Three Evaluation Criteria's Towards a Comparison of Two Characters Segmentation Methods for Handwritten Arabic Script

Fadoua Bouafif Samoud Samia Snoussi Maddouri Hamid Amiri
Signal, Image and Information Technology Research Lab
National School of Engineers of Tunis ENIT
BP37, Belvedere 1002 Tunis, Tunisia
Fadoua_Bouafif@yahoo.fr *Samia.Maddouri@enit.rnu.tn* *Hamid.Amiri@enit.rnu.tn*

***Abstract*-This paper presents three evaluation criteria's for a comparison of two characters segmentation methods for handwritten Arabic words. The first segmentation method is based on a combination between the projection and the minima and maxima of the contour of the image. The second method is a combination between Hough Transform (HT) and Mathematical Morphology (MM) operators. These methods are developed, evaluated and compared with reference to IFN/ENIT-database in comparison of three evaluation criteria's. The first criterion is based on the segments positions (SP). The second criterion is based on the segments numbers (SN). The third is based on the recognition rates by Transparent Neural Network (RR).**

Keywords-Arabic word; handwriting; segmentation; characteres.

I. INTRODUCTION

Many research have been published in the area of segmentation and recognition of handwritten Arabic script [1], but unfortunately the result of Arabic handwriting recognition and segmentation still have not reached a required level. The reasons for that relate to the nature of Arabic writing, where most words are written cursively and sometimes depend on the character, some characters can be connected with others. Most characters have three or four shapes according to their position in the word: “Isolated or Single”, “Beginning”, “Middle” and “End”. Also external objects are used in Arabic writing like "dots", “Hamza” and movements that make the task of segmentation more complicated. In additional, characters that do not touch each other's but occupy a shared horizontal space increase the difficulty of segmentation. The extraction of the characters from the handwritten Arabic word is the pivotal stage of recognition step. Several segmentation methods are presented in literature such as skeletization method [2, 3, 4], a both contour [4] and skeleton segmentation [5], a projection method [2], and hybrid methods [1, 4]. Nearly all these methods need a recognition step to validate the segmentation stage.

Our contribution in this work is a proposition of two segmentation methods without recognition stage and new criteria used to validate the segmentation points of characters. Indeed, two segmentation methods are achieved and evaluated on the IFN/ENIT database in comparison to three evaluation criteria’s. The comparison concerns not only methods but also criteria’s. The first segmentation method based on a combination between the projection and the minima and maxima of the contour of the image. The second method is a combination between Hough Transform (HT) and Mathematical Morphology (MM) operators**.** The first evaluation criterion is based on segments numbers (SN). The second criterion is based on segments positions (SP). The third is based on the recognition rates by Transparent Neural Network (RR).

This paper is organized as the following: section two focuses on proposed characters segmentation methods of words, section three, emphasizes on evaluation of each method on IFN/ENIT-database by the use of different criteria’s in order to choose the efficient one for structural description and transparent neural network recognition of Arabic script [6], and finally conclusion and future work in section four.

II. CHARACTERS SEGMENTATION

In this section, we propose two character segmentation methods. The first one based on the contour, we introduce minima and maxima of the boundary of the image which are between the upper and the lower baseline deduced from a combination between the min-max and the projection methods. The second method, based on the HT and MM (HT-MM), the introduction Harris detectors is needed.

A. Detection of baselines

Baseline is an artificial line composed by a sequence of aligned pixels that connect the maximum

978-0-7695-4774-9/12 $26.00 © 2012 IEEE
DOI 10.1109/ICFHR.2012.283

black pixels of the characters in the word. Three types of baselines exist: lower baseline, upper baseline and median baseline. Different baseline extraction methods abound in literature [6]. The proposed baseline detection method is applied on the binary word without slant correction. It is based on the HT in order to detect the median baseline. The horizontal projection stage is applied on the Hough space in order to extract the lower and upper lines. These tow lines divide the word into three parts: (1) Ascender and upper diacritic points above the upper baseline; (2) Descender and lower diacritic points under the lower baseline and (3) the main content of the word between the two baselines. Fig. 1 illustrates the three baselines extracted by HT and horizontal projection.

