Zone Classification Evaluation Criteria

Different types of metrics have been used for the performance evaluation of document analysis and classification in the classification phase. These metrics are defined below:

#### 2.2.1. Text recognized as Text (TRT)

The ratio of the total area of text zones in the ground truth which is recognize as text.

#### 2.2.2. Text recognized as Non Text (TRNT)

The ratio of the total area of text zones in the ground truth which is recognize as non text.

#### 2.2.3. Non Text recognized as Non Text (NTRNT)

The ratio of the total area of non text zones in the ground truth which is recognize as non text.

#### 2.2.4. Non Text recognized as Text (NTRT)

The ratio of the total area of non text zones in the ground truth which is recognize as text.

#### 2.2.5. Percentage Accuracy

$$Acc = \frac{TRT+NTRNT}{total\_area\_of\_all\_zones} * 100 \quad (5)$$

#### 2.2.6. Percentage of Error

$$Acc = \frac{TRNT+NTRT}{total\_area\_of\_all\_zones} * 100 \quad (6)$$

### 3. Experimental work

In this section we present our experimentation results of the proposed document classification technique, XY cut, and RLSA methods. Each method was tested for segmentation, and zone classification.

#### 3.1. Document Segmentation

Based on the performance measures defined in section 2.1, we evaluated the performance of the three algorithms for page segmentation.

Table 2 shows the performance of the proposed, XY cut, and RLSA algorithms. The proposed algorithm has the best performance in merged zone error measure. XY cut shows the best performance in missed zones measure. Our proposed algorithm gives the worst results in the missed zones measure. This is due to one of the steps, which is rescaling the image, leading to loss of some pixels. Rescaling the image step divides the image into n\*n window, the pixels of the window are removed if the number of foreground pixels in the window are smaller than or equal to three.

#### 3.2. Zone Classification

The classifier, which was used to classify each region into text and non text, is Neural Network (Multilayer Perceptron- Back propagation). The Neural Network contains one hidden layer of six neurons. The transfer functions of the hidden layer and output layer are is the Log-sigmoid and Linear transfer functions respectively.

Twenty six features were extracted from each region. We evaluated each feature independently. The initial information of zone classification (viz. number of pages, number of zones, and etc.) are presented in Table 3. Table 1 shows the results of each feature in the proposed, XY cut, and RLSA approaches respectively using the Neural network classifier. Some features have better accuracy in some approaches. For example the “the Mean of Horizontal Projection” feature has good accuracy for XY cut and low accuracy for the proposed and RLSA approaches. The best feature in the proposed approach is ‘Background Vertical Run Length Mean’. In the case of XY cut approach the best accuracy is achieved with ‘Foreground Right-diagonal Run Length Variance’. While the ‘Foreground Left-diagonal Run Length Mean’ feature is the best for the RLSA approach. As shown in Table 1, the features, whose accuracy are zero in the proposed approach, are the circularity, the foreground horizontal run length variance, the foreground left-diagonal run length variance, and the foreground right-diagonal run length variance.

**TABLE 1**  
**RESULTS OF EACH FEATURE OF THE PROPOSED APPROACH USING THE NN CLASSIFIER. THE FEATURES ARE RANKED (BEST AT TOP)**

