Skew Detection/Correction and Local Minima/Maxima Techniques for Extracting a New Arabic Benchmark Database

Husam Ahmed Al Hamad
Department of Information Technology
Qassim University
Qassim, Saudi Arabia

***Abstract*—We propose a set of techniques for extracting a new standard benchmark database for Arabic handwritten scripts. Thresholding, filtering, and skew detection/correction techniques are developed as a pre-processing step of the database. Local minima and maxima using horizontal and vertical histogram are implemented for extracting the script elements of the database. Elements of the database contain pages, paragraphs, lines, and characters. The database divides into two major parts. The first part represents the original elements without modifications; the second part represents the elements after applying the proposed techniques. The final database has collected, extracted, validated, and saved. All techniques are tested for extracting and validating the elements. In this respect, ACDAR proposes a first issue of the Arabic benchmark databases. In addition, the paper confirms establishment a specialized research-oriented center refers to learning, teaching, and collaboration activities. This center is called "Arabic Center for Document Analysis and Recognition (ACDAR)" which is similar to other centers developed for other languages such as English.**

Keywords—ACDAR; Arabic benchmark database; Arabic scripts; document analysis; handwriting recognition; skew detection and correction

I. INTRODUCTION

Arabic language is spoken by hundreds of millions of people around the world. It profoundly influenced many cultures, including the Western culture, for many centuries. Although it is one of the most important languages in the world throughout its long history, it still lags behind many other languages as far as information technology resources and applications are concerned. As a result, the so-called "digital gap" is greater for Arabic language than other languages such as English, for instance.

Automatic recognition of handwritten words remains a challenging task even though the latest improvements of recognition techniques and systems are very promising. The term handwriting refers to some artificial graphical marks containing a message in a given human language [1]. The concept of handwriting has always existed, for the purpose of expanding people's memory and facilitating communication together [2] and much of the human culture may be attributed to the advent of handwriting. Because of the fact that only humans can perfectly understand and recognize the handwritings of others, one computationally challenging task resides in the attempt to imitate the human ability to read and recognize handwriting [3]. Consequently, automatic recognition of handwritten words remains a difficult task even though the latest improvements of recognition techniques and systems seem to be promising. For the purpose of automating Arabic scripts processing, numerous contributions have made in the area of handwritten script segmentation and recognition [4]. However, no outstanding results were reached so far, as OCR Arabic processing is still facing serious issues. One of the reasons is that Arabic language is considerably harder than Latin counterpart [5]. Therefore, in the area of automatic recognition of Arabic handwriting, many works have still to be done. One of the most important requirements for the development and comparison of recognition systems is a large database together with ground truth information. Compared to Latin scripts where handwritten words and numbers have publicly available for a long time (e.g. CEDAR [6], NIST[1]) the situation for Arabic is quite different. Others implement large databases that are not available to the public [7], or unreliable databases that concern only one Arab country (e.g. IFN/ENIT [8]).

Although many research efforts have done, so far in the recognition of handwritten Arabic script [4] until now they have not reached satisfactory results for the following reasons [2].

- Arabic words are overlapped and written always cursively, i.e., more than one character can be written connected to each other.
- Arabic writing uses many types of external objects, such as 'dots', 'Hamza', 'Madd', and diacritic objects, these external objects make the task of line separation and segmentation scripts more difficult.
- An Arabic character can have more than one shape according to its position in the word, i.e., initial, middle, final, or as a standalone character.
- Arabic writing uses many ligatures, especially in handwritten text.
- Other characters have very similar contours and are difficult to segment and to recognize especially when non-characters and external objects are present in the

[1] NIST database, http://www.nist.gov/srd/

scanned image.

II. HISTORICAL BACKGROUND

Earlier surveys discussed recognition and segmentation of both handwriting and machine-print, with much emphasis on machine-print. Unfortunately, only a small and unreliable database is available for Arabic Language today. However, in 1980, Nouh et al. suggested a standard Arabic character set to facilitate computer processing [9]. Standard and reliable databases were developed many years ago for the recognition of handwriting in Latin scripts. Among these databases, the CEDAR database (Center of Excellence for Document Analysis and Recognition) was released in 1993 [6]. It contains images of approximately 50,000 alphanumeric characters, 5,000 city names, 5,000 state names, and 10,000 ZIP codes. Each image was scanned from mail in a working post office at 300 pixels per inch in 8-bit grayscale on a high quality flatbed digitizer. The data were unconstrained for the writer, style and technique of preparation. These characteristics help overcome the limitations of earlier databases that contained only isolated characters or were prepared in a laboratory setting under prescribed circumstances. In addition, the database is divided into explicit training and testing sets to facilitate the sharing of results among researchers as well as performance comparisons.

In 1999 Al ISRA Arabic database [10] collected from 500 students, it contains words, digits, signatures, which is has limitation because it does not contain paragraphs. Another database lunched in 2002 is IFN/ENIT [4, 11], it was developed at the Institute of Communications Technology (IFN) at Technical University Braunschweig in Germany and the Ecole Nationale d'Ingenieurs de Tunis (ENIT) in Tunisia. It consists of 26,459 images of the 937 cities names and towns in Tunisia, written by 411 different persons filled forms with about 26400 names containing more than 21,0000 characters. For each name some information are coded such as the sequence of character shapes, some style information, and the baseline are coded. It is used for recognition of data entry, mail sorting, and other recognition tasks. The images are partitioned into four sets so that researchers can use and discuss training and testing data in this context. The database has certainly many advantages and some of its drawbacks is the fact that it is written for Tunisia only and therefore contains only Tunisian cities and names and does not cover other Arab countries. It also lacks reliable training and testing sets. As a result, it is not widespread among researchers.

One of the efforts that addressed the handwriting recognition problem is in writing personal checks. One such system, developed a decade ago, is AHDB (Arabic Handwritten DataBase), a database containing samples from 100 different writers, including words used for numbers [12]. In 2003 AI-Ohali et al., developed CENPARMI images databases from 3,000 checks and implemented at the Center for Pattern Recognition and Machine Intelligence provided by a banking corporation [13].

These databases contain numeric amounts written in words, sub-words, Indian digits, and numeric amounts written with Indian digits. Notably, Indian digits are the numeric digits normally used in Arabic writing, as opposed to “Arabic numerals” ordinarily used in Latin script. The Indian digits database contains 15,175 samples, the legal and courtesy databases 2,499 samples, and the sub-words database contains 29,498 samples.

