83International Journal of Multimedia and Ubiquitous Engineering
Vol.11, No.10 (2016), pp.83-92
http://dx.doi.org/10.14257/ijmue.2016.11.10.07

Arabic Handwriting Text Recognition Based on Efficient Segmentation, DCT and HOG Features

Alia Karim Abdul Hassan* and Mustafa S. Kadhm*

Computer Science Department, University of Technology/ Baghdad
must.salam@yahoo.com

Abstract

Writing in its different forms, printed and manuscript has always been a tool essential in human communication, and as ubiquitous in most areas of its operations. It is used to store and archive knowledge. Thereby, human has always developed techniques for sustainability across generations. Indeed, with the advent of new information technologies and electronics computers, and further increase the power of machines, automated processing. Besides, one of the important system is the handwriting recognition. Handwriting text recognition system based on efficient segmentation, DCT and HOG Features is proposed. The proposed system depends on the segmentation of the text into words only. Moreover, the system achieved best recognition accuracy 96.317% based on the used methods and SVM classifier.

Keywords*: Arabic Text, HOG, Feature Extraction, DCT, Segmentation*

1. Introduction

Handwriting recognition is a complex computer process, aims to translate a printed or handwritten text in a digitally encoded text, so understandable by a machine, by transmitting thereto the ability to read. The handwriting recognition relates more precisely all the related tasks with the mass processing of paper documents. Therefore, it focuses on the applications large repetitive with rather large databases namely: automatic processing of administrative files, automatic sorting of mail, the reading the amounts and bank checks, processing of addresses, the forms processing, interfaces without keyboard, the analysis of the written gesture, reading Heritage documents, indexing and search library archives of information in databases.

In order to develop pattern recognition methods and systems, a large amount of sample patterns are essential. In the same way as off-line character pattern databases such as IAM [1], AHDB [2], IFN/ENIT [3], and KHATT [4] and so on, have been playing significant roles for off-line handwriting recognition. Therefore, in this paper AHDB dataset has been selected for our proposed system. The dataset has the most popular Arabic words that written by many writers.

Furthermore, segmentation is the first step in handwriting recognition systems to make each text word clear and unique from the others. Preprocessing is the second step which helps to reduce the variability of handwriting by correct these factors and it will help to enhance the accuracy of and recognition. The third step in recognition system is the features extraction which extract a helpful information from the image text word to distinguish it from the other words. The last step of the recognitions is the classification which make the decision to sign the text word to its desired class then convert it into editable text. [5]

ISSN: 1975-0080 IJMUE

2. Proposed System

The proposed work for handwriting text recognition has several major steps. Each of the recognition step affect the accuracy and the performance of the recognition. First of all the input images grouping together to be as collection of documents which will pass through several process as shown in Figure1.

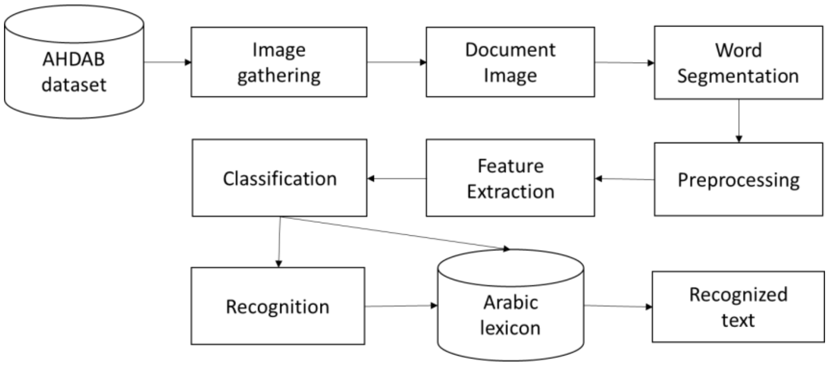

Figure 1. The Proposed Work

The proposed system involves several steps which are; preprocessing, feature extraction, classification and recognition. Besides that, each step has it benefits for the recognition process. Here the proposed method steps described in details:

2.1. Text Segmentation

Segmentation is a critical and decisive stage in several systems recognition. It is defined as the operation that seeks to break down a text image pseudo images of individual symbols. The result of this operation is a form isolated from an image and that could be a word or character. However, this separation is not always possible. Generally, the performance of the segmentation affects directly the reliability of the overall system. The input image to the system will be as shown in Figure2.