Feature Name	Feature No	Correct Classification				Misclassification				Acc. %	Err. %
		Text / Text		Non Text / Non Text		Text / Non Text		Non Text / Text			
		Area	%	Area	%	Area	%	Area	%		
Back. Ver. RL Mean	Feat.15	8139196	95.123	7182201	96.022	417274	4.877	297579	3.978	95.542	4.458
Stand. Dev. of fore.	Feat.1	8112140	94.807	7189524	96.119	444330	5.193	290256	3.881	95.419	4.581
Back. Left-diag. RL Variance	Feat.20	8326327	97.310	6971037	93.198	230143	2.690	508743	6.802	95.392	4.608
Foreground Mean	Feat.3	8051506	94.098	7245188	96.864	504964	5.902	234592	3.136	95.388	4.612
Back. Hor. RL Mean	Feat.11	8458313	98.853	6831637	91.335	98157	1.147	648143	8.665	95.346	4.654
Background Mean	Feat.4	8040136	93.966	7247653	96.897	516334	6.034	232127	3.103	95.333	4.667
Back. Right-diag. RL Variance	Feat.24	8171802	95.504	7087345	94.753	384668	4.496	392435	5.247	95.154	4.846
Back. Hor. RL Variance	Feat.12	8458313	98.853	6799073	90.899	98157	1.147	680707	9.101	95.143	4.857
Back. Ver. RL Variance	Feat.16	8445990	98.709	6710420	89.714	110480	1.291	769360	10.286	94.513	5.487
Back. Right-diag. RL Mean	Feat.23	7669156	89.630	7262536	97.096	887314	10.370	217244	2.904	93.112	6.888
Back. Left-diag. RL Mean	Feat.19	7644971	89.347	7274386	97.254	911499	10.653	205394	2.746	93.035	6.965
# of Con. Comp.	Feat.2	8488069	99.201	5632926	75.309	68401	0.799	1846854	24.691	88.057	11.943
Fore. Hor. RL Mean	Feat.9	6055136	70.767	5865490	78.418	2501334	29.233	1614290	21.582	74.335	25.665
Aspect Ratio	Feat.5	3874195	45.278	7246911	96.887	4682275	54.722	232869	3.113	69.350	30.650
Mean of Ver. Proj.	Feat.8	4602532	53.790	6417562	85.799	3953938	46.210	1062218	14.201	68.720	31.280
Fore. Right-diag. RL Mean	Feat.21	8100953	94.676	1977607	26.439	455517	5.324	5502173	73.561	62.849	37.151
Fore. Left-diag. RL Mean	Feat.17	8204950	95.892	1760999	23.543	351520	4.108	5718781	76.457	62.146	37.854
Area of Large Hor. Blank Blocks	Feat.25	7513995	87.817	2004251	26.796	1042475	12.183	5475529	73.204	59.355	40.645
Area of Large Ver. Blank Blocks	Feat.26	6332385	74.007	2801905	37.460	2224085	25.993	4677875	62.540	56.960	43.040
Mean of Hor. Proj.	Feat.7	8134241	95.065	303158	4.053	422229	4.935	7176622	95.947	52.615	47.385
Fore. Ver. RL Variance	Feat.14	236	0.003	7477624	99.971	8556234	99.997	2156	0.029	46.631	53.369
Fore. Ver. RL Mean	Feat.13	2260066	26.414	4946944	66.138	6296404	73.586	2532836	33.862	44.942	55.058
Circularity	Feat.6	0	0.000	7479780	100.00	8556470	100	0	0.000	46.643	53.357
Fore. Hor. RL Variance	Feat.10	0	0.000	7479780	100.00	8556470	100	0	0.000	46.643	53.357
Fore. Left-diag. RL Variance	Feat.18	0	0.000	7479780	100.00	8556470	100	0	0.000	46.643	53.357
Fore. Right-diag. RL Variance	Feat.22	0	0.000	7479780	100.00	8556470	100	0	0.000	46.643	53.357

**TABLE 2**  
**COMPARING THE PAGE SEGMENTATION ALGORITHMS**

Page Segmentation Algorithms	Number of segmented zones	Merged Zones		Missed Detected Zones	
		Number Of Merged Zones	Ratio of the overlapping areas of the merged zones to the total size of zones	Number Of Missed Zones	the ratio of missed areas to total size of zones
Proposed	47543	276	0.814%	1557	1.938%
XY cut	5846	279	1.112%	566	1.444%
RLSA	38744	424	1.466%	1220	1.776%

**TABLE 3**  
**THE INITIAL INFORMATION OF ZONE CLASSIFICATION**

Attributes	Approaches			Ground Truth
	Proposed	XY Cut	RLSA	
Number of Pages	398	398	398	398
Number of Train Pages	278	278	278	278
Number of Test Pages	120	120	120	120
Total Num. of Zones	47543	5846	38744	6074
Num. of Text Zones	33889	5061	28285	5111
Num. of Non Text Zones	13378	506	10035	963
Num. of Merged Zones	276	279	424	--
Num. of Text Train Zones	22593	3543	18857	3570
Num. of Non Text Train Zones	8919	354	6690	661
Num. of Non Text Test Zones	4459	152	3345	302
Num. of Text Test Zones	11296	1518	9428	1541

#### IV. CONCLUSION AND FUTURE WORK

In Document segmentation, a proposed segmentation algorithm was presented to segment documents into homogenous regions. The proposed segmentation algorithm consists of three steps, namely rescaling the image, finding the boundaries of foreground pixels, and assigning regions. In rescaling the image, the image is divided into a number of  $n \times n$  pixel windows to produce a scaled image. In the scaled image, each window is represented as one background pixel (white) if the number of black pixels in the window are less than a threshold otherwise as foreground (black). In finding the boundaries of foreground pixels, the boundaries of each connected component of the scaled image are found, then each connected component is assigned as a region. To evaluate the proposed algorithm, we used two types of region-based error measurements: merged and missed zones errors. In merged zones errors measure, the proposed algorithm has the best performance compared to XY cut and RLSA implemented algorithms. On the other hand, our proposed algorithm gives the worst performance in the missed zones measure as the rescaling steps that lead to loss of some pixels.