In 2009 ADAB database (Arabic DAtaBase) with Arabic online handwritten words has used by Haikal, *et al* [14] at the first time, the database was developed for Arabic online handwritten scripts in a cooperation between the Institute for Communications Technology (IfN) and the Ecole Nationale d’Ing`enieurs de Sfax (ENIS). The database written by more than 130 persons, it consists of 15158 Arabic handwritten words, 937 Tunisian town/village names. The database contains in additional special tools for the collection of the data and verification of the ground truth. These tools give the possibilities to record the online written data, to save some writer information, to select the lexicon for the collection, and re-write and correct wrong written text.

Although the recognition accuracy for separated handwritten numerals and characters has improved significantly in recent years, the final frontier remains the accurate recognition of handwritten Arabic scripts. The pursuit of more accurate recognition rates continues to encourage researchers in the field. It must also be mentioned that along with the challenging nature of the handwritten word recognition problem, immense potential lies in the commercial sector to make these systems available. So, an important bulk of work is required to undertake a serious research and meet its multiple challenges.

III. BENCHMARK DATABASE

One of the most important components in ACDAR center is the benchmark database; often recognition algorithms have tested using one type of database, especially in the case of off-line handwriting recognition. ACDAR is concerned with both off-line and online handwriting recognition. It proposes a common benchmark database of Arabic handwritten scripts, which is essential for research on handwritten Arabic word recognition. The first issue of this database has hosted in the center.

ACDAR began to work with the off-line Arabic handwriting recognition, which is may divided into segmentation-based and holistic ones. In general, the former approach uses a strategy based on the recognition of individual characters or patterns whereas non-segmentation based deals with the recognition of the word image as a whole [15]. In the online case, the handwriting has captured and stored in digital form *via* different means. Usually, a special pen has used in conjunction with an electronic surface. As the pen moves across the surface or paper, the two-dimensional coordinates of successive points have represented as a function of time and have stored in order [1]. It is generally, information is not easy to recover from handwritten words written on a non-digital medium such as accepted that the online technique of recognizing handwriting has so far achieved better results than off-line. This may be attributed to the fact that more information may be captured in the online case such as the direction, speed and the order of strokes of the handwriting. At the end, ACDAR's database will be made freely available to researchers.

A. Data collection

ACDAR started recognizing the paragraphs, lines, words, and characters. The handwriting papers have wrote by 113 distinct writers and scanned in a RGB-scale. The writers are variant in age, education, background, genders, and countries; Figure 1 shows a snapshot form that contains the instruction and personal details of the writers, Table I shows the statistical data that collected from all writers. Figure 2 illustrates the comparisons of database contains.

As a result, two paragraphs contain all shapes of Arabic characters have wrote by those writers. Figure 3 shows the original two paragraphs have requested to write by the persons, Figure 4 displays in blue the position of the different characters shapes, the second paragraph has required for collecting more samples.

Fig. 1. Form of the personal details of each writers

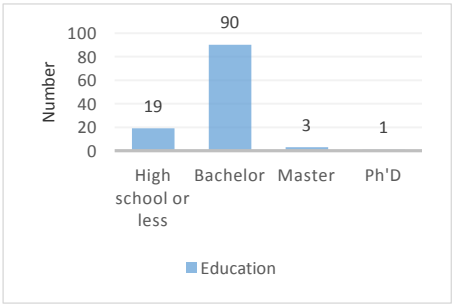
TABLE I. STATISTICAL DATA OF 113 WRITERS

Education	High school or less	Bachelor	Master	PhD
Number of writers	19	90	3	1

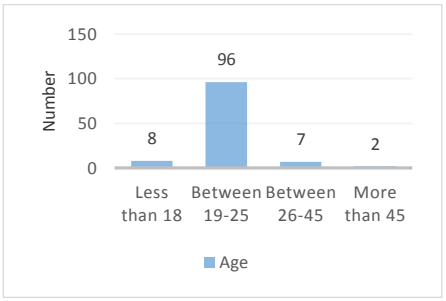
Age	Less than 18	Between 19-25	Between 26-45	More than 45
Number of writers	8	96	7	2

Gender	Male	Female
Number of writers	72	41

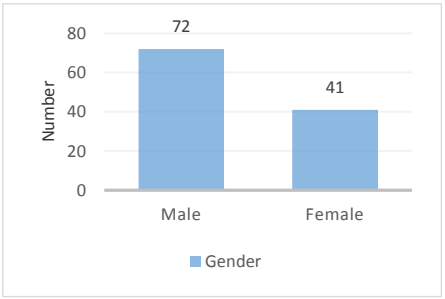
Country	Saudi Arabia	Jordan	Algeria	Syria	Egypt	Yemen
Number of writers	85	20	2	3	1	2



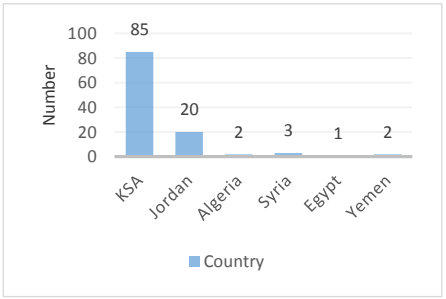
a) education



b) age



c) gender



d) country

Fig. 2. Comparisons of the statistical, a) education, b) age, c) gender, and d) country

Lines, words and characters have also extracted and saved in the database; verification phase has also investigated before the final adoption of the samples for quality purposes. In summary, each writer has wrote 358 words (first paragraph contains 162 words, second paragraph contains 196 words), both of them are 1,916 characters, on average each writer has wrote 30 lines. In total, all writers have wrote 226 paragraphs, 3,390 lines, 40,454 words, and 216,508 characters. Number of 57 writers are identified what they wrote as a training set, also 56 are writers identified what they wrote as a testing set. Table II illustrates the numbers of images have collected before extraction and validation processes.

Fig. 3. Two paragraphs cover all shapes of Arabic characters

Fig. 4. Position of all Arabic characters shapes covered by only the first paragraph

TABLE II. NUMBERS OF COLLECTED IMAGES BEFORE EXTRACTION AND VALIDATION PROCESSES

Details / writers	Each writer	Training set by 57 writers	Testing set by 56 writers	Total (training and testing sets) 113 writers
Paragraphs	2	114	112	226
Lines	30 as average	1,710	1,680	3,390
Words	358	20,406	20,048	40,454
Characters	1,916	109,212	107,296	216,508

The images have scanned in 200, 300 dpi resolution in RGB-scale images [2, 16, 17, 18]. Two version of the scanned documents have saved in the database, before and after preprocessing. Paragraphs, lines, words, and characters have extracted and saved as well.

