Touching Component Segmentation for Arabic Manuscript Recognition

Nabil Aouadi and Afef Kacem Echi
Latice Laboratory, Ensit University,
Tunis, Tunisia

Abstract: *This Segmentation of manuscripts into text-lines and words is an important step to make recognition systems more efficient and accurate. One of the problems making this task crucial is the presence of touching components which are connections between characters of successive text-lines, words of the same text-line or characters of a word. This work proposes an automatic system for Arabic manuscript recognition. The proposed system is based on a stochastic model of type HMM (Hidden Markov Model). First, it segments the manuscript into text-lines and words while solving the touching component problem using the shape context descriptor. Then, it extracts some structural features from word images and trains a Markovian classifier to recognize them. The performance of the proposed system is assessed using samples extracted from historical handwritten documents. The obtained results are encouraging. We achieved an average rate of recognition of 87%.*

Keywords: *Text-line segmentation, touching component, shape context, arabic handwritten word recognition, structural features, HMM*

Received August 1, 2015; Accepted September 25, 2016

1. Introduction

The need to transliterate large number of Arabic manuscripts into machine readable form has motivated many works on off-line recognition of Arabic script. But, a lot of problems are posed for systems of Arabic handwriting recognition such as: segmentation, skew angle of text-lines, overlaps, ligatures, spaces between words also a diacritic may be placed above or below the body of the character and change the meaning of the word, etc. This work is an attempt to overcome some of these problems. The separation of overlapping and touching text-lines within handwritten Arabic documents is still a hard task. The difficulty rises from the characteristics of the handwritten documents especially when they contain touching components. A touching component (TC) is the connected component which is produced when the adjacent characters touch each other. That's some parts ofthe characters are connected horizontally/left - right or vertically/ up-down. These ambiguously related components can be found in rich text or old documentsand theyoccurred due to presence of noise in images, writing style variation, low ink quality and aging, etc. (see Figure. 1). The segmentation of TCs means separation of the connected components into individual characters.The connection generally occurs between consecutive text-lines or words of the same text-line or even between characters of a word.In this work, we propose a complete system for Arabic manuscript recognition. It consists of four main steps: 1) TC extraction and segmentation based on the most similar model that approximates the TC, 2) Text-line and word extraction, 3) Structural feature extraction from word images and 4) Word recognition using a Markovian classifier. The paper is organized as follows; in section 2, we present some related works. In section 3, we detail the different steps followed by the proposed system. Experimental results are reported in section 4 and some conclusions are drawn in section 5.

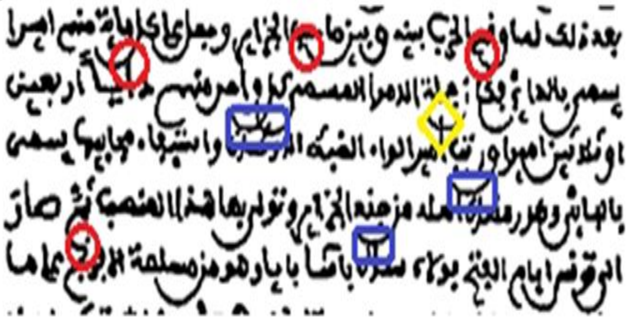


Figure 1.Samples of TCs in Arabic Manuscripts

2. Related Works on Offline Handwritten Word Recognition

This section deals with related Off-Line word recognition in Arabic handwriting. Afteranalyzingdifferent approaches (On-Line / Off-Line / Global / Analytical) for the automatic recognition of Arabic script.A. Amin [1], concludes that fundamental problems that affect the performance of an Arabic handwriting recognition system are: diacritics, styles writing, pseudo-words. In Mahjoub et al. [2], a new system for an offline handwritten Arabic word recognition based on coupled HMMs was proposed. In their system, each handwritten word image was transformed into two sequences of feature vectors that would be the observations to be given to a Dynamic Hierarchical Bayesian Network model. This system was tested on the IFN/ENIT database and achieved a recognition rate of 67.9%. In [3], AlKhateeb et al. propose a new method for offline word recognition using HMM by extracting intensity features from each word. Then, a combined system based on the HMM classifier was developed using these features for classification. Experiments were carried out using the IFN/ENIT database and results were very promising.Masmoudi and Amiriin [4], extract occlusions, empty spaces and diacritics from each word image. Once the word has been represented as a feature vector, its class determines the highest probability of the observed pattern. Method achieved recognition rate about 96%. In this study, S. Masmoudi and H. Amiri note that, the unique problem that affects performance of such system is that when the size of the vocabulary increases. According to Jayech et al [5], two major problems in Arabic handwriting recognition system are considered: segmentation and recognition. In their paper, authors present a methodology that segments a word and recognizes it. Recognition was based on dynamic hierarchical Bayesian networks preceded by a feature extraction algorithm that extracts the character and sub-character features from the segmented word. Tests have been done on a corpus extracted from the IFN/ENIT database. The recognition rate is about 81.14%.