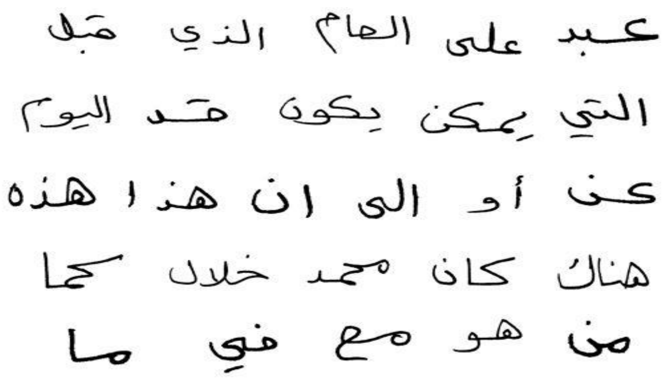

Figure 2. Input Document Image

TO perform the segmentation, two features in the Arabic handwriting text are considered. The first feature is the longest Arabic word has eleven characters and the shortest Arabic word has only two characters. The second feature is that, the distances between the words in the Arabic text are bigger than the distances between the sub-words which belong to the same word. Therefore, according to these features a proposed segmentation algorithm is taking place through draw a rectangle around each word in the text image and cutting each part to save it as an image. Algorithm 1 explains the proposed segmentation steps.

Algorithm 1: Text Segmentation (TS)
Input: Handwriting text image
Output: Segment text (word)

Step1: Read the input image
Step2: Apply Sobel filter on the image to detect the edge
Step3: Dilate the edge image
Step4: If there are hole in the image then
    Fill the holes
Step5: End IF
Step6: Scan the image and label all the objects from left to right
Step7: For each labeled object
Step8: Find the center of mass (centroid)
Step9: End For
Step10: Scan the image from left to right and draw a rectangle around each object
    based on the centroid
Step11: Draw the obtained rectangles around the objects in the original image
Step12: Crop the objects insides the rectangles and save each one as standalone image
Step13: End

By applying the Algorithm 1 there will be several handwriting segments (words). Each segment given a name to represent the text that appeared in the segment image then pass it to the preprocessing phase. Figure 3 present the main stages of the proposed segmentation method.

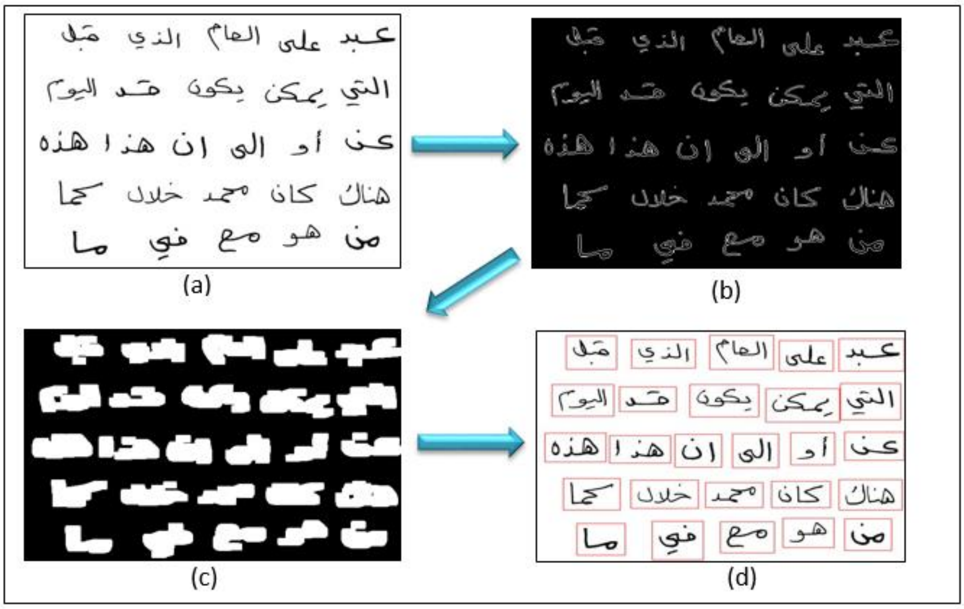

Figure 3. Text Segmentation (a) Original Image, (b) Edge Image, (c) Deleted Image, (d) Segmented Image

2.2. Preprocessing

Preprocessing consist of image transformations. The preprocessing goal is to facilitate the characterization of the image to be recognized either by eliminating the noise or reducing the amount of information to process to keep only the most information significant. Reducing the amount of information to be treated may be obtained from operations to bring the line thickness to a single pixel. The input to the handwriting text recognition system is a grayscale text image which has the Arabic word. During preprocessing the image converted to binary by thresholding method. The benefit of the thresholding is reducing the image diamantine to make it easy to process.