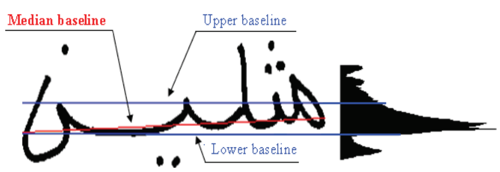


Figure 1. Median, upper and lower baseline detected by HT and horizontal projection

B. Min-Max-projection segmentation method

Min-Max method is used to detect critical points from the boundary of a word. From an initial point, we go through the contour and extract minima and maxima. This method is based on the segmentation of characters having structural primitives. To estimate the position of the primitive, we search for the existence of global maxima or minima, a loop or diacritic dots. The existence of such primitives indicates the existence of corresponding letter. After this first estimation of character zones position, an improvement of segment limits is given by the extraction of the structural primitives.

1) Primitives of handwritten Arabic script: During structural primitive's extraction of handwritten Arabic script, we extract ascender, descender, loops, high and low diacritic dots.

Ascender: It is defined as a character having a high greater than a threshold and a width lower than its height. This threshold depends on the size of the script, as estimated in formula (1):

$$\text{Threshold} = \text{upper baseline} + 2*(\text{upper baseline} - \text{lower baseline}) \tag{1}$$

Local maxima, detected below global maxima, throw the upper contour of the Pieces of Arabic Word (PAW), indicates the existence of an ascender, if the distance between these two maxima is greater than the threshold of formula (1). These kind of local maxima are then used to improve the estimated boundaries of the segmented zone. Fig. 2 illustrates the improvement of the second PAW segmentation. The lower boundary of the first ascender from the right to the left is moved from its estimated position illustrated by discontinue vertical line to the right position illustrated by a bold line. The same improvement is given to the upper boundary of the third ascender of the whole word.

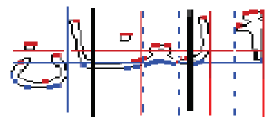


Figure 2. Estimation of ascender zones and their improvement

Descender: In Arabic script the descender can be only at the end of a PAW or as an isolate character. By the extraction of this primitive we can improve the lower boundary at the left of the segmented zone. Fig. 3 illustrates the improvement of descender zones. In fig. 3(a), PAW's are processed separately. In the first PAW at the right, the lower boundary of the descender at the end is moved from bold vertical line to discontinue line. For the second PAW the two limits of the isolated descender are improved. Fig. 3(b) evinced the final result obtained after improvement and superposition of the processed PAW's. Treating PAW's separately resolves the overlapping problem due to descenders.

Diacritic points and loops: In the case of loop or diacritic points, we marked the intersection between the zone boundaries and the loop or diacritic points. Then we try to move the boundary in order to eliminate this intersection. Fig. 3 illustrates a correct extracted loop at the beginning of the first PAW and lower diacritic points in its middle. Higher diacritic point is detected in the second PAW presented as isolated letter and contributes to the improvement of the zone boundaries.

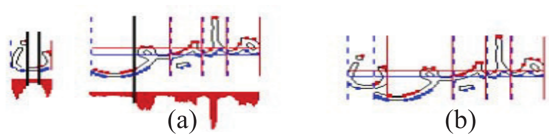


Figure 3. Descender zones estimation and improvement

2) Zones without structural primitives: The zones without primitives present ambiguities. These zones are detected during characteristics extraction stage and can be in the beginning, in the middle or at the end of a PAW.

Central zones: These zones exist between already extracted zones having global structural primitives as shown in fig. 4. Boundaries of these zones are deduced from the lower limit of the previous zone and the higher limit of the next zone. We can have in some case more than one character.