3. Proposed System

The proposed system, described here, has been built on a subset of the Tunisian National Archive collection. We are interested by documents that are from the 19th century and correspond to enumeration registers at Tunisian protectoral period. Figure 2(a) displays a small portion of this manuscript: a set of lines composed of list of names. These documents are written in one author's hand. Our system segments these manuscripts into text-lines and recognizes words composing them. Below is a description of the different steps followed by our system.

3.1. Preprocessing

The used manuscripts were binarized using several methods. We found that the best result is obtained using Sauvola's method [6]followed by a morphological operation to restitute thinned characters (see Figure 2).

3.2. Touching Component Extraction

Arabic alphabet is composed of 28 letters and 21 of them have an ascender or a descender which causes touching or overlapping between words. Thus, TCs in Arabic documents can happen when the inter-lines spacing is small or when we use calligraphy with big jambs. To extract these TCs, we applied the Ouwayed and Belaïd's method [7]. But, for connections that appear between words or characters of the same text-line, we propose a new extraction method based on curve convexity analysis. The proposed method consists of three main steps: i) Baseline extraction, ii) Text-line Skeletonization and iii) Junction point location.

3.2.1. Extraction of TCs Between Text-Lines

In Ouwayed and Belaïd [8], text-lines are a sequence of connected components belonging to the same alignment. TCs were detected when connected components run simultaneously into two adjacent text-lines (see Figure. 3(a)). First, authors extract text-line skeleton using Zhang and Suen thinning algorithm as described in [9] (see Figure. 3(c)). Then, they look for intersection points **Ip** of each connected component near the minima axis (valley in horizontal histogram projection profile) between two text- lines (see Figure. 3(b)). The TC is the connected components centered on **Ip** (see Figure. 3(d)).

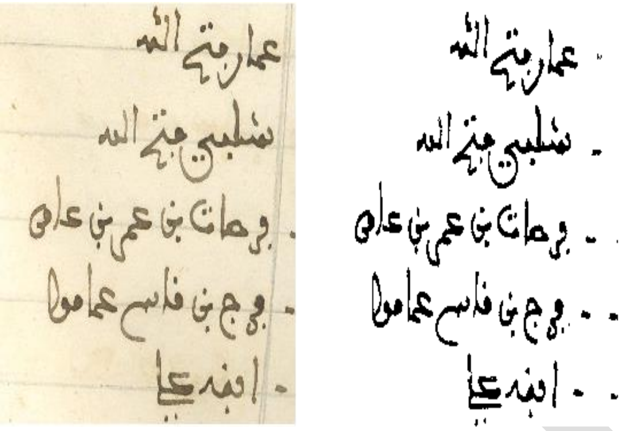
3.2.2. Extraction of TC Occurring in the Same Text-Line

In Arabic handwriting, parts of words in text-lines are connected at terminal letters under baseline. To detect these TCs, the text-line lower part skeleton is extracted using Zhang and Suen thinning algorithm [8]. The skeletonized image is formed by several branches of thickness equal to one pixel around the intersection point. Junction points are intersection points that have at least three neighbor pixels while TCs are connected components around them. A junction point, resulting from an overlap between two disjoint letters, is usually associated to a convex curve turned to right and end up at the baseline. Baseline is quite tricky to locate especially in case of Arabic script which is, in contrast to the Latin script, has not major accumulation of black pixels in a line. This is mainly due to letter extensions or horizontal ligatures. We referred to line support in the used manuscripts, used by the author to write words which perfectly coincide with baseline.To avoid invalid junction points which do not lead to real TCs, only large branches were retained (ignoring diacritical components and noise level (see Fig. 4). For these branches, we analyze their convexity [19]. In case where a concave segment in some valid branches exists, we will consider that the branch $B=\{(x,y)\in\mathbb{R}^2\}$ contains a real TC, only if there is a sufficient number (a threshold equals to 310) of pairs (x,y) that satisfies convexity equation (see Equation 1).