In the proposed work Otsu method [6] has been used to for thresholding purpose. After that, some noise appear due to the thresholding. 3X3 median used to remove undesired information from the binary image as shown in Figure4.

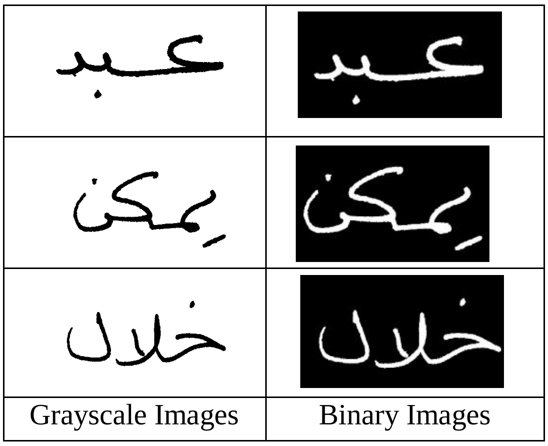

Figure 4. Image Thresholding

Black space around the written word in an image does not help in any recognition process. So this unwanted black space around the word was eliminated. To eliminate this black space, bounding boxes were used. From each side of the binary image, the first pixel of the written word was located. This produced four points which formed the boundaries of the bounding box. The black area around this box could then be eliminated using these four values as shown in Figure4.

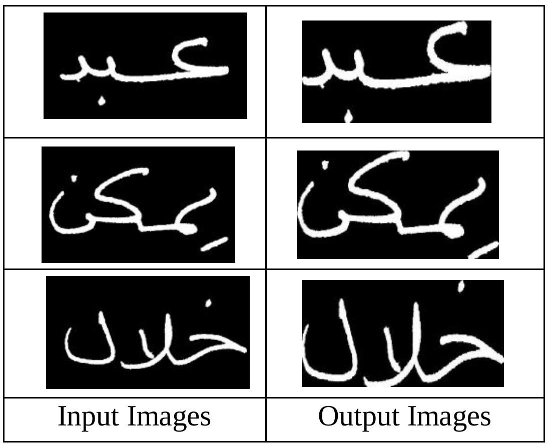

Figure 4. Black Space Elimination

The last step in preprocessing is the image normalization. AHDAB dataset has various image sizes. It important to make all the image in the dataset in the same size and make the recognition process fast. After testing several sizes (32x32, 64x64 and 128x128) the 128*128 size gave best recognition rate. Therefore, all the dataset images normalize into size 128*128.

2.3. Features Extraction

The objective of the feature extraction stage is to capture the most relevant and discriminate characteristics of the handwriting text image to recognize. The selection of good features can strongly affect the classification performance and reduce the computational time. In other words, it's possible to choose a set of features that denote the significant differences from one class to another, these selected features, consequently, result in an easier classification task. The features used must be suitable for the application and for the applied classifier. In this paper, two types of feature extraction methods has been used which are: Discrete Cosine Transform (DCT) and Histogram of Oriented Gradient (HOG).

2.3.1. Discrete Cosine Transform (DCT)

The DCT transformation matrix being Orthogonal, which accompanied by inversion method in order to return in space. So after making changes in the frequency domain, eliminating variations of the picture almost invisible to the human eye, one returns to a representation of pixels. Given an image f (i, j), its 2D DCT transform is defined as follows:

$$f(u,v) = \alpha(u)\alpha(v)\sum_{i=0}^{I-1}\sum_{j=0}^{J-1} f(i,j)\cos\left[\frac{(2i+1)}{2I}\right]\cos\left[\frac{(2j+1)}{2J}\right] \quad (1)$$

The inverse transform is defined by:

$$f(i,j) = \sum_{U=0}^{N-1}\sum_{V=0}^{N-1} \alpha(u)\alpha(v) f(u,v))\cos\left[\frac{(2i+1)u\pi}{2N}\right]\cos\left[\frac{(2j+1)v\pi}{2J}\right] \quad (2)$$

Where

$$\alpha(u) = \alpha(v) = \begin{cases} \frac{1}{\sqrt{2}} & for \quad u,v \quad = 0 \\ \frac{2}{\sqrt{2}} & for \quad u,v \quad \neq 0 \end{cases} \quad (3)$$

Due to its strong capability to compress energy, the DCT is a useful tool for pattern recognition applications. The DCT can contribute to a successful pattern recognition system with classification techniques such as Support Vector Machine and Neural Network [7].