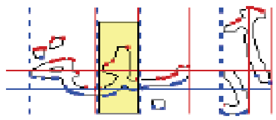


Figure 4. Estimation of the ambiguous zones in the middle of a PAW

First Zones: The higher boundary of a zone in the beginning of a PAW corresponds to the first black pixel while browsing the image from right to left. The lower boundary could be deduced from extracted global primitive in a second position of the processed PAW. Fig. 5 shows this step. If boundary detected is unknown, we search for a minimum above the upper baseline.

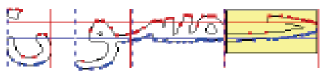


Figure 5. Estimation of the ambiguous begin zones of the PAW without structural primitives

Last zones: We apply the same process of the first zone to detect the last zone of the PAW. The lower boundary is easy to extract by browsing image from left to right (fig. 6). The upper boundary is estimated from the global extracted primitive or as a first minimum above the upper baseline. Within such fig. 6, limits of the begin zone of the word can also be estimated.

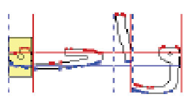


Figure 6. Estimation of ambiguous last zones of the PAW

C. HT-MM segmentation method

The present method is composed by two stages: extraction of PAW's and detection of characters.

1) PAW's detection: Handwritten Arabic word can have some discontinuities due to the fact of the pen up and the binarization stage. According to these discontinuities we can detect more than the real number of PAW in a given word. The application of morphological filter can connect some parts of PAW. A labeling stage is then applied in order to associate a label color to each connected component. The labeled component can be an isolated character or a diacritical dots or characters set. The PAW extraction needs to eliminate the components existing below lower baseline and above upper baseline. Fig. 7 illustrates the different steps of PAW detection. Fig. 7(a) presents the original image, where discontinuity is surrounding. The smoothing and labeling stage are shown in fig. 7(b), where the correction of discontinuity is surrounded and the baselines are presented. Fig. 7(c) shows two extracted PAW's where the diacritic points localized above upper baseline are eliminated.

(a) (b) (c)

Figure 7. PAW's detection

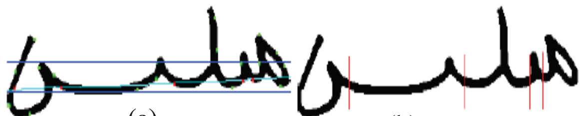
2) Characters segmentation: The Harris Corner Detector [8] is probably the most widely used interest point detector thanks to its strong invariance to scale, rotation and illumination variations, as well as image noise. The detector is based on the matrix $C(x, y)$ which is computed over a pxp patch for each interest point at position (x, y) as given in formula *2*.

$$C(x, y) = \begin{pmatrix} \sum I_{x^2} & \sum I_x I_y \\ \sum I_x I_y & \sum I_{y^2} \end{pmatrix} \quad (2)$$

Where, *Ix* and *Iy* are the image gradient in horizontal and vertical directions. Let λ_1 and λ_2 be the Eigen values of the matrix C(x, y), we define the auto-correlation function *R* as equation 3.

$$R = \lambda_1 \lambda_2 - k(\lambda_1 + \lambda_2)^2 \quad (3)$$

This function will peak sharply, if both of the Eigen values are high. This means that shifts in any direction will produce a significant increase, indicating that it is a corner. A typical value for k is 0.04. After the PAW's extraction steps, characters segmentation points are generated by Harris Corner Detector. Indeed, for each black pixel we calculate the Harris autocorrelation function R given by the formula 3. We keep only local maxima for which Harris map R is superior to a threshold S. For each spatial maximum, we add the corner to a list. Finally, we remove from this list, corners too close to each other whose separate distance is below D. The median zone between the upper and the lower baselines contains valleys described as horizontal segments connecting adjacent peaks. For each point P detected by Harris in the median zone, we calculate the distance in relation to median baseline. P is a segmentation point if D is blow a threshold T. After a training step on test database, we choose the following values: S=30, D=12 and T=5. Fig. 8 illustrates steps of characters PAW segmentation. Fig. 8(a) presents the baselines detected by HT. The interest points detected by Harris are shown in different color and detected segmentation points are presented vertical lines. The result of characters segmentation is presented in fig. 8 (b).