In document classification, some of the features are used by other researchers.

The features, that show the best accuracy individually, are Background Vertical Run Length (RL) Mean, Standard Deviation of foreground, Background Left-diagonal RL Variance, Foreground Mean, Background Horizontal RL Mean, Background Mean, Background Right-diagonal. RL Variance, Background Horizontal. RL Variance, Background Vertical RL Variance, Background Right-diagonal RL Mean, and Background Left-diagonal RL Mean

Some directions for further improving the performance of our system can be listed as follows:

1) The missed zone errors in the proposed segmentation algorithm need to be reduced. This error is due to missing some pixels in the rescaling the image step. As a future work, updating and enhancing the rules of the rescaled images may help in reducing most of the missed zone errors.

2) Extend the work for labelling text regions

to title, abstract, footnote, caption or references.

3) Extend the work to label the non text regions to logos, forms, or etc.

4) Applying features selection techniques to select only the features that may provide high accuracy.

5) Using extra classifiers in zone classification.

#### REFERENCES

1. H. Jaekyu, R. M. Haralick, and I. T. Phillips, "Recursive X-Y cut using bounding boxes of connected components," in Proceedings of the Third International Conference on Document Analysis and Recognition (Volume 2) - Volume 2 (ICDAR '95), 1995, vol. 2, pp. 952-955.
2. G. Nagy, S. Seth, and M. Viswanathan, "A prototype document image analysis system for technical journals," Computer (Long. Beach. Calif.), vol. 25, no. 7, pp. 10-22, Jul. 1992.
3. J. Liu, Y. Y. Tang, and C. Y. Suen, "Chinese document layout analysis based on adaptive split-and-merge and qualitative spatial reasoning," Pattern Recognit., vol. 30, no. 8, pp. 1265-1278, Aug. 1997.
4. J. Liang, I. T. Phillips, J. Ha, and R. M. Haralick, "Document Zone Classification Using Sizes of Connected-components," in Document Recognition III, SPIE'96, 1996, pp. 150-157.
5. N. Otsu, "A Threshold Selection Method from Gray-Level Histograms," IEEE Trans. Syst. Man Cybern., vol. 9, no. 1, pp. 62-66, 1979.
6. S. A. Mahmoud, "Pergamon Arabic Character Recognition Using Fourier Descriptors and Character Contour Encoding," Pattern Recognition, vol. 27, no. 6, pp. 815-824, 1994.
7. S. Inglis and I. H. Witten, "Document Zone Classification Using Machine Learning," Proc. Digit. Image Comput. Tech. Appl., 1995.
8. Y. Wang, I. T. Phillips, and R. M. Haralick, "Document zone content classification and its performance evaluation," Pattern Recognit., vol. 39, no. 1, pp. 57-73, Jan. 2006.
9. [9] A. G. AL-Hashim, "Arabic database for automatic printed Arabic text recognition research and benchmarking," MSc Thesis. KFUPM, Dhaharan, Saudi Arabia, 2009.
10. A. G. Al-Hashim and S. A. Mahmoud, "Benchmark Database and GUI Environment for Printed Arabic Text Recognition Research," WSEAS Trans. Inf. Sci. Appl., vol. 7, no. 4, pp. 587-597, 2010.
11. A. G. Al-Hashim and S. A. Mahmoud, "Printed Arabic Text Database ( PATDB ) for Research and Benchmarking," in Proceedings of the 9th WSEAS international conference on Applications of Computer Engineering, 2010, pp. 62-68.
12. F. Shafait, D. Keysers, and T. M. Breuel, "Performance Evaluation and Benchmarking of Six-Page Segmentation Algorithms," IEEE Trans. Pattern Anal. Mach. Intell., vol. 30, no. 6, pp. 941-954, 2008.