The key steps of the techniques that were developed in this research is shown in Figure 5, the Figure shows briefly how we extracted all paragraph, lines, words, and characters. The first step is scanning the original documents with 200 and 300 dpi in RGB-scale, next step is preprocessing the scanned images, skew detection / correction, thresholding, and remove the noise using filtering are investigated in this stage. Next, start extracting process of the database; this stage includes developing a set of techniques to get the best extraction results of lines, words and characters. Finally, the last step is validating step, all extracted elements underwent to the evaluation process, if the element successfully passed this stage, then it will save into the benchmark database, otherwise it will discard. As mentioned before, this research aims to build the first issue of ACDAR database and test the proposed algorithms have developed in this research.

B. Pre-processing

Many techniques have developed to perform further processing to allow superior recognition. Thresholding and filtering which they aim to eliminate and remove any noise or any small ascenders. Skew detection and correction technique that aims to adjust slopes of the paragraphs and lines. Next sub-sections are explain in details the parts of the pre-processing.

1) Thresholding and Filtering

The first step of preprocessing is thresholding (binary format); it uses as prior to further processing. Thresholding involves the conversion of a grey-scale image (0–255) into a binary image (0–1). This format will be easier to manipulate an image without levels of color in some researches, in additional the processing will be faster, less computationally expensive and will allow for more compact storage. The goal of using the thersholding is to determine the segmentation points of the lines, words, and the characters. Determine the segmentation points from the grey-scale image will be easier than color the image; the same points have extracted were applied on the RGB-scale. There are of course many of the defects such as loosing features from image. However, since the goal of this stage is only to determine the segmentation points, the effect will be the lowest grades possible. *rgb2gray* function was used to converts RGB images into grayscale by eliminating the hue and saturation information while retaining the illumination. The definition *rgb2gray* is shown in the following equations. The *im2bw* function was also used to convert this grayscale image to binary format (matrix). The output binary image BW has values of 0 as a foreground pixel (black) for all pixels in the input image and 1 as a background pixel (white) for all other pixels. All images were converted using the previous technique so that only binary images remained and could be used for further processing.

$$g = w_r I_r + w_g I_g + w_b I_b \quad (1)$$

$$\text{s.t. } w_r + w_g + w_b = 1, \quad (2)$$

$$w_r \geq 0, w_g \geq 0, w_b \geq 0, \quad (3)$$

where, g is a constraint linear combination of R, G and B channels of input color image I,

I_r, I_g, and I_b are the inputs,
w_r, w_g, and w_b : weights sum to 1, and they are non-negative numbers.

Next, elimination of the elements noise; the goal of this technique is to remove the noise as well as small foreground objects that were not part of the writing. Once the component of word image as matrix were identified, it was possible to perform various useful operations. *imfilter* function has used, it performs multidimensional filtering according to the specified options like *fspecial* function to create 2-D special filters that used ‘disk’ function to returns a circular averaging filter 'pillbox' within the square matrix of side (2 * radius + 1). Gaussians function [19] is applied, it was used at the lowest degree possible in order to not lose the features of the scripts

as much as possible. The function aims to make the word image more smoothly, and to eliminate any small ascenders or descanters noise between the lines. The following equation shows the one per direction using Gaussians function.

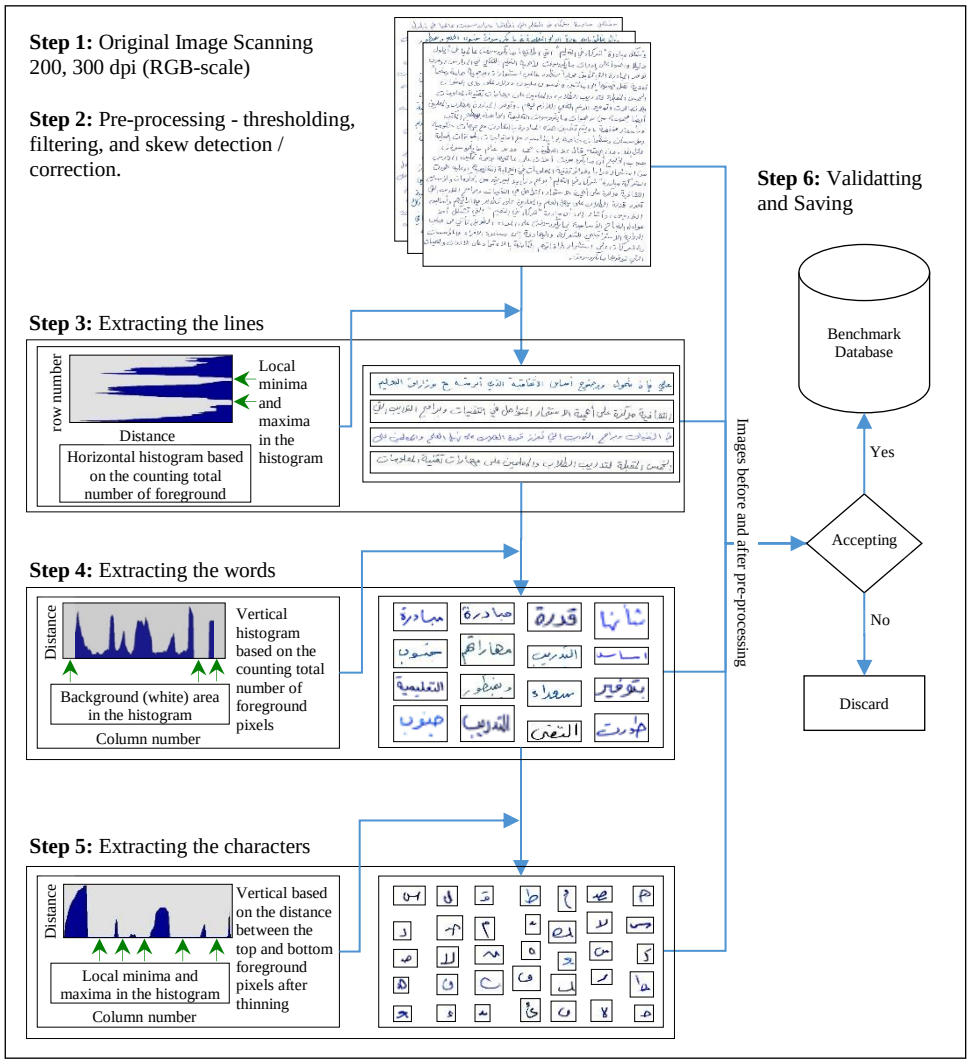


Fig. 5. Steps of extracting and validating the benchmark database

$$g(x,y) = \frac{1}{2\pi\sigma^2} \quad . \quad e^{-\frac{x^2+y^2}{2\sigma^2}} \qquad (4)$$

where, σ is the standard deviation of the Gaussian distribution,

x is the distance from the origin in the horizontal axis,

y is the distance from the origin in the vertical axis.