(a) (b)

Figure 8. Characters Detection by HT and Harris

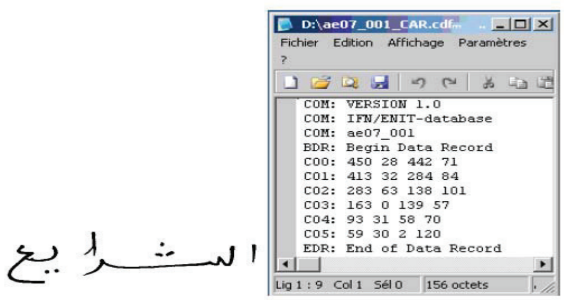
III. EVALUATIONS AND COMPARISON

Characters segmentation methods are evaluated on the set-a of the IFN/ENIT Database [9].

D. *Improvement of IFN/ENIT database*

From 2002 only the set-a of the IFN/ENIT database has been automatically labeled and manually verified. In this labeling step we found the corresponding postal code and printed name of the image, the description of real letters and their real number, shape and position, the number of PAW's and the position and the quality of the upper and the lower baselines. To can evaluate segmentation by the IFN/ENIT database we add positions of characters in the image.

A manual segmentation of 9730 characters was performed from 1250 images of the set-a. An example of manual segmentation results is given in fig. 9. In fig. 9(a) the main word. In fig. 9(b) we give associated manually files describing the segmented word in characters. The word is composed by six characters coded by: C00, C01, C02, C03, C04, C05.



(a) (b)

Figure 9. File description of manual segmented characters

E. *Evaluation criteria's*

The main problems of Arabic segmentation methods are: firstly, the method of evaluation and in many cases uses recognition step. Secondly, evaluation is not done on the same data that is why even comparison cannot be done. Three steps of evaluation are proposed and compared in this paper. The first evaluation step is based on the comparison between the number of characters automatically extracted by each of the two methods, and the manual characters extracted number. The second evaluation method is based on the characters positions in the word. The last evaluation method is based on the impact of each of the segmentation methods on the recognition rate of a transparent neural network based on structural description of recognized words.

1) Segments Numbers criteria (SN): The word is considered to be correctly segmented if the correct number of characters extracted automatically is equal to the number of characters extracted manually. It is considered over segmented if the correct number of characters extracted automatically is less than the number of characters extracted manually and under segmented else. In Fig. 10 we present the obtained results, the extraction rate is lower than 30%. This low rate is explained by many reasons. The first reason is that the Min-Max-projection method is based on threshold value to distinguish between diacritic points and PAW's. The second reason is that it depends on the quality of the baseline extraction method. Indeed, global maxima are considered to be greater than upper baseline and global minima are lower than lower baseline. A work on baseline extraction methods is done in [7] in order to resolve this problem.

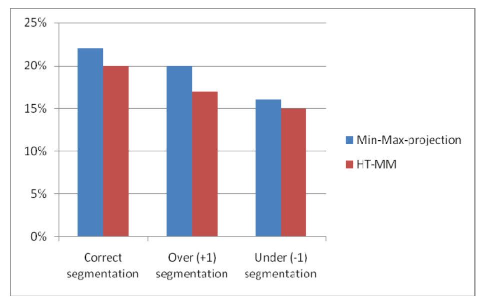


Figure 10. Rate extraction of the two segmentation methods by the use of segments number

The baseline quality can also have his impact on the bad extraction rate of HT-MM method based also on this concept for choice of Harris segmentation points. Also, we are not sure that all words having correct number of segments are correctly segmented.