Let X be a convex set in a real vector space and let B:X→ℝ, be a function, B is convex if:

$$\forall\ (x,y)\ \text{couples of points} \in X\times X \text{ and } \lambda \in\ ]0,1[$$
$$B(\lambda x + (1-\lambda)^* y) \le \lambda B(x) + (1-\lambda) B(y) \qquad (1)$$

To evaluate the performance of TC detection, we tested our system on 1500 text-lines. We achieved a rate of 95% for detecting the main junction point which corresponds to real TCs occurring between words in the same text-lines, while a rate around 94% for TCs between words of two successive text-lines. Most errors are due to the fixed threshold and skeletonization result.



(a) Original image

(b) Sauvola's method followed by morphological operation.

Figure 2.Binarization of the documents using several methods.

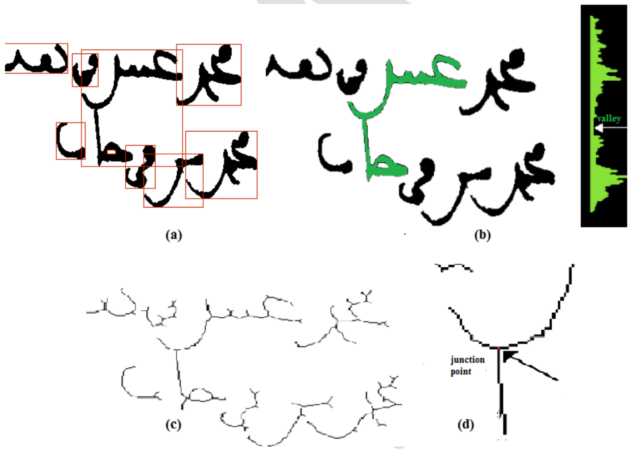
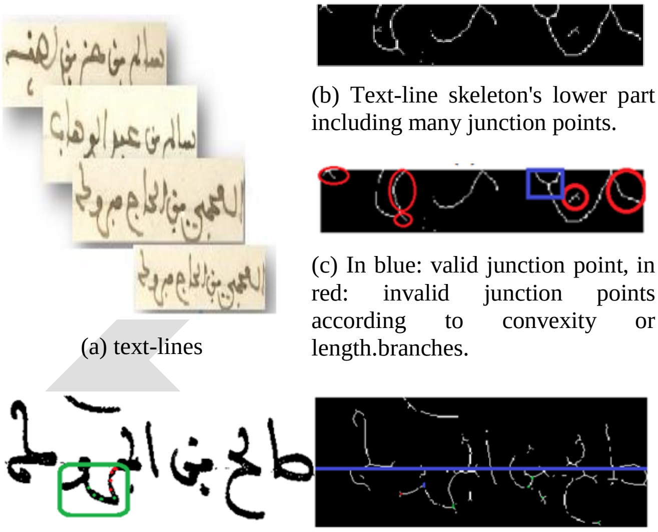


Figure3.Extraction of TCs that occurring between words of two consecutive text-lines.

3.3. TC Segmentation

To segment a localized TC, we approximate it to one of the models, stored in a dictionary with their known segmentation (model's part A and model's part B belonging respectively to the first and the second character). Thus, there are two fundamental steps before TC can be segmented: 1) a recognition step to find the most similar model for an input TC and 2) an approximation step to estimate the transformation aligning the selected model to the TC. Finally, we adjust the midpoints of the most similar model's parts to segment the TC.



(a) text-lines

(b) Text-line skeleton's lower part including many junction points.

(c) In blue: valid junction point, in red: invalid junction points according to convexity or length.branches.

(d) An example of a valid TC in a branch with convex and concave appearance

Figure 4.Semi-convexity branches analysis in text-lines skeleton's lower part.

