Word-Based Handwritten Arabic Scripts Recognition Using Dynamic Bayesian Network

Jawad H AlKhateeb

Faculty of Science and Information Techonlogy, AlZaytoonah University of Jordan
Ammnan Jordan

Jawad.kh@alzaytonah.edu.jo

***Abstract*— In this paper, multi-class classification system is of handwritten Arabic words using Dynamic Bayesian Network (DBN) is proposed, in which technical details are presented in terms of three stages, i.e. preprocessing, feature extraction and classification. Firstly, words are segmented from inputted scripts and also normalized in size. Then, features are extracted from each normalized word, where a set of new features for handwritten words is proposed based on a sliding window approach. The sliding window is moving across the mirrored word image. Finally, these features are then utilized to train a DBN for classification. The proposed system has been successfully tested on database (version v2.0p1e) consisting of 32492 Arabic words handwritten by more than 1000 different writers, and the results were promising and very encouraging.**

Keywords*— Off-line handwritten recognition; feature extraction; dynamic Bayesian Network (DBN); IFN/ENIT database.

I. Introduction

Handwritten recognition (HWR) plays essential roles in many applications, such as office automation, cheques verification, mail sorting, and a large variety of banking, business as well as natural human-computer interaction.

In general, this task can be divided into on-line based or off-line based systems. For on-line applications, the computer can trace the process of writing, hence the strength and sequential order of each segment when it is written can be recorded for recognition. While in offline applications, only a digital image is available hence it is more difficult. In this paper, we emphasize on offline recognition of handwritten Arabic text.

Arabic is written by more than 250 million people [1-3]. Unlike many other languages such as Latin, Chinese, and Japanese scripts which have been widely investigated, recognition of handwritten Arabic text remain a challenge task as there is very limited work reported. By nature, Arabic script is cursive, which makes its recognition rate lower than that of Printed Latin. Arabic text is written from right to left, and it has 28 basic letters in which 16 of them have dots. The dots can be one dot, two dots, or three dots. The dot(s) can be below or above the baseline and accordingly form different semantics. Therefore, detection of the baseline is one of the most important steps for Arabic text recognition.

Basically, there are two different categories of systems for the recognition of Arabic scripts, i.e. segmentation based and segmentation free based. In the first category, words need to be further segmented into characters or letters and these characters are then used for recognition, this is known as analytical approach. While, the second category does not need segmentation and the word images are taken as a whole on recognition, this is known as global approach where the recognition is globally performed. The global approach makes the recognition process simpler by avoiding the difficulty in character segmentation.

The recognition rate of a system depends on different factors. Two very important factors are the quality of input images and effective preprocessing. Once the sample image is acquired, pre-processed is required to enhance the signal for better performance. Pre-processing usually includes many relevant techniques like thresholding, skew/slant correction, noise removal, thinning, baseline estimation and segmentation of words. Except for baseline detection, we also focus on words segmentation.

The work for Arabic script recognition has started more than three decades ago. Al-Muallim and Yamaguchi [4] proposed a structural recognition technique for Arabic handwritten words which were segmented into strokes. The strokes were classified and combined into characters according to their features. However, their system showed a failure in most cases due to incorrect segmentation of words. Amin and Alsadoun [5] proposed techniques using binary tree to segment printed Arabic text into characters. Amin and Alsadoun [6] proposed recognition of hand printed Arabic characters using neural network. Abuhaiba [7] dealt with some problems in the processing of binary images of handwritten text documents, such as extracting lines from pages, which is found to be powerful and suitable for variable handwriting. Abuhaiba et al. [8] introduced a novel offline cursive Arabic script recognition system to recognize offline handwritten cursive script having high variability based on segmentation based system. In their system, a single component strokes were extracted. Khorsheed [9] presented a new method on off-line recognition of handwritten Arabic script, in which segmentation into characters is not required. The method decomposed the skeleton of the word into an observation sequence, and then a single hidden Markov model (HMM) with structural features is employed for classification. HMM is also used in Alma'adeed et al [10] for unconstrained Arabic handwritten word recognition. In Alma'adeed [11], a complete scheme for unconstrained Arabic handwritten word recognition based on a neural network is proposed.

In this paper, we proposed a system for handwritten Arabic word recognition, in which techniques are discussed in details in terms of preprocessing, feature extraction, and classification. Unlike traditional methods which recognize the whole image or letters, we conduct word-based processing for both effectiveness and flexibility. In addition, our method is validated using the IFN/ENIT database, where DCT features of words are inputted to train a neural network for classification.