2) Segments Positions criteria (SP): Here, the comparison is made only according to the X axis segment positions of IFN/ENIT database words. Fig. 11 gives extraction rate of each method in comparison to manually IFN/ENIT segmentation point called MSP and already presented in section *D*. The segmentation points which is between MSP+2 and MSP-2 pixels is considered to be good extracted. It is deemed acceptable if it is between MSP+2 and MSP+4 or between MSP-2 and MSP-4. The choice of 2 and 4 is due to the fact that the width of an Arabic character can be at least 8 pixels in the case of isolated alif. It is, however, considered bad if it is greater than MSP+4 pixels or lower than MSP-4 pixels.

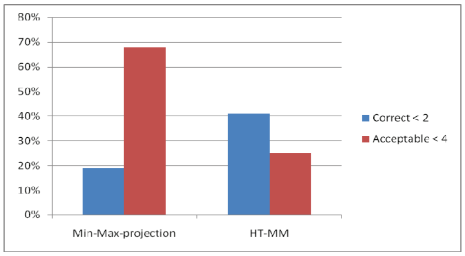


Figure 11. Extraction rate of the two segmentation methods by the use of segments positions

The result of fig.11 shows that the HT-MM method achieves an extraction rate of 41% but the Min-Max-projection method attempts only 19%. However, this upshot is not the same with an average pixel error<4. The main reason of the poor obtained result of Min-Max-projection segmentation method is that the objective of this method is the extraction of structural primitives and the estimation of their positions in the beginning, in the middle, in the end or isolated towards a global recognition by TNN, and not the detection of the exact limits of characters, in order to recognize them separately. The second raison is due to the reduced number of 1250 words used for evaluation. The third reason is the choice of the manual segmentation position. We can have several correct segmentation points in comparison to different point of view. So, when a SP evaluation method is chosen, an agreement about manual SP should be done. In fig. 12, we present two possibilities of manual segmentations in fig. 12(a) and 12(b). In fig. 12(c) the word is correctly segmented by HT-MM and in fig. 12(d) it is correctly segmented by Min-Max-projection method. As chosen in the figures, some characters will be considered correctly segmented by one of two manual segmentation and not by the other.

3) Recognition Rate criteria (RR): Recognition rate is evaluated using a classifier based on a Transparent Neural Network (TNN) [8]. This TNN is composed by 4 layers: structural primitive's layer, letter's layer, PAW's layer and recognized word's layer. It proceeds by global vision of structural primitives and local vision by Fourier descriptors for zones without structural description. These two visions need an estimation of boundaries zones, described by global structural primitives, or required a local processing. The quality of the primitive extraction has a direct impact on recognition. This is shown in fig. 13 where the recognition rate decreases from 92% by the use of manual and correct structural primitives extraction to about 17% by the use of automatic extraction rate based on Min-Max-projection segmentation method. The improvement of PAW's number and primitive's position improves primitive description and then RR. In fig. 13, the recognition rate, it increases from 17% to 70% by HT-MM segmentation method.

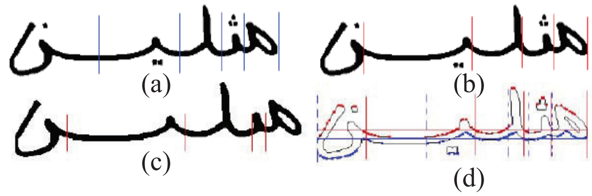


Figure 12. Impact of manual segmentation position on the evaluation of the segmentation rate

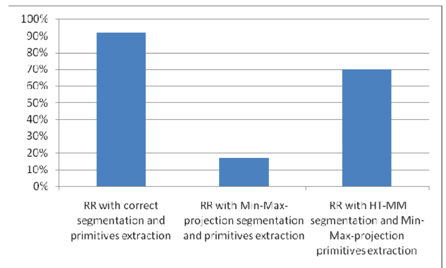


Figure 13. TNN Recognition rate obtained for each segmentation method

F. Comparison

After these three evaluation criteria, we can deduce that recognition rate is the most objective possibility to evaluate segmentation methods. The number of segments automatically extracted can be in some case equal to the correct number of characters but this number does not correspond to the correct segments.