3.3. Recognition and Transformation Step

Recognition and transformation precede the segmentation stage. Their object is to find a similar connection already segmented and estimate the transformation aligning it to the input TC. The system will be trained with several types of touching components called models and a segmented connection database is constructed. These models can be saved without any hierarchical structure or can be organized into levels with a representative element for each group. In the first case the research is a browsing of the whole training set but with a defined structure that can be achieved using clustering algorithms [9], the research is more intelligent and the performances are ameliorated especially the treatment time. Another property of this set is the dynamism, thereby it can be static (fixed on a training phase) or dynamic so it can be extended by some of the input connections having a satisfying segmentation result. To find the appropriate model, we compare the input connection to elements of the training set and we select the most similar. In this work, our comparison is based on the shape matching method proposed by Belongie[10]. It covers both stages of recognition and transformation since, similarity is computed by solving the correspondence between shapes with shape context descriptor and then estimating the aligning transform with the Thin Plate Spline (TPS) function [11].

3.3.2. Segmentation Step

In this stage, we dispose of the most similar model for a given TC and the estimated TPS transformation parameters that map this model on the TC. We exploit the correct segmentation of the model and their midpoints to segment the TC. For more details, see [12, 13, 14, 15]. Figure 5 and 6 show respectively some TCs and text-line segmentation.

3.4. Structural Feature Extraction

Generally, there are two types of features: statistical and structural. Structural features are intuitive aspects of writing, such as loops, branch-points, end-points and dots. Statistical features are numerical measures computed over images or regions of images. They include pixel densities, histogram of chain code directions, moments and Fourier descriptors, but they are not limited to them. The most important statistical features, used for word representation, are derived from distribution of points. We believe that words can be represented by structural features with high tolerance to distortions and style variations. This type of representation may also encode some knowledge about word structure or may provide some knowledge as to what sort of components make up that word. Extracting structural features that represent words is a difficult task, because of the high degree of variability and imprecision. In this subsection we propose a feature extraction method for handwritten Arabic word recognition. The main objective is to maximize the word recognition rate, using some structural features such as loops, jambs, stems, legs and diacritics.

According to Amin in [1], Arabic manuscripts reveal the complexity of the task, especially for the features choice (discontinuity of the writing, multiple connections of sub word, complex ligatures, etc.). In Azizi et al. [16], global structural features (descenders, ascenders, unique dot below the baseline, unique dot above the baseline, two dots, number of connected components, etc.) combined with statistical features (density measures, word subdivisions), could be efficient to represent words. Parameters such as lower and upper convergence of baselines are used, by El-hajji et al. [17] to derive a subset of baseline dependent features.

In our feature extraction algorithm, we proceed without any word segmentation. It is about to detect the presence of letters without delimiting them and thus have a global vision of words while avoiding word segmentation problems. We first pre-process extracted words in order to remove noise. To extract global and structural features, we consider their position in the word (at the beginning, in the middle or at the end, in the upper, central or lower bands) as shown in Figure 7.

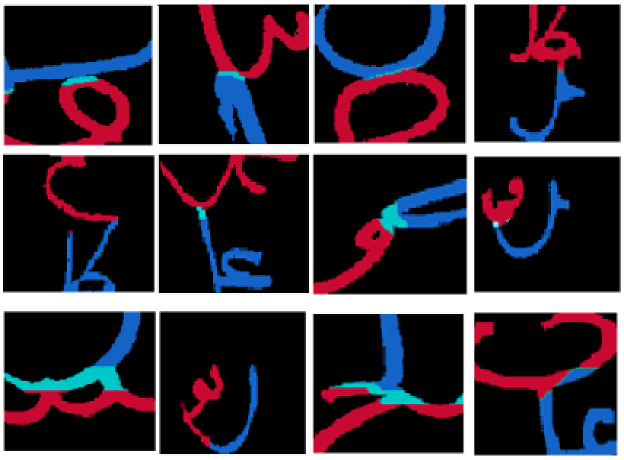
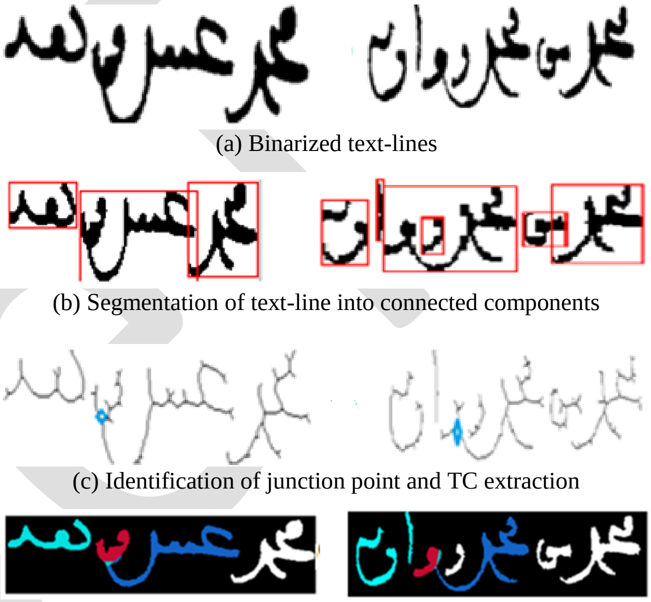


Figure 5.Some TCs segmentation results.