2) Skew Detection and Correction

Before extracting the lines, words, and characters from the documents images, skew of the paragraphs should be detect and then correct, the technique uses projections of an image matrix along specified directions, Hough transform [20] algorithm is applied to detect and correct the slopes. Hough Transform is the linear transform for detecting straight lines, the straight line is described as $y = mx + b$ where the parameter m is the slope of the line, and b is the intercept (*y-intercept*). Before start to apply the Hough transform algorithm, the document image should be prepared. So, a set of steps were used, at the beginning Threshold of the image to binary was applied. Next, in order to obtain a clear base line for all lines in the page, the punctuation marks (dots) and small stroke have removed from the image. Then, dilate image is also applied to close the internal gaps between the characters and words as well. Closing operation was performed upon the horizontal line element and merging the words of the lines. The text lines now look likes rectangles, to apply the Hough transform one-step is remained, this step is thinning the image includes all horizontal rectangles. To find the skew of the image, the mean and standard deviation of slopes were calculated, any bad data conceders far away from the standard deviation was removed, then the average of the good slopes was calculated, therefore the skew can be calculated by using the angle of the slope. The skew correction has applied using the negative of this angle. The below equations shows the Hough transform technique, the line

equation can be written as shown in equation (5), rearranged the equation shown in equation (6), and equation (7) shows formula of an point on the image with coordinates. Figure 6 shows the steps of skew detection and correction have developed in this research; Figure 7 shows samples of skew detection and correction for one paragraph and one line.

$$y = \left(-\frac{\cos\theta}{\sin\theta}\right)x + \left(\frac{r}{\sin\theta}\right) \quad (5)$$

$$r = x\cos\theta + y\sin\theta \quad (6)$$

$$r(\theta) = x_0\cos\theta + y_0\sin\theta \quad (7)$$

where, r distance between the line and the origin, is determined by θ,

θ is the angle of the vector from the origin to this closest point.

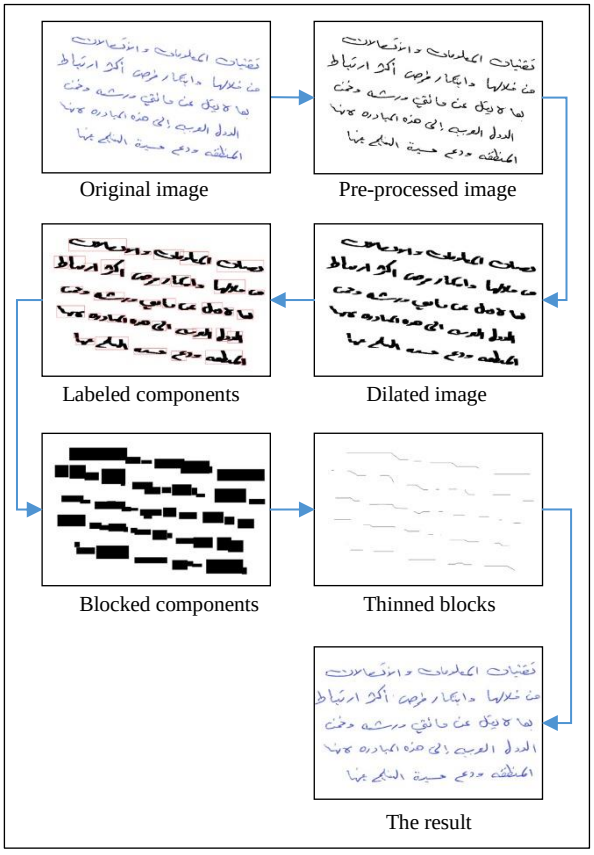


Fig. 6. Steps of skew detection and correction

C. Extracting the database

Local minima and maxima of horizontal and vertical histograms have used for determining the segmentation points SPs for extracting the lines, words and characters. The concept of using the horizontal histogram is for extracting the line image. Horizontal histogram is formed by counting the total numbers of foreground pixels (black color) for each row from left to right in the paragraph image; the segmentation points have located based on the white color (background pixel) or the distance between two successive local maxima and one local minima with almost no foreground pixels. Figure 8 illustrates technique of extracting the lines images based on horizontal histogram.

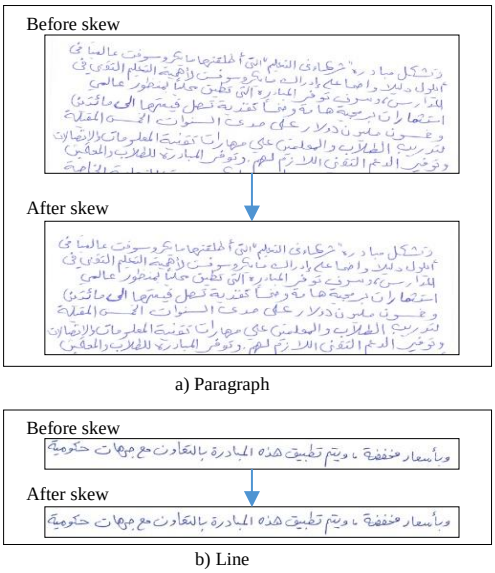


Fig. 7. Sample of skew detection and correction, a) a paragraph b) a line

Using the same technique, but now by applying the vertical histogram to extract the word images. Vertical histogram has formed by counting the total numbers of foreground pixels (black color) for each column from top to bottom in the line image. The segmentation points have located based on the white color (background pixel) or the distance between two successive local maxima and one local minima with almost no foreground pixels. Figure 9 illustrates technique of extracting the words images based on vertical histogram.