II. IFN/ENIT DATABASE

Any recognition system needs a large database to train and test the system. Real data from banks or the post code are confidential and inaccessible for non commercial research. Although some work was conducted in Arabic handwritten words, but generally they had small databases of their own or the presented results on databases which were unavailable to the public. Consequently, there was no benchmark to compare the results obtained by researches. The IFN/ENIT database, available for free, is very important in this context as it has been used as a standard test set in such a context [12].

The IFN/ENIT database, available for free (www.ifnenit.com) for non commercial research, is very important in this context as it has been used as a standard test database in such a context [12]. The IFN/ENIT was published by the Institute of Communication Technology (IFN) at Technical University Braunschweig in Germany and the Ecole National d'Ingenieurs de Tunis (ENIT) in Tunisia. The database consists of 937 Tunisian town/villages names together with their postcode. In total more than 1000 different people were selected as writers to put their names. Also each writer was asked to fill one or more than one form with handwritten pre-selected names of Tunisian town/villages with the corresponding postcode. All the forms were scanned with 300dpi and converted to binary images. The images are divided into five sets so that researches can use some of them for training or testing, respectively.

III. THE PROPOSED METHOD

In this paper, a system for word-based handwritten Arabic text recognition is proposed where three main stages are included: preprocessing, feature extraction, and classification. The recognition is carried out to identify the word. This is done by preprocessing the word image resulting in a normalized image which removes all the variation in the images. Then, features are extracted using the sliding window approach. Finally, the DBN is applied in order to classify an unknown word by deciding to which class it belongs.

A. Preprocessing

The main goal for preprocessing is to enhance the inputted signal and to represent it in a way which can be measured consistently for robust recognition. Here the preprocessing stage involves scanning the paper document, noise removal, image binarization, line and word segmentation, and baseline estimation. These steps are strongly dependent on the quality of the document.

In the IFN/ENIT database, words are separated cropped out during the development stage [12]. Therefore, the only thing we need to do is estimation of the baseline and normalization. Despite of manually cropped words in the database, we have also investigated how to generally segment words and estimate the baseline and details can be found in [15].

Fig. 1 One sample images of a word (a), normalized image (b), and its Mirror Image (c).

As for normalization, it is essential to remove the variation in the handwritten images for consistent analysis and robust recognition. Among many algorithms proposed for this purpose, the skeletonization technique is the most popular one used, and in our system the normalization algorithm in [14] is employed. According to one sample images in binary format in Fig. 1(a), its normalized image and the mirror image are shown in Fig. 1(b) and Fig. 1(c), respectively. Please note pixel features will be extracted from the normalized image as described below.

B. Feature Extractions

Feature extraction is to remove the redundancy from the data and gain a more effective representation of the word image by a set of numerical characteristics, i.e. extracting most essential information from raw images. To cope with the characteristics that how Arabic texts are written, sliding windows/frames technique is widely used from right to left to extract features for off-line recognition [16]. In this paper, the sliding window technique used in speech recognition [16] has been adopted, yet applied to mirrored word image (MWI) after normalization in size to speed both training and testing process.

The pixel density features are extracted from the mirror image as shown in Figure 3(c). In this phase the feature vectors for each word mirror image is performed by applying a horizontal sliding window having the same height of the word image, a width of three pixels and a one pixel overlap. Since the size of each word image is 45×270, the word mirror image is divided into fifteen horizontal uniform frames; the sliding window is shifted across the word mirror image from left to right, and the feature vector is computed for each window strip as illustrated in Figure 2.

Fig. 2 Regions used for feature extraction and sliding window

Starting from the first pixel of the word, a sliding window is applied to the MWI to calculate the number of non-background pixels. The horizontal sliding window has the same height of the word image, three pixels in width with one overlapped pixel. When the sliding window is moving from left to right, as shown in Figure 3, each MWI is divided into fifteen uniform strips/frames horizontally. From these window strips, in total 30 features are extracted as follows.

Firstly, the first fifteen features (F_1 – F_{15}) are determined as average intensity of the pixels in each strip, i.e.

$$F_i = (\text{Avg.pixel intensity in the } i^{\text{th}} \text{ vertical area}) \mid i \in [1,15] \tag{1}$$

Then, average of these 15 features is used as the sixteenth feature F_{16}, which denotes overall mean intensity of the whole word image.

$$F_{16} = \sum_{i=1}^{15} F_i / 15 \tag{2}$$

Afterwards, the mean intensity of each consecutive pair of strips is extracted as fourteen additional features (F_{17}-F_{30}) as follows.

$$F_{i+16} = (F_i + F_{i+1})/2, \quad i \in [1,14] \tag{3}$$

In addition, several structure-like features are also extracted including number of connected regions n_r, number of connected regions (dots) below the baseline n_b, and number of connected regions above the baseline n_a. These are called structure-like features as to some degree they represent topological structure of the image.