The two main problems of SP criteria is firstly, the hard task of manual segmentation and associated positions organization and saving. Secondly, as we had explained above, the manual segmentation points should be taken into consideration by segmentation methods which would be evaluated in comparison to these manual points. The recognition rate needs only a recognition system depending from segmentation step, to be evaluated.

The main problems of comparison with other segmentation methods are in the first hand, the data used for evaluation is not the same. In the second hand, the segment evaluation method is not explained clearly. For instance, for the method based essentially on skeleton and proposed in [4] by Dinges and all, they use their own database for evaluation. For evaluation, they use a recognition step by a neural network. In [2], Elaiwat, uses the IFN/ENIT database for the evaluation of his segmentation method but he do not explain neither the set of data used nor the evaluation method.

IV. CONCLUSION

Two segmentation methods of handwritten Arabic script in characters are presented in this paper. The first method based on the analysis of global and local minima and maxima of the contour (Min-Max). The second segmentation method is based on the combination of HT and MM (HT-MM). To compare the efficiency of these two segmentation methods, we are based on three criteria's. The first criterion is the number of segmented characters (SN). The second criterion is the limits positions of characters in the word (SP). The third criterion is the recognition rate (RR) of Transparent Neural Network (TNN) based on structural primitives. Evaluation is done on the set-a of IFN/ENIT database. The detection rate of SP is 41% by HT-MM and 19% by Min-Max. The SN method achieved a detection rate lower than 30%. The RR method varied between 70% by HT-MM and 17% by Min-Max method. This RR attempts the 92% when the structural primitives and segmentation are correctly done. Indeed, the Min-Max-projection method as well as the HT-MM one needs the extraction of the baseline to distinguishes between diacritic and the main components of the word to be segmented. The quality of the baseline can have a great effect on the quality of segmentation step.

We should focus in the future in preprocessing and normalization steps of two segmentation methods in order to improve each one. A comparison to other developed segmentation method is also one of our perspectives on the same database. In the field of evaluation criteria we think that by a combination between SN and SP we can have more correct interpretation.

References

[1] H. A. Al Hamada, R.Abu Zitar, "Development of an efficient neural-based segmentation technique for Arabic Handwriting recognition", Pattern Recognition, vol 43, pp 2773–2798, 2010.

[2] S.Elaiwat, M.AL-abed Abu-zanona, "A Three Stages Segmentation Model for a Higher Accurate off-line Arabic Handwriting Recognition", World of Computer Science and Information Technology Journal, Vol. 2, No. 3, 98-104, 2012.

[3] T. Sari, M. Sellami, "State-of-the-art of off-line Arabic handwriting segmentation", *International Journal of Computer Processing of Oriental Languages*, vol. 20, pp. 53-73, 2007.

[4] L. Dinges, A. Al-Hamadi, M.Elzobi, Z.Al Aghbari and H.Mustafa, "Offine Automatic Segmentation based Recognition of Handwritten Arabic Words", International Journal of Signal Processing, Image Processing and Pattern Recognition, Vol. 4, No. 4 , pp 131-143, 2011.

[5] S. Wshah, Z. Shi and V.Govindaraju, "Segmentation of Arabic handwriting based on both contour and skeleton segmentation", *ICDAR*, pp. 793-797. 2009.

[6] S. Snoussi Maddouri, H. Amiri, A. Belaid, C. Choisy, "Combination of local and global vision modeling for Arabic handwritten words recognition", *IWFHR*, pp. 128-135, 2002.

[7] S. Snoussi Maddouri, H. El Abed, F. Bouafif Samoud, K. Bouriel, N. Ellouze, "Baseline extraction : comparison of six methods on IFN/ENIT database", *ICFHR*, 2008.

[8] C. G. Harris, M. Stephens, "A combined corner and edge detector", *Vision conference*, pp. 147-151, 1998.

[9] M. Peschwitz , S. Snoussi Maddouri, V. Margner, H. Amiri. "IFN/ENIT-Database of handwritten Arabic words", *CIFED*, pp. 129-136, 2002.