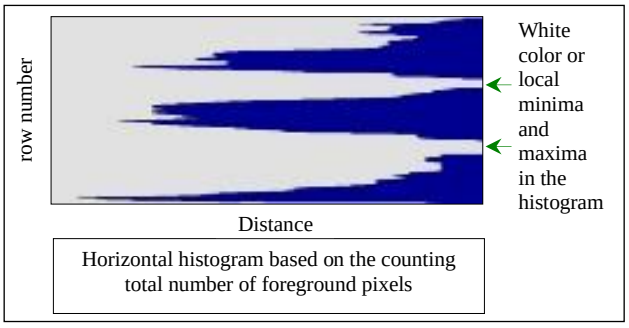


Fig. 8. Extracting the lines images based on horizontal histogram

Likewise, extracting the characters images technique uses also the vertical histogram, which has calculated based on the distance between the top and bottom of foreground pixels for the word image after thinning. Extracting of the characters from the words is required to remove the punctuation marks (dots). The dots here consider a major obstacle to identify the correct segmentation points of the characters. After determining the segmentation points, the dots will recover. Figure 10 illustrates technique of extracting the word images based on vertical histogram before thinning and Figure 11 after thinning.

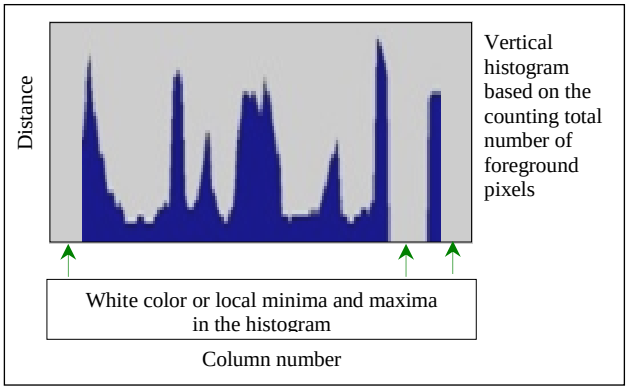


Fig. 9. Extracting the words images based on horizontal histogram before thinning

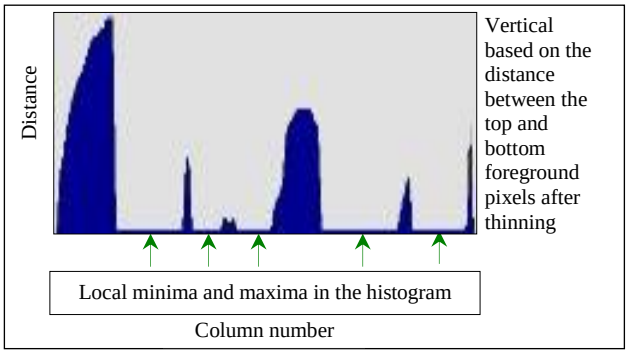


Fig. 10. Extracting the word images based on horizontal histogram after thinning

The following equations show how the histogram has calculated based on the total number of the foreground pixels.

$$pr(r_k) = \frac{n_k}{n} \quad (8)$$

$$s_k = T(r_k) = \sum_{j=0}^{k} p_r(r_j) = \sum_{j=0}^{k} \frac{n_j}{n} \quad (9)$$

where, K=0, 1, …., L-1,
N is total number of pixels in the image,
L is total number of possible grey levels in the image.

IV. Experimental Results

As a result of all the previous steps, in addition to the final stage of verification processes, the first issue of ACDAR database is now available. ACDAR database contains 208 pages, 208 paragraphs, 2,969 lines, 32,890 words, and 158,872 characters, the database is divided into two sets one for training and the second for testing. The details of the first issue of ACDAR database after extraction and validation process is shown in Table III. Table IV summarizes a comparison between the results of some databases use the Arabic handwritten scripts. Diversified samples from the ACDAR database have published in ACDAR's website under this link http://www.acdar.org/DBsamples.php.

Figure 12 displays samples of handwritten paragraph that written by one person with its printed text, Figures 13 to 16 display samples of complete free handwritten paragraph. More samples for characters, words, and paragraphs see http://www.acdar.org/ DBsamples.php.

TABLE III. Final Database after Extracting and Validating Processes

Details / writers	Each writer	Training set by 51 writers	Testing set by 53 writers	Total (training and testing sets) 104 writers	Percentage from the original
Paragraphs	2	102	106	208	92.0%
Lines	Average 30	1,467	1,502	2,969	87.6%
Words	358	16,214	16,676	32,890	81.3%
Characters	1,916	78,584	80,288	158,872	73.4%

TABLE IV. Comparison Between Some of Handwritten Databases Used Arabic Letters

Database	Details	Year	Writers
Al-Isra [10]	▪ 500 sentences ▪ 37,000 words ▪ 10,000 digits ▪ 2,500 signatures	1999	500
AHDB [12]	▪ 10,000 words for check processing	2002	100
IFN/ENIT [8]	▪ 26,459 Tunisian city names	2002	411
Khedher and Abandah [21]	▪ 48 pages of text	2002	48
IFHCDB [22]	▪ 52,380 characters ▪ 17,740 numerals	2006	–
ADBase / MADBase [22]	▪ 70,000 digits	2007	700
Alamri et al. [24]	▪ 11,375 words ▪ 21,426 characters ▪ 46,800 digits strings ▪ 13,439 numerical ▪ 1,640 special symbols	2008	328
On/Off LMCA [25]	▪ 500 words ▪ 1,00,000 characters ▪ 30,000 digits	2008	55
Al Hamad et al. [2]	▪ 20 Pages ▪ 40 paragraphs ▪ 500 words ▪ 620 characters	2010	10
ACDAR – First issue	**▪ 208 Paragraphs/Pages ▪ 2,969 Lines ▪ 32,890 words ▪ 158,872 characters**	**2014**	**113**

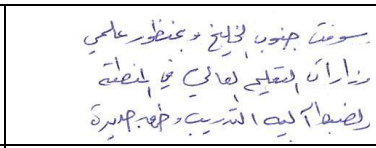
handwritten	
printed	سوفت جنوب الخليج وبمنظور علمي وزارات التعليم العالي في المنطقة لضبط آلية التدريب وطرق جديدة