The features extracted by the sliding window can be summarized as follows:

Algorithm FEATEXTWINDOWSLIDE
for k=1 to number of images
Img_in=Read the normalized image
Img_in=resize the Img_in into 45×270
Mirror_image=flip_left_right (Img_in)
Divide the Mirror_image into 15 horizonatal frame (width =3 pixels)
Apply Sliding window width = 3 and overlap = 1 from right to left

```
featX( 1:15,:) = featX( 1:15,:) /  9;
featX( 16  ,:) = featX( 16  ,:) /135;
featX(17:30,:) = featX(17:30,:) / 18;
end
```

C. Classification

Classification, as defined by Duda and Hart, means the assignment of a physical object or event to one several pre-specified categories. In our system, the definition means to assigning unknown Arabic word to one of the N classes representing the City names from the IFN/ENIT database. Classification methods vary and depend on the nature and the type of the extracted features. Our system uses the DBN architecture for recognition. Dynamic Bayesian networks (DBNs) are widely used in speech recognition. However, little research has been directed towards OCR.

An extension to Bayesian networks, a DBN is a method to model the probability distributions over semi-infinite collections of random variables, Z_1, Z_2,….. Likforman-Sulem et al [17] presented a new approach for off-line printed character recognition based on DBNs. Their model consists of coupling two HMMs in various DBN architectures. The image rows and image columns of the coupled HMMs were used as the main observations. Their system has been evaluated using various DBN architectures and achieved a recognition rate of 98.3% with the vertical HMM, and 93.7% with the horizontal HMM. However, when testing degraded letters the recognition rate went down as far as 93.8% with the vertical HMM and 88.1% with the horizontal HMM.

The DBN models have been considered to have two observation streams. The indices i=1, 2 represent the two streams. The variables X_i and Y_i represent the respective hidden state and the observation attributes in each stream. The processes modeled by DBNs have been assumed to be first order Markovian and stationary. This means that the model parameters are independent of *t* and the parents of any variable X_t^1 or Y_t^1 belongs to the time slices *t* or *t*-1 only. Figure 3 shows an example of an unrolled DBN for an observation sequence for an image whose width is 3 pixels.

Given a word image and its size denoted by [r, c] respectively, the unrolled DBN for such an image will be repeated c times which is 270 in this paper.

In general, the DBNs and the graphical models have several advantages over HMMs in terms of increasing the flexibility in the state-space factorization and structuring

In this paper, the application of DBNs to handwritten Arabic word recognition is investigated. To the best of my knowledge, this is the first time that a DBN has been created to carry out this type of recognition. The coupled HMMs architectures is represented as a single DBN [18].

Several coupled HMMs architectures can be constructed by adding directed edges between the two streams within the same time slice [18]. In order to enhance the influence of the vertical stream, the edges are directed from the vertical stream to the horizontal one. Experimentally, it has been found that the vertical HMM is more reliable than the horizontal one [18]. Due to the fact that both streams are synchronized at each time slice, it is required that both observation sequences in the proposed coupled HMMs architectures have the same length. Therefore, all the normalized word images are resized to be 45×270 .

In the coupled models, there are two states: vertical and horizontal states. The vertical states correspond to the column observations, while the horizontal states correspond to the row observations. Like the classic left right HMMs, a transition to the vertical state X_t^1 depends only on the preceding state value X_{t-1}^1. However, a transition to the horizontal state X_t^2 depends on both the preceding state value X_{t-1}^2 and the current vertical state value X_t^1. The observation dependences are expressed by the dependences between the horizontal and vertical states.

Figure 3 shows three main coupled architectures; the state coupled model ST_CPL, the general coupled model GNL_CPL, and the auto regressive coupled model AR_CPL. These models were suggested by Likforman-Sulem and Sigelle [18]. The ST_CPL model is obtained by adding the directed edges between the hidden state nodes of both vertical and horizontal HMMs as shown in Figure 3a. The GNL_CPL model is obtained by adding an edge from hidden states in the horizontal stream X_t^2 to the observation variables in the vertical stream Y_t^1 as shown in Figure 3b. The AR_CPL model is obtained by coupling both vertical and horizontal streams as shown in Figure 3c. More details about these three models can be found in [18].

In this paper, the AR_CPL model was chosen to be used since it is superior to other coupled models, having achieved the highest recognition rate [18].