(a) Binarized text-lines

(b) Segmentation of text-line into connected components

(c) Identification of junction point and TC extraction

(d) text-line segmentation into parts of Arabic words (PAWs) based on TC segmentation result.

Figure 6.Text-line segmentation into parts of Arabic words (PAWs) based on TC segmentation result.

To extract structural features from word image, the word needs to be partitioned into three bands: the upper, the central and the lower bands. The central band is delimited by the upper and the lower lines. These lines are located using the baseline position which divides word image into inferior and superior parts. For these parts, we respectively compute the upper and the lower bands. 50% of the superior part and 30% of the inferior part are respectively considered for the upper and the lower bands. This choice is made because: 1) letter stems are generally higher than their legs, 2) words in the handled manuscripts are written by a single author and 3) the height of the letters, without stems, does not exceed 50%ofthe upper band of the word image. Afterwards,structural features(stems, legs, diacritics and loops) are respectively extracted from the upper, the lower and the central bands as shown in Figure 8.

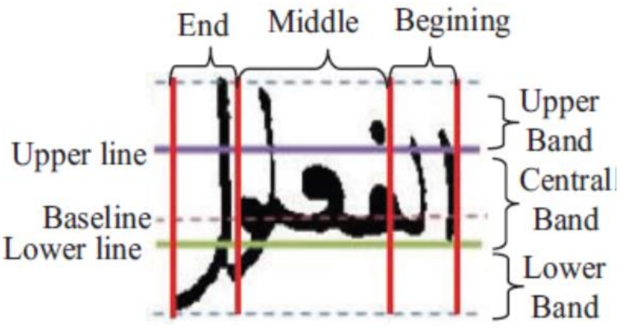
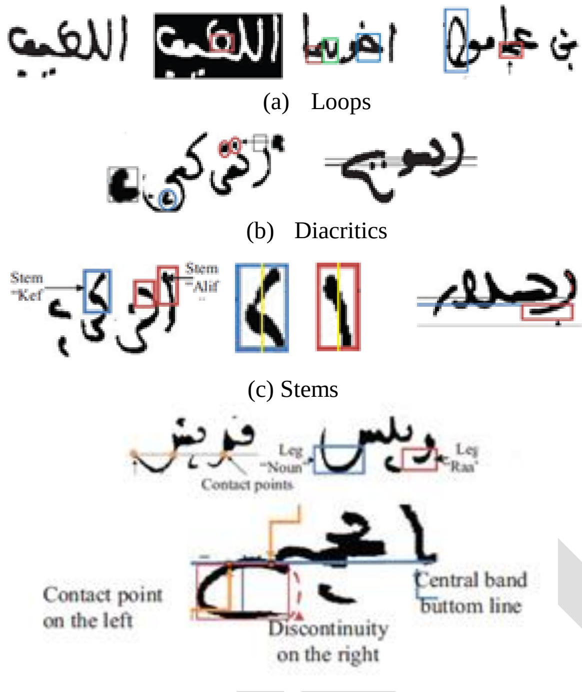


Figure7.possible positions of features extraction.



(a) Loops

(b) Diacritics

(c) Stems

(d) Legs

Figure 8.Different feature type extractions.

Dividing then words into three zones, from right to left, serves to classify extracted features according to their position in the word: in the beginning (the first quarter), in the middle (the second and the third quarters since Arabic word is generally elongated in the middle) and at the end (the last quarter) of the word. Word description is then performed from right to left as a sequence of structural features gathered from each band. For more details about structural feature extraction, see [18]. Table 1 presents all global structural features we extracted for a given word.

To evaluate features extraction results, we computed the Levenshtein distance, or edit distance, which is a string metric for measuring the amount of difference between two sequences. This distance is defined as the minimum number of edits needed to transform one sequence into the other, with the allowable edit operations being insertion (case of feature extracted in superfluous), deletion (case of not extracted feature), or substitution (case of not correctly extracted feature) of a single feature. In Table 2, E, T and D respectively refer to sequences of extracted features and truth description features and the Levenshtein distance.