Fig. 11. Sample of part free handwritten paragraph

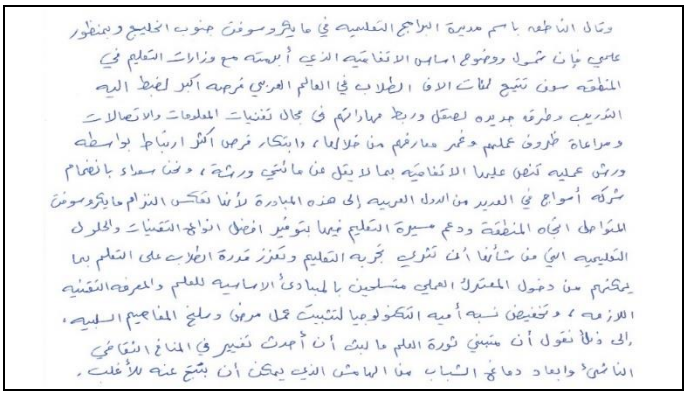


Fig. 12. Sample (a) of complete free handwritten paragraph 1

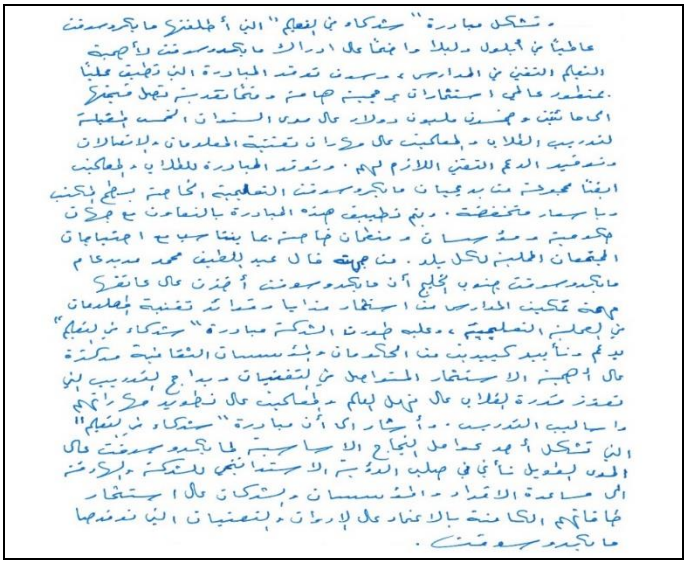


Fig. 13. Sample (a) of complete free handwritten paragraph 2

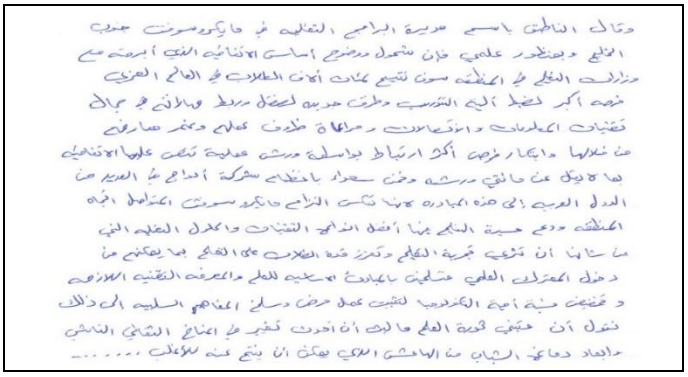


Fig. 14. Sample (b) of complete free handwritten paragraph 1

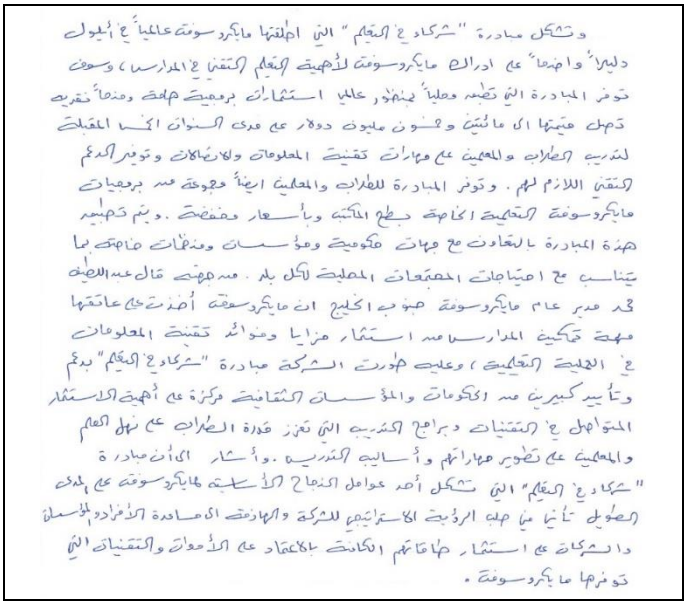


Fig. 15. Sample (b) of complete free handwritten paragraph 2

Figures 17 and 18 show samples of free handwritten lines extracted from the paragraphs.

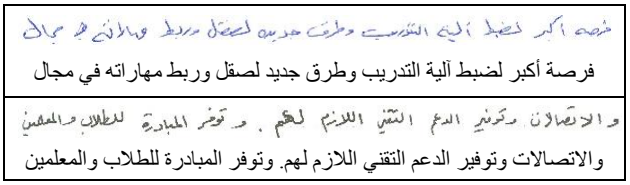


Fig. 16. Samples of ACDAR free handwritten lines

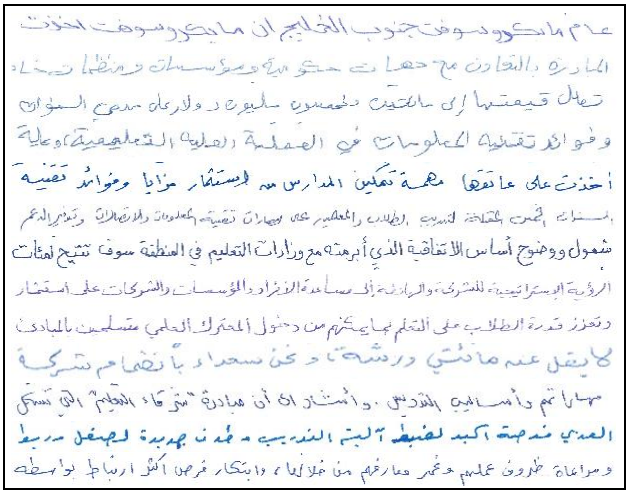


Fig. 17. Samples of ACDAR free handwritten lines written by different persons

In addition, the first issue of the ACDAR database contains different samples of Arabic handwritten words; Figure 19 shows samples of handwritten words with their printed text, Figure 20 shows samples of free handwritten extracted from the lines image.