In the proposed system the DCT applied for the whole image that produced from the previous phase. The output of the DCT is an array of DCT coefficients.

The features are extracted in a vector sequence by arranging the DCT coefficient in zigzag order, so that most of the DCT coefficients away from the beginning are small or zero. After testing the coefficients it found that the best number of DCT coefficients to represent the handwriting word as feature vector is the first 50 coefficients. The main steps to get the DCT features present in Algorithm 2.

Algorithm 2: **DCT_FEXT**
Input: Binary image
Output: DCT Features

Step1: Read the input image
Step2: Convert the image into gray
Step3: Compute DCT for the input image
Step4: Convert the DCT image into 1D array by zigzag order
Step5: Choose the first 50 DCT coefficients as a features
Step 6: Save the result features in 1D array
Step7: End

2.3.2. Histogram of Oriented Gradient (HOG)

Histogram of Oriented Gradient (HOG) was first proposed by Dalal and Triggs [8] for human body detection but it is now one of the successful and popular used descriptors in computer vision and pattern recognition. In order to extract the HOG features, the obtained binary images convert to gray. Besides, a proposed edge detection mask filter in Figure5 is used to find the image magnitude and gradient.

$$\begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \qquad \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix}$$

Figure 5. Proposed Edge Detection Filter

At each point, the approximations of the horizontal and vertical gradients and directions can be combined as in equations 2.9 and 2.10 to get an approximation of the gradient norm [16]:

$$g = \sqrt{g_x^2 + g_y^2} \tag{4}$$

Where g is gradient magnitude

$$\theta = \tan^{-1}\left(\frac{g_x}{g_y}\right) \tag{5}$$

Where θ is gradient direction

The image divides into small square cells 6×6 with bin=9 directions then computes the histogram of gradient directions based on the central differences. An example of getting the HOG illustrates in Figure 5.

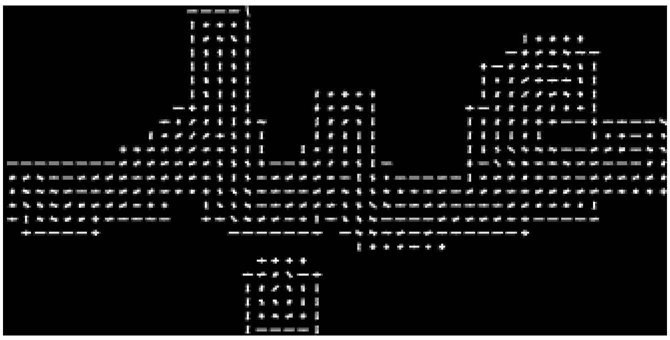

Figure 5. HOG

For improve accuracy, the local histograms have been normalized based on the contrast and this is the reason that HOG is stable on illumination variation. By applying this step, the total size of the feature set in the feature vector will be (6x6x9) = 324. The normalization factor can be one of the following:

$$\text{L2-norm: } f = \frac{v}{\sqrt{\|v\|_2^2 + e^2}}$$

$$\text{L1-norm: } f = \frac{v}{\|v\|_1 + e}$$

$$\text{L1-sqrt: } f = \sqrt{\frac{v}{\|v\|_1 + e}}$$

Where v is non-normalized vector containing all histograms in in given block, $\|v_k\|$ its k-norm for k= 1, 2 and e small constant (not influence the result).

2.4. Classification and Recognition

After the feature extraction, the major task is the make decision to classify the word to which class it belongs. There are various classifiers that can applied in text recognition. The most important and more effective classifier is Support Vector Machine (SVM).

2.4.1. SVM Classifier

The origin of support vector machines (SVM) dates back to 1975 when Vapnik and Chervonenkis have proposed the principle of structural risk and size for VC characterize the ability of a learning machine. At that time, this principle has not found a place and it does not yet exist a solid apprehended classification model to be usable. It took until the year 1982 that offers Vapnik SVM a first classification based on structural risk minimization [9-10].

SVM commonly used with linear, polynomial, RBF and sigmoid kernels. A multiclass SVM classification (libsvm) has been used in the proposed system [11] with different kernels of 1) linear, 2) polynomial, 3) RBF, 4) sigmoid and it achieves a very high recognition accuracy.