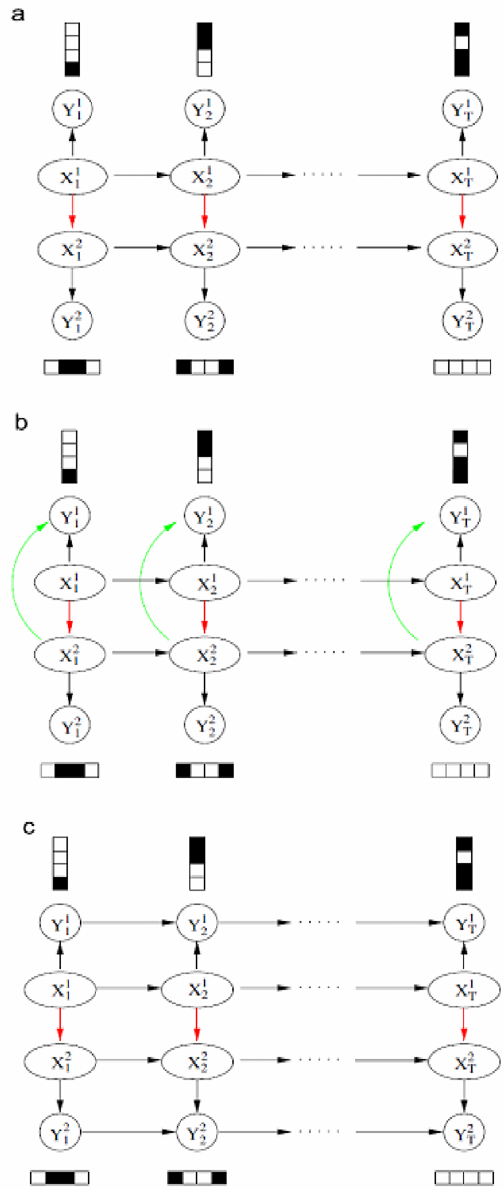

Fig. 3 Coupled architectures representing a single DBN

IV. EXPERIMENTAL RESULTS

In order to evaluate the performance of our recognition system, experiments are conducted on the IFN/ENIT database which contains 32492 Arabic words handwritten by more than 1000 different writers and divided into five sets a, b, c, d, and e [12]. Currently, only a small set of 500 words are used in our experiments.

We have considered a subset of the dictionary that occur more than 8 times in the whole database. Consequently, there was no benchmark to compare the results obtained by researchers. The IFN/ENIT database [12], available for free, is very important in such a context for benchmarking.

The DBN experiments have been conducted using the BayesNet Toolbox for Matlab [19], which provides the source code to perform several operations on BNs and DBNs. The system learns the DBN parameters by using the EM algorithm. The training set was created by randomly extracting five samples of data per class from the IFN/ENIT database. The set e is used for testing. A sliding window technique was used to extract the features to be used in the DBN.

To test the effectiveness of the DBN, the pixel features extracted using the sliding window were mapped into the DBN. However, the DBN requires a balanced database for training and testing which is not a property of the IFN/ENIT database. To overcome this problem, the training and testing samples from the IFN/ENIT were chosen randomly. The training and testing experiments were repeated five times and each time the training and testing samples were selected randomly. Table 1 summarizes the DBN recognition results.

TABLE I
DBN RECOGNITION USING AR_CPL MODEL

Exp.	**Recognition rate (%)**
1	65.46
2	67.86
3	65.32
4	66.27
5	67.86
Average	66.56

A new group of features using pixel density was mapped into the AR_CPL DBN classifier model. The DBN results are compared with six other systems tested with the same data set and conditions in the ICDAR 2005 Arabic handwriting competition. Figure 4 summarizes this comparison.

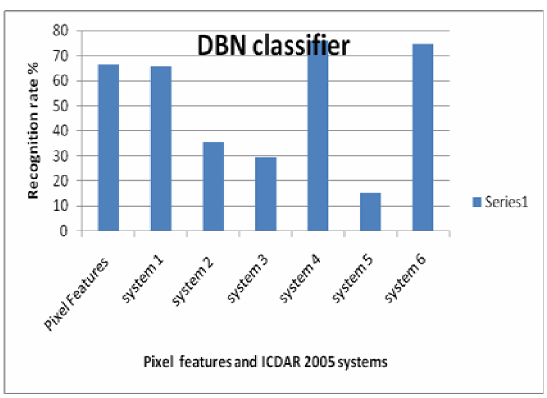

Fig. 4 Results from DBN and ICDAR 2005 systems compared

V. CONCLUSIONS

This paper proposed a system to classify Arabic handwritten word using DBN based on density features using sliding window technique. The system has been applied to the IFN/ENIT database of handwriting words written by different writers. Employing DBN as a recognition tool in handwritten Arabic word recognition system has shown good results which produces good rate of recognition rates. The proposed system relies on the DBN classification and density features. This system can be applied to other patterns with slightly adaptation.