A sample of the characters that extracted from the words image and wrote by one person is shown in Figure 21; Figure

22 shows samples of free handwritten characters wrote by many writers.

Handwritten	Printed	Handwritten	Printed
الذي	الذي	وقال	وقال
ودعم	ودعم	ومراعاة	ومراعاة
أحد	أحد	التعليم	التعليم
الدول	الدول	امواج	امواج
الطلاب	الطلاب	أساس	أساس
مساعدة	مساعدة	وزارات	وزارات
ارتباط	ارتباط	وابتكار	وابتكار
باسم	باسم	وغمر	وغمر

Fig. 18. Samples of ACDAR free handwritten words

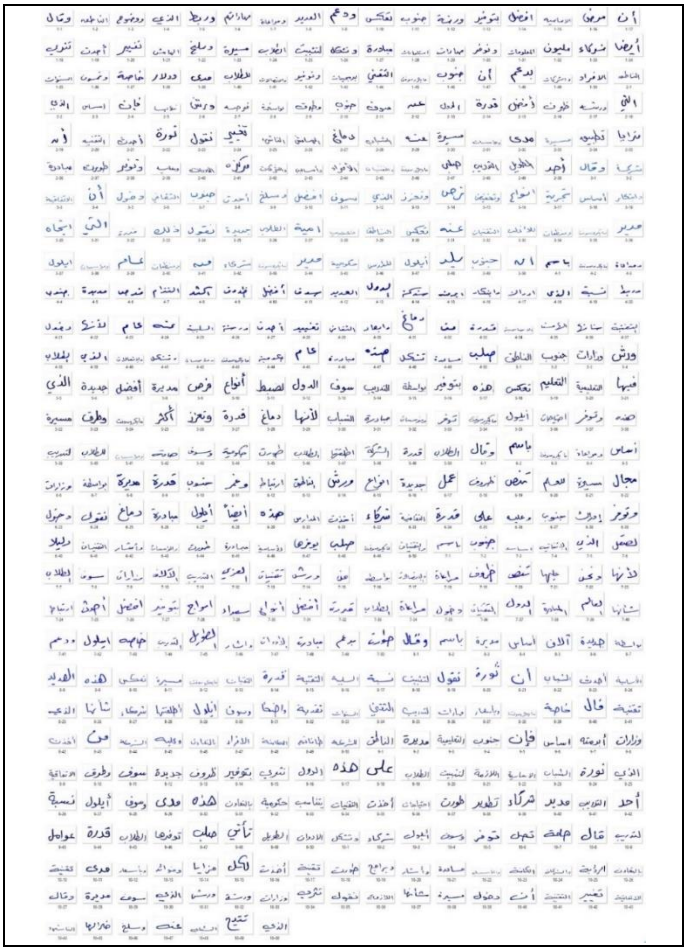


Fig. 19. Samples of ACDAR free handwritten characters written by different persons

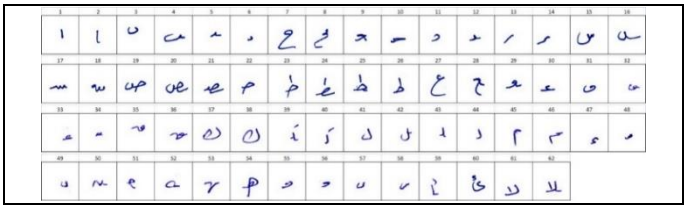


Fig. 20. Sample of ACDAR character that wrote by one person

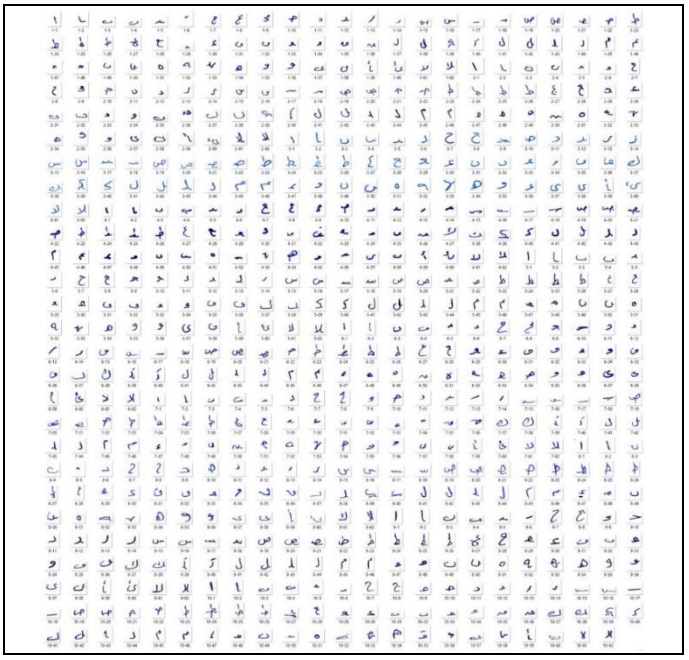


Fig. 21. Samples of ACDAR free handwritten characters written by different persons

V. ACDAR – A NEW CENTER

As any other center in this area, ACDAR [26] contains a set of internal sections describing the main objectives of the center and its contents, these sections cover many activities in the area of document analysis and recognition such as teach courses, research, publications, resources, people, contact details. While analysis of documents and handwriting recognition continues to be our primary interest, we propose research and software development projects involving diverse digital document types.

ACDAR is dedicated to re-build and re-structure a reliable and standard a benchmark database and set of integrated tools for handwritten Arabic scripts within, it is newly established a website (http://www.acdar.org). The website includes details about the center, and a sample of the first issue of the benchmark database. In the conceptual framework of ACDAR, many functions would give ACDAR its identity, mission, and direction. These centered on the benchmark database, research, training, and collaboration with the community. More details about ACDAR center see Al Hamad *et al* [26].

VI. CONCLUSION

The paper presents new techniques for extracting the first issue of a new benchmark database, it has written by 113 distinct writers with different ages, cultures, and genders. Two paragraphs cover all shapes of Arabic characters have scanned with different resolution; the final database contains 208 pages, 208 paragraphs, 2,969 lines, 32,890 words, and 158,872 characters. Half of the database assigns as training set; another part assigns as testing set. For extracting and validating the proposed database, the research has developed and tested a set of new techniques. An example of these techniques are pre-processing of the images such as

thresholding, filtering, local minima and maxima of vertical and horizontal histogram for the segmentation, in addition, developing skew detection / correction technique, etc. The techniques have examined and tested through several experiments in order to use them later for creating a comprehensive database that we seek to cover all Arab countries. The paper also displays a comprehensive details of forming a new center for analysis and recognition Arabic handwritten scripts, the center calls "ACDAR". Functions and activities of the center have identified and explained in detail.

