Available online at www.ejournals.uofk.edu

UofKEJ Vol. 8 Issue 1 pp.30-35 (February 2018)

UNIVERSITY OF KHARTOUM ENGINEERING JOURNAL (UOFKEJ)

Extended Feature Extraction Technique (Edge Direction Matrixes) For Online Arabic Handwriting Recognition

Mohamed Mosadag Albadawi, Hozeifa Adam Abd Alshafy

College of Computer Science and Information Technology, Karary University, Khartoum, Sudan.
(Email: m.mosadag@gmail.com), (Email: hozeifa.adam@gmail.com)

ABSTRAC: Recognition of Arabic handwriting has attracted the interest of researchers for many years. Until now it has been a challenging research area due to many issues. The feature extraction is an essential stage in the recognition systems of handwriting. The main idea behind this paper is to study EDMs (Edge Direction Matrixes) as a feature extraction technique for Online Arabic Handwriting Recognition. In this study, SUSTOLAH datasets will be used, in which datasets of online Arabic handwriting are presented in Sudan University of Science and Technology. In this paper, satisfactory results have been achieved, where the value of the correlation/regress coefficient for the differences between the variant handwritten characters is found to be -0.01322.

Keywords: *On-line Arabic handwriting, feature extraction, edge direction matrixes (EDMS).*

INTRODUCTION

In recent years, online handwriting recognition has become a central issue for Tablet PCs, PDAs screens, and hand-held computers. Online handwriting recognition is a task of mapping the handwriting words and characters to their text-based representations.

In contrast to other languages, Arabic handwriting presents a lot of challenges [1]. Handwriting of the same characters can vary in both their sizes and styles even when written by the same person. The context-based difference in one person's handwriting style presents another problem that faces Arabic handwriting recognition. Moreover, handwriting style has been affected by many factors such as the writer mood and the writing state.

Great research effort has been put in the area of Arabic handwriting recognition [4] [6] [7] [11]. Nevertheless, limited progress has been achieved in this area. The Arabic language has 28 alphabet characters. Some of these characters take different formats depending on the character location within a word (beginning, middle or end) and they may differ in size.

Moreover, there are many Arabic characters which can be associated with a single, double, or triple dots, or zigzag,

As shown in Figure 1, the stages of pre-processing, feature extraction, and classification enact the key phases for the handwriting recognition systems. The Feature extraction phase has an important role in the task of the recognition. It acts to take the input pattern and transforms it to some data. This data is then used in the later recognition phases in order to categorize known or unknown patterns.

The purpose of this paper is to study the EDMs as a feature extraction technique within the context of online Arabic handwriting recognition. We intend in this study to enhance the EDMs technique and then evaluate the enhanced technique with the use of some online Arabic handwriting characters which are taken from SUSTOLAH dataset. SUSTOLAH is a dataset of online Arabic handwriting which is presented in Sudan University of Science and Technology Online Arabic [10].

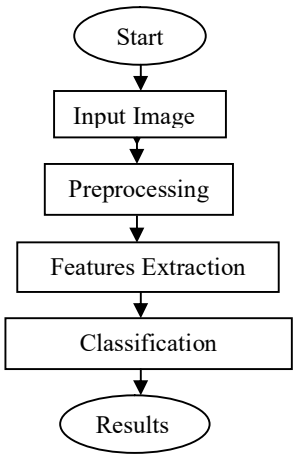


Fig. 1. The proposed Recognition System Flowchart

EDGE DIRECTION MATRIXES (EDMs) FEATURE EXTRACTION TECHNIQUE

EDMs technique is a general unvarying descriptor which uses linear relationships. As an input, the user employs a digitizer to handwrite a character on the screen, and then the coordinates of the handwriting points are kept in a serial. A bounding frame is drawn around the handwritten character and its features are taken for the purpose of the recognition.

Based on the pixel neighborhood connection, the EDMs technique employs the adjacent pixels which lay between 0 to 315 degrees. The technique was proposed in (2011), then later in (2013) it has been developed by a hybrid combination with Local Binary Pattern (LBP) [8] [9]. EDMs have two versions which are named as EDM1 and EDM2. Figure 2 (a) and (b) show the EMDs matrices which depict the neighboring relationship of the image pixels. These matrices are 3×3 matrices. The relationship involves the adjacent pixels corresponding to the angles of 0, 45, 90, and 135 degrees.