Table 1.Extracted structural features.

Description	Code	Description	Code
Loop at the Beginning of the word	LB	Stem Alif at the Beginning	SAB
Loop in the Middle of the word LM	LM	Stem Alif at in the Middle	SAM
Loop in the End of the word LE	LE	Stem Alif at the End	SAE
One diacritic Point Up at the Beginning	1PUB	Stem Kef at the Beginning	SKB
Two or three diacritic Points Up at the Beginning	2PUB	Stem Kef at in the Middle	SKM
One diacritic Point Up in the Middle	1PUM	Stem Alif at the End	SKE
Two or three diacritic PointsUp in the Middle	2PUM	Stem Kef at the Beginning	LNB
One diacritic Point Up at the End	1PUE	Stem Kef at in the Middle	LNM
Two or three diacritic Points Up at the End	2PUE	Stem Kef at the End	LNE
One diacritic Point Down at the Beginning	1PDB	Leg "Noun" at the Beginning	LRB
Two or three diacritic Points Down at the Beginning	2PDB	Leg "Noun" in the Middle	LRM
One diacritic Point Down in the Middle	1PDM	Leg "Noun" at the End	LRE
Two or three diacritic Points Down in the Middle	2PDM	Leg "Ra" at the Beginning	LHB
One diacritic Point Down at the End	1PDE	Leg "Ra" in the Middle	LHM
Two or three diacritic Points Down at the End	2PDE	Leg "Ra" at the End	LHE

Table 2. Example of features extraction results for some words

Word image	E	T	D
	SAM, SAB, SAB, LM,LRE, 1PDE, 1PDE,1PDM, 1PDB	SAM, SAB, SAB, LM, LRE, 1PDE, 1PDE, 1PDM, 1PDB	**0**
	LE, LHE, LNB, 1PDM, 1PDB	LE, LHE, LNB, 1PDM, 1PDB	**0**
	SAM, 2PUE, 2PUE,**1PDM**	SAM, 2PUE, 2PUE,**2PDM**	**1**
	SKM, SKB, LRM,LRB, 2PUM, 2PDM,1PUB	LRM, LRB, 2PUM, 2PDM, 1PUB	**2**

Table 3 displays evaluation results of structural feature extraction using two databases: personal names, extracted from registers of the national archive of Tunisia, and Tunisian city names from the public database IFN-ENIT [19].

Table 3.Evaluation result

Data set	Similarity rate
Personal names (116)	0.89
IFN-ENIT (534)	0.78

As shown, in Table 3, despite the difficulties specific to the Arabic and ancient handwriting of personal names, the extraction results are better than those concerning Tunisian city names of the IFN-ENIT database. This can be explained by the different writing styles as Tunisian city names are written by several people while registers, we deal with, are written by the same person. Most of feature extraction errors can be attributed to the writing style and the poor quality of some data samples for instance:

- A group of two diacritic points can be written in the form of one or two related components. A group of three points may result in one, two or three related components depending on the writing style,
- A group of two or three letters is linked vertically at the beginning of the word,
- The same letter can be written in two different ways at the end of the word. Consequently the same word can be written in different ways.

3.5. Word Recognition

In this step, each word is modeled by a unique HMM. As Arabic is written from right to left, HMM's topology is sequential from right to left. We used 3 states where each state corresponds to a zone word or a column (beginning, middle and the end, see Figure 9). The training step is performed by the Baum-Welch algorithm. Word classification is performed based on discriminant model found among all the developed HMMs. Note that when dealing with higher length words, we will see more structural features. Loop in the HMM structure tells about staying in the same state, which means that the proposed HMM takes care of small and large words. Therefore, the HMM, could eliminate paths that are not promising early by computing probability from the first observations. Each word is modelized by a unique HMM. To recognize a given word, its image is tested on all HMMS and it is assigned to the HMM class which gives the highest probability as shown in Figure 10 for the word "محمد".

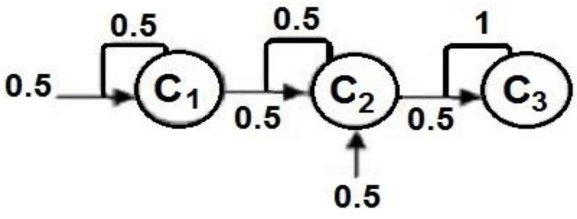


Figure 9. HMM structure.