REFERENCES

[1] Plamondon, R., S.N. Srihari, "On-line and Off-line Handwriting Recognition: A Comprehensive Survey," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 22, 2000, pp. 63–84.

[2] Al Hamad, H.A., R. Abu Zitar, "Development of an Efficient Neural-based Segmentation Technique for Arabic Handwriting Recognition," Pattern Recognition, vol. 43(8), 2010, pp. 2773-2798.

[3] Blumenstein M., "Intelligent Techniques for Handwriting Recognition. School of Information Technology," PhD Dissertation, Griffith University-Gold Coast Campus, Australia, 2000.

[4] Lorigo, L., V. Govindaraju, "Off-line Arabic Handwriting Recognition: A Survey," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 28(5), 2006, pp. 712–724.

[5] Hamid, A., R. Haraty, "A Neuro-Heuristic Approach for Segmenting Handwritten Arabic Text," ACS/IEEE International Conference on Computer Systems and Applications, AICCSA, vol.1, 2001, pp. 1–10.

[6] Hull, J., "Database for Handwritten Text Recognition Research, Center of Excellence for Document Analysis and Recognition (CEDAR)," Department of Computer Science, State University of New York at Buffalo, Buffalo, New York, http://www.cedar.buffalo.edu/ Databases/CDROM1, 1993.

[7] Kharma, N., M. Ahmed, R. Ward, "A New Comprehensive Database of Hand-written Arabic Words," Numbers and Signatures used for OCR Testing. IEEE Canadian Conference on Electrical and Computer Engineering, 1999, pp. 766-768.

[8] Pechwitz, M. et al., "IFN/ENIT – Database of Handwritten Arabic Words, Proc. of CIFED," 2002, pp. 129-136.

[9] Nouh, A., A. Sultan, R. Tolba, "An Approach for Arabic Characters Recognition," J. Eng. Sci, vol. 6, 1980, pp. 185–191.

[10] N. Kharma, M. Ahmed, R. Ward, "A new comprehensive database of handwritten Arabic words," numbers, and signatures used for OCR testing, Canadian Conference on Electrical and Computer Engineering, 1999, pp. 766–768.

[11] Pechwitz Mario, et al, "IFN/ENIT – Database of Handwritten Arabic Words," Institute for Communications Technology (IFN), Technical University Braunschweig, Germany, Ecole Nationale d'Ingénieur de Tunis (ENIT), BP 37 le Belvédère 1002, Tunis. IFN/ENIT, 2002, http://www.ifnenit.com/.

[12] Alma'adeed, S., D. Elliman, C. A. Higgins, "A Database for Arabic handwritten Text Recognition Research," Eighth International Workshop on Frontiers in Handwriting Recognition, 2002, pp. 485-489.

[13] AI-Ohali, Y. M. Cheriet, and C. Suen, "Databases for Recognition of Handwritten Arabic Cheques," Pattern Recognition, vol. 36, 2003, pp. 111-121.

[14] Haikal El Abed, et al, "Online Arabic Handwriting Recognition Competition," 10th International Conference on Document Analysis and Recognition ICDAR, 2009, DOI 10.1109/ICDAR.2009.284.

[15] Fan, X., B. Verma, "Segmentation vs. Non-Segmentation Based Neural Techniques for Cursive Word Recognition," An Experimental Analysis International Journal of Computational Intelligence and Applications, vol. 2(4), 2002, pp. 377–384.

[16] Al Hamad, H.A., "Over-segmentation of handwriting Arabic scripts using an efficient heuristic technique," IEEE International Conference on Wavelet Analysis and Pattern Recognition (ICWAPR), 2012, pp.180-185.

[17] Al Hamad, H.A., "Neural-Based Segmentation Technique for Arabic Handwriting Scripts," WSCG 2013, 21st International Conference on Computer Graphics, Visualization and Computer Vision, indexed by Thomson Reuters/ISI-WoS, Czech, June, 2013.

[18] Al Hamad, H.A., "Use an Efficient Neural Network to Improve the Arabic Handwriting Recognition," IEEE International Conference on Signal and Image Processing Applications (ICSIPA), 2013.

[19] Nixon M. S. and Alberto S. Aguado, "Feature Extraction and Image Processing. Academic Press," 2008, pp. 88.

[20] Richard O. Duda and Peter E. Hart, "Use of the Hough transformation to detect lines and curves in pictures," ACM, vol. 15(1), 1972, pp. 11-15, doi:10.1145/36:1237.361242.

[21] Khedher, M., Abandah, G., "Arabic character recognition using approximate stroke sequence," Arabic Language Resources and Evaluation - Status and Prospects Workshop, 3rd International Conference on Language Resources and Evaluation (LREC'02), 2002.

[22] Mozaffari S., Faez k., Faradji F. Ziaratban M, Golzan S. M., "A comprehensive isolated Farsi/Arabic character database for handwritten OCR research," In Proceedings of the 10th International Workshop on Frontiers in Handwriting Recognition (IWFHR), 2006, pp. 385–389.

[23] El-Sherif E., Abdelazeem S., "A two-stage system for Arabic handwritten digit recognition tested on a new large database," In Proceedings of the International Conference on Artificial Intelligence and Pattern Recognition (AIPR'07), 2007, 237–242.

[24] Alamri H., He C. L., Sue C. Y., "A new approach for segmentation and recognition of Arabic handwritten touching numeral pairs," Proceedings of the International Conference Computer Analysis of Images and Patterns (CAIP). Lecture Notes in Computer Science, vol. 5702, Springer, 2009, pp. 165–172.

[25] Kherallah Monji , Elbaati A., El Abed H., Alimi A. M., "The on/off (LMCA) dual Arabic handwriting database," 11th International Conference on Frontiers in Handwriting Recognition (ICFHR), 2008.

[26] Al Hamad, H.A., Hamdi-Cherif A., "The Arabic Center for Document Analysis and Recognition (ACDAR) - Structure and Perspective," European Conference of COMPUTER SCIENCE (ECCS '12), 2012.