(a)

(x-1,y-1)	(x,y-1)	(x+1,y-1)
(x-1,y)	s(x,y)	(x+1,y)
(x-1,y+1)	(x,y+1)	(x+1,y+1)

(b)

135°	90°	45°
180°	Scoped Pixel	0°
225°	270°	315°

Fig. 2. Matrices of Neighboring Relationship in EDMs

EDM1 assigns values to the matrices locations, such that each of the matrices locations contains a value that measures the occurrence number for the relationship corresponding to the location. The following algorithm is used to the relationships occurrences of EDM1:

EDM1 Estimation Algorithm:

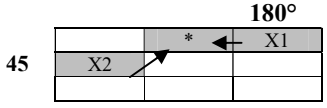
```
1. Procedure EMD1_Estimation (Area[n, m], EMD1[3, 3])
2. For i :=1 to n do
3. For j:=1 to m do
4.  Begin
5.  IF Area[i, j] = black THEN
6.  Begin
7.  EMD1[2, 2]  := EMD1[2, 2]+1
8.  IF (Area[i+1, j] = black)
9.  THEN
10. EMD1[2, 3]  := EMD1[2, 3]+1
11. IF (Area[i+1, j–1] = black)
12. THEN
13. EMD1[3, 1]  := EMD1[3, 1]+1
14. IF (Area[i, j–1] = black)
15. THEN
16. EMD1[2, 1]  := EMD1[2, 1]+1
17. IF (Area[i–1, j–1] = black)
18. THEN
19. EMD1[1, 1]  := EMD1[1, 1]+1
20. IF (Area[i–1, j] = black)
21. THEN
22. EMD1[1, 2]  := EMD1[2, 3]+1
23. IF (Area[i–1, j+1] = black)
24. THEN
25. EMD1[1, 3]  := EMD1[3, 1]+1
26. IF (Area[i, j+1] = black)
27. THEN
28. EMD1[2, 3]  := EMD1[2, 1]+1
29. IF (Area[i+1, j+1] = black)
30. THEN
31. EMD1[3, 3]  := EMD1[1, 1]+1
32. End
33. End
34. End EMD1_Estimation
```

EMD2 is a 3×3 matrix that describes the relationship between the black points in the handwritten curves. It is the second order of the relationship (whereas the EDM1 is the first order) which is determined as follows:

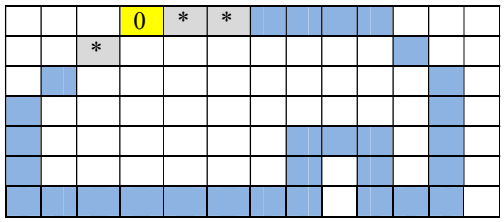
- Discerningly order EDM1. The smallest positions (curves) should firstly be assigned in the order when they have a same value.
- For each black point in the handwriting area,
- Find the relationship between its two neighboring black points, and then approach this relationship ship to the available ones.
- Increase the EDM2 position which corresponds to the relationship.

The first-order relationship (i.e., EDM1) is shown in Figure 3 (c) and the second-order of the relationship (i.e., EDM2) is presented in Table 1.

(a)



(b)



(c)

	135°	90°	45°
180°	5	11	5
	17	33	17
	5	11	5
	225°	270°	315°

Fig. 3. (a) The Two Neighbouring Pixels and (b) the Pattern and (c) it is EDM1

Table 1. The Second-order of the Relationship

Order	angle	value
1	0	17
2	180	17
3	90	11
4	270	11
5	45	5
6	135	5
7	225	5
8	315	5

EDM1 and EDM2 are used to estimate 14 features which characterize the handwriting images. PR, Hom, Weight, ED, ER, are some of these features.

1) PR: It is the feature of pixel regularity which explains the distribution of the handwriting directions.

$$\mathbf{PR(\theta)} = \frac{\mathbf{EDM1(x,y)}}{\mathbf{EDM1\ (2,2)}}$$

Where θ is an angle such as 45° or 135° and (x, y) is the angle related occurrences in EDM1

2) **Hom:** It the feature that presents the probabilities of the handwriting directions.

$$\mathbf{Hom(\theta)} \frac{\mathbf{EDM1(x,y)}}{\sum_{\mathbf{x,y-1}}^{\mathbf{x,y-3}} \mathbf{EDM1(x,y)}}$$

3) **Weight:** Compared with the total number of the handwriting points, this feature introduces the intensity of the handwriting boarders.

$$\mathbf{Wight} = \frac{\mathbf{EDM1(2,2)}}{(\sum_{\mathbf{x,y-0}}^{\mathbf{x,y-height\ width}} \mathbf{(boarder(x,y))}}$$

4) **ED:** This feature estimates the main form of the handwriting image, where this form is determined by the maximum occurrence of the handwriting curves.

$$\mathbf{ED} = \mathbf{max(EDM!\ (x,y))}$$

Such that the ED concerns only with the curves of 0°, 45°, 90° and 135°.