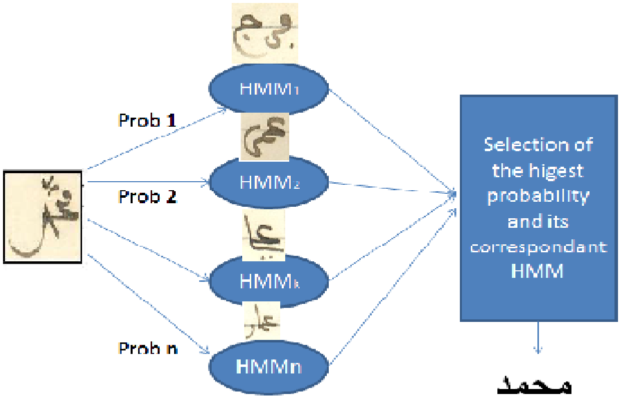

Prob 1
Prob 2
Prob n
HMM1
HMM2
HMMk
HMMn
Selection of the higest probability and its correspondant HMM
محمد

Figure 10. The proposed HMM structure for word recognition.

4. Experimentation

To evaluate our recognition system, we used a database containing 234 different namesfor training, while 120 words extracted automatically from text lines according to the segmentation process have been used for test. Table 4 shows the basic allocation for some names.

Table 4.Used database composition.

Word	Data base	Training set	Testing set
صالح	32	21	10
محمد	70	47	23
عمار	31	21	10
الباجي	3	2	1
باباي	2	1	1
عياد	10	7	3
قلعية	2	1	1
العربي	10	7	3
فرحات	26	7	19
قمرة	10	7	3

Table 5 shows the performance of the proposed system: achieve an average rate of recall and precision about 89%.

Table 5. Recognition rates.

Rappel	Precision	F-measure	Top1 (%)	Top2 (%)	Top3 (%)	Top4 (%)
0.87	0.85	0.86	87.25	95.64	97.48	98.25

These experiments have shown how robust are the proposed models since we dealt with old and high degraded manuscripts. The obtained results show that the proposed HMM has higher recognition rates when structural features extraction is correct.

5. Conclusion

For manuscript segmentation into text-lines and words, we firstly proposed to extract TCs between successive text-lines or words of the same text-line. Then, we segmented the TCswith reference to a set of models, stored in a dictionary with their known segmentation, using shape context descriptor, an interpolationfunction: the thin plate spline transformation (TPS) and the central points of the most similar model's parts. For word recognition, we extracted some structural features from word imageand trained a classic right-left HMM. Experiments are carried on a set of ancient Arabic manuscripts. The obtained results are encouraging: an average recognition rate is around 0.87. This shows the effectiveness of the TC segmentation method and opensopportunities in the field of text-line segmentation, feature extraction and selection, and classifiers for Arabic manuscript recognition.

References

[1] Amin, A.," Off line Arabic Character Recognition - a survey", *InProceedings of the 4th International Conference on Document Analysis and Recognition,* pages. 596–599, 1997.

[2] Essoukri Ben Amara, N., Ghanmi, N., Jayech, A.N., and Mahjoub, M.A., " Proposition d'un modèle de réseau bayésien dynamique appliqué à la reconnaissance de mots arabes manuscrits ", *Journées Francophones sur les réseaux bayésiens (JFRB 2012), 11-13 Mai, îles Kerkennah, 2012.*

[3] AlKhateeb, J.H., Al-Muhtaseb, H., Jiang, J., and Ren, J., "Offline handwritten Arabic cursive text recognition using Hidden Markov Models and re-ranking*",Pattern Recognition Letters, vol. 32, pp.1081– 1088, 2011.*

[4] Amiri, H., and Masmoudi, S., "Reconnaissance de mots arabes manuscrits par modélisation markovienne*",In Proceedings of 2`eme Colloque International Francophone sur l'Ecrit et le Document (CIFED'2000), Lyon, (France), July 3 – 5, 2000.*

[5] Essoukri Ben Amara, N., Jayech, K.,Mahjoub, M..A., and Trimech, N.," Dynamic Hierarchical Bayesian Network for Arabic Handwritten Word Recognition", *Fourth International Conference on Information and Communication Technology and Accessibility (ICTA), 2013.*