The final step is the recognition which is matching the selected class by the SVM with the character ASCII and find the desired word in the Arabic lexicon.

3. Experimental Results and Discussions

The proposed method is implemented using Matlab R2015a version, under windows7 64-bit Operating System, with RAM 6GB, CPU 2.50GHz core i5 and it achieved fast and effective results.

The proposed dataset has 2913 handwriting word images. Each word has 105 images written in different style. In the handwriting word recognition system 70% of the dataset used for training purpose (2044) and 30% for testing (896) and it achieved 96.3% recognition accuracy.

An important step to make the mathematical computing simple and fast a feature normalization (scaling) has been used to make the features ranges [0 1] by applying the following formula:

$$A' = \frac{A - \mathrm{Min}\,(A)}{\mathrm{Max}(A) - \mathrm{Min}\,(A)} \tag{6}$$

Where A is an original value, A' is the normalized value.

Moreover, in the proposed system SVM classification work with different kernels and each kernel achieved different accuracy. Besides that, there are an important parameters which make the SVM work perfectly.

The most important parameters in SVM are: cost(c) and gamma ( ). After many testing of the system the best values of the parameters was c = 8 and $\gamma = 0.0625$.

Table 1. Comparison between Different Kernels of SVM

SVM Kernels	Linear	Polynomial	RBF
Recognition Accuracy	92.63%	**96.317**%	91.5%

4. Conclusion

In this a paper, a proposed a high accurate handwriting text recognition system is present. The system use 70% of the dataset for training and 30% for testing and obtained high accuracy with SVM polynomial kernel. The high accuracy achieved by several factors starting from the efficient preprocessing stage with the use of Otsu method then with efficient proposed segmentation and feature extraction methods DCT , HOG and finally with more accurate recognition classifier.

References

[1] U.-V. Marti and H. Bunke, "A Full English Sentence Database for Offline Handwriting Recognition", Proceedings of the 5th International Conference on Document Analysis and Recognition (ICDAR 1999), **(1999)**, pp. 705-708.
[2] S. Al-Ma'adeed, D. Elliman and C. Higgins, "A Data Base for Arabic Handwritten Text Recognition Research", Proceedings of the 8th International Workshop on Frontiers in Handwriting Recognition (IWFHR 2002), **(2002)**, pp. 485-489.
[3] S. Mozaffari, H. E. Abed, V. Maergner, K. Faez and A. Amirshahi, "IfN/Farsi-Database: a Database of Farsi Handwritten City Names", Proceedings of the 11th International Conference on Frontiers in Handwriting Recognition (ICFHR 2008), **(2008)**, pp. 397-402.
[4] S. A. Mahmoud, I. Ahmad, M. Alshayeb, W. G. Al-Khatib, M. T. Parvez, G. A. Fink, V. Margner and H. E. Abed, "KHATT: Arabic Offline Handwritten Text Database", Proceedings of the 2012 International Conference on Frontiers in Handwriting Recognition (ICFHR 2012), **(2012)**, pp. 449-454.
[5] S. Mori, H. Nishida and H. Yamada, "Optical Character Recognition", (John Wiley & Sons, NY), **(1999)**.
[6] N. Otsu, "A threshold selection method from gray level histograms", IEEE Trans. on Systems, Man and Cybernetics, vol. 9, **(1979)**, pp.62-66.
[7] J. Jiang, Y. Weng and P. Li, "Dominant colour extraction in DCT domain", Image and Vision Computing, vol. 24, **(2006)**, pp. 1269-1277.
[8] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection", In CVPR, **(2005)**.
[9] C. Cortes and V. Vapnik, "Support-vector networks", Mach Learn, vol. 20, no. 3, **(1995)**, pp. 273-297.
[10] V. Vapnik, "The nature of statistical learning theory", Springer; **(1999)**.
[11] C.-C. Chang and C.-J. Lin, "Libsvm: A library for support vector machines", ACM T. Intell. Syst. Technol., vol. 2, no. 3, **(2011)**, pp. 1-27.

Author

Mustafa Salam, received the B.S. degrees in Software Engineering from Al-Mansour University College, Baghdad, Iraq 2009 and M.S. in Information Technology from University of Tun Abdulrazak, Malaysia 2012. During 2012-2013, he worked as Instructor in Ishik University, Turkey .Currently, PhD candidate at University of Technology, Baghdad, Iraq.