There are three main drawbacks for the DBNs used in this research:

1) Long training time required for some applications is one of the main drawbacks of DBN.
2) The DBN is not suitable for an imbalanced data such as the IFN/ENIT database. To overcome this drawback, a random reading for the images in each folder is done.
3) The entire images have to be equal in size.

REFERENCES

[1] A. Amin. "Offline Arabic character recognition: The state of the art". *Pattern Recognition*, vol. 3, pp. 517-530, 1998.

[2] L. M. Lorigo and V. Govindaraju, "Offline Arabic handwriting recognition: a survey", *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 28, pp. 712-724, 2006.

[3] M.S. Khorsheed," Off-Line Arabic Character Recognition – A Review", Pattern Analysis & Applications, vol.5, pp. 31-45, 2002.

[4] H. Al-Muallim and S Yamaguchi. "A method of recognition of Arabic cursive handwriting". *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 9, pp. 715-722, 1987.

[5] A. Amin and H. Alsadoun, "A new segmentation technique of Arabic text.", *IEEE Trans. Pattern Recognition*, Vol.2, pp. 441-445, 1992.

[6] A. Amin and H. Alsadoun, "Hand printed Arabic Character Recognition System", IEEE Trans. Pattern Recognition, Vol. 2, pp536-539, 1994.

[7] I. S. I. Abuhaiba, M. J. J. Holt, and S. Datta, "Processing of binary images of handwritten text documents," *Pattern Recognition*, vol. 29, pp. 1161-1177, 1996.

[8] I. S. I. Abuhaiba, M. J. J. Holt, and S. Datta, "Recognition of Off-Line Cursive Handwriting," *Computer Vision and Image Understanding*, vol. 71, pp. 19-38, 1998.

[9] M. Khorsheed, "Recognising handwritten Arabic manuscripts using a single hidden Markov model", Pattern Recognition Letters, vol. 24, pp. 2235-2242, 2003.

[10] S. Alma'adeed, C. Higgens, and D. Elliman, "Off-line recognition of handwritten Arabic words using multiple hidden Markov models", Knowledge-Based Systems, vol. 17, pp. 75-79, 2004.

[11] S. Alma'adeed, "Recognition of Off-Line Handwritten Arabic Words Using Neural Network", proc. Of the Geometric Modeling and Imaging –New Trends, 2006.

[12] M. Pechwitz, S. S. Maddouri, V. M¨argner, N. Ellouze and H. Amiri, "IFN/ENIT - Database of Arabic Handwritten words",Colloque International Franco-phone sur l'Ecrit et le Document (CIFED), 2002, pp 127–136.

[13] M. Pechwitz, and V. Margner. "Baseline Estimation for Arabic Handwritten Words". *International Workshop on Frontiers in Handwriting Recognition*, pages 479-484, 2002.

[14] M. Pechwitz, and V. Margner. "HMM Based Approach or Handwritten Arabic Word Recognition Using the IFN/ENIT Database". Proc. Int'l conf. Document Analysis and Recognition, pp. 890-894, 2003.

[15] J. AlKhateeb, J. Ren, S. S. Ipson and J. Jiang: "Knowledge-based Baseline Detection and Optimal Thresholding for Words Segmentation in Efficient Pre-processing of Handwritten Arabic Text**"**. To appear in the Proc. 5th Int. Conf. Information Technology: New Generation (ITNG008) (2008)

[16] Husni, A. A.-M., Sabri, A. M., and Rami, S. Q., 2008. Recognition of off-line printed Arabic text using Hidden Markov Models. *Signal Process.*, 88**,** pp. 2902-2912

[17] Likforman-Sulem, L. & Sigelle, M. (2007) Recognition of Broken Characters from Historical Printed Books Using Dynamic Bayesian Networks. *Ninth International Conference on Document Analysis and Recognition, 2007. ICDAR 2007.*

[18] Likforman-Sulem, L. & Sigelle, M. (2008) Recognition of degraded characters using dynamic Bayesian networks. *Pattern Recognition,* 41**,** 3092-3103.

[19] Murphy, K. P. (2002) Dynamic Bayesian Networks: Representation, Inference, and Learning; PhD thesis. *Computer Science.* University of California- Berkeley