[6] Pietikainen, M., Sauvola, J., Seppanen, T., and Haapakoski, S., "Adaptive Document Binarization", *Proc. of 4th Int. Conf. On Document Analysis and Recognition, Ulm, Germany, pp.147—152, 1997.*

[7] Belaîd, A., and Ouwayed,N.,"Separation of overlapping and touching lines within handwritten Arabic documents", *Proc. of the 13th International Conference on Computer Analysis of Images and Patterns*, Mûnster (North Rhine-Westphalia), *Germany, pp.237-244, 2009.*

[8] Suen, C.Y., and Zhang, T.Y., "A Fast Parallel Algorithm For Thinning Digital Patterns",*Proc in Research Contributions Image Processing and Computer Vision, 1984.*

[9] Satu, SE , "Graph clustering survey". *In: Elsevier, 2007.*

[10] Belongie, S., Malik, J, and Puzicha, J., "Shape Matching and object Recognition Using Shape Context",*IEEE transactions on patternanalysis and machine intelligence, pp.509—522, 2002.*

[11] Bookstein, F. L., "Principal Warps:Thin-Plane Spline and the Decomposition of deformations", *In IEEE transactions on pattern analysis and machine intelligence (1989).*

[12] Aouadi, N., Amiri, S.,andKacemEchi, A.,"Segmentation of connected Component in Arabic Handwritten Documents", *Proc. of CIMTA, Elsevier, Vol. 10, pp.~738—746, 2013.*

[13] Aouadi, N., Belaîd, A., and Kacem, A., "Segmentation of Touching Component in Arabic Manuscripts". *Proc. of IEEE 14th Int. Conf on Frontiers in Handwriting Recognition, pp.452–457, 2014.*

[14] Aouadi, N., Belaîd, A., and Kacem, A., " A Recognition based Approach for segmenting Touching Components in Arabic Manuscripts". *Proc. of IEEE 13th Int. Conf ICDAR 2015 Nancy France, 2015.*

[15] Aouadi, N., Kacem, A., "A Proposal for a touching component Recognition based Approach for segmenting Touching Components in Arabic Manuscripts". *Pattern Analysis and Applications, Vol 19, Number 1, DOI 10.1007/s10044-016-0543-15, (2016).*

[16] Azizi, N., Farah, N., Sellami, M., and Tarek, K., "Arabic Handwritten Word Recognition Using classifiers Selection and features Extraction/Selection". *Recent Advances in Intelligent Information Systems, ISBN 978-83-60434-59-8: , pp. 735–742, 2009.*

[17] El-Hajj, R., Likforman-Sulem, L.,andMokbel, C.,"Arabic handwriting recognition using baseline dependant features and hidden Markov modeling", *In Proc. of International Confernce on Document Analysis and recognition (ICDAR 05), pp. 893 – 897, 2005.*

[18] Aouadi, A., Aouiti, N., Kacem, A., and Khemiri, A., "Système à base de MMC pour la reconnaissance de noms propres manuscrits Arabes ", *Proc. of CollogueInternational sur le documents électronique, Tunisia, 2012.*

[19] Website of IFN-ENIT database, http://www.ifnenit.com.

[20] Aouadi, N., Belaîd, A., and Kacem, A., "Localization Of Touching Letters In Arabic Handwritten Documents". *Proc. of IEEE 15th Int. Conf on Frontiers in Handwriting Recognition, 2016.*

Nabil Aouadi received the engineer diploma degree in 1993 in Computer Sciences from the Faculty of Science of Tunis and MS in 2008 from College of Science and Technology of Tunis. Currently he is a PhD student and a member of LaTICE: Laboratory for Technologies of Information and Communication at the High School of Sciences and Techniques of Tunis. He has authored over 12 articles in various national and international journals and conference proceedings.

Afef Kacem Echi received M.Sc. and Ph.D. degrees in Computer Sciences from the National School of Computer Sciences of Tunis in 1997 and 2001 respectively. Since 2000, she has been an assistant in the computer science department at the Faculty of sciences of Monastir, and was appointed Assistant Professor there in 2002. Dr.Kacem is a responsible member of the research area: Analysis and Recognition of Handwriting and document in LaTICE: Laboratory for Technologies of Information and Communication at the National High School of Engineers of Tunis. She has authored over 50 articles in various national and international journals and conference proceedings.