5) **ER:** This feature provides the ratio of a specific handwriting curve in relation to all directions in EDM2.

$$.\mathbf{ER(\theta *)} = \frac{\mathbf{EDM2(x,y)}}{\mathbf{EDM2(2,2)}}$$

EXTENDED EDGE DIRECTION MATRIXES (EEDMs) FEATURE EXTRACTION TECHNIQUE

We have developed a new version of EDMs which we named as EEDMs (extended edge direction matrixes). In our effort to produce EEDMs, we brought new formats and styles to EDMs. The main feature which differentiates EEDMs and EDMs is that: the EDMs work based on closer neighboring relationship of the pixels, while the EEDMs take a second neighboring relationship for the pixels in addition to the closer neighboring. Both EEDM1 and EEDM2 are presented in 5×5 matrix so as to hold the neighboring relationship style of EEDMs. Figure 4 (a) and (b) shows the general structure of EEDMs. We have also used two algorithms to determine the EEDMs. The first of these algorithms is a modified version of the EDM1 Estimation Algorithm, and the second algorithm is an enhanced version of the procedure that determines the EDM2.

(a)

(x+2,y-2)	(x-1,y-2)	(x,y-2)	(x+1,y-2)	(x+2,y-2)
(x+2,y-1)	(x-1,y-1)	(x,y-1)	(x+1,y-1)	(x+2y-1)
(x+2,y)	(x-1,y)	S(x,y)	(x+1,y)	(x+2,y)
(x+2,y+1)	(x-1,y+1)	(x,y+1)	(x+1,y+1)	(x+2,y+1)
(x+2,y+2)	(x-1,y+2)	(x,y+2)	(x+1,y+2)	(x+2,y+2)

(b)

135°	112.5°	90°	67.5°	45°
157.5°	135°	90°	45°	22.5°
180°	180°	Scoped Pixel	0°	0°
202.5°	225°	270°	315°	337.5°
225°	247.5°	270°	292.5°	315°

Fig. 4. (a) Locations of the Neighboring Pixels in EEDMs (b) Angles of the Neighboring Pixels in EEDMs

RESULTS AND DISCUSSION

We have selected a several isolated characters which are taken from SUSTOLAH, the datasets of online Arabic handwriting which has been presented in Sudan University of Science and Technology [10].We chose those characters randomly some of them are similar in their styles and outlook, and others are different (ا, أ, إ, ب, ت, س, ح) presented in Figure 5.

Fig. 5. The Different Shapes of the Selected Isolated Handwriting Characters

We have used MATLAB for the implementation and evaluation of the technique which is presented in this paper. We use the preprocessing part of the work which present in [11].After that we look two sequences of the handwriting direction (x and y), which are generated by the preprocessing stage, as inputs. Applying the EEDMs and EDMs algorithms, each one as a stand-alone to give us the occurrence of each angle orders. Lastly, we applied the previous equations which resulted in the features that we needed to extract.

After the application of these techniques is completed, the results are ready for assessment. It is then possible to evaluate the performance of the EEDMs technique.

The technique performance is measured by comparing the rates (R) of correlation and regress between the results of the chosen characters (EEDMs vs. EDMs).This is to assess the extent of the relationship values extracted from each correlation or downhill with the help of Excel program and using the Peterson equation:

$$R = \frac{N\Sigma xy - (\Sigma x)(\Sigma y)}{\sqrt{(N\Sigma x^2 - (\sum \Sigma x)^2)(N\Sigma y^2 - (\Sigma y)^2)}}$$

Whenever the correlation between the feature results of the different characters is less, which is good for us because it makes the mission easy for the classification phase and increased the percentage of recognition.

Table 2 presents the results of experiment using our EEDMs and Table 3 presents the results of experiment and tests using EDMs, Both on the same characters.

Table 2. The Experiments Results of the EEDMs on the Selected Characters

Characters	ا	أ	إ	ب	ت	س	ح
Weight							
	19.2115	14.0704	14.4783	9.8911	9.1651	10.6277	8.0565
Edge Direction							
	22	20	9	21	21	62	21
pixel regularity							
0°	0	0.0050	0.0060	0.0160	0.0210	0.0110	0.0210
22.5°	0	0.0040	0.0050	0.0060	0.0100	0.0070	0.0030
45°	0.0010	0.0120	0.0070	0.0090	0.0160	0.0240	0.0050
67.5°	0.0070	0.0020	0.0030	0.0020	0.0070	0.0190	0.0010
90°	0.0220	0.0200	0.0090	0.0210	0.0210	0.0621	0.0110
112.5°	0	0.0040	0.0080	0.0050	0.0100	0.0070	0.0050
135°	0	0.0030	0.0030	0.0100	0.0130	0.0080	0.0110
157.5°	0	0.0030	0.0020	0.0130	0.0050	0.0060	0.0090
Homogeneity							
0°	0	0.2778	0.4000	0.9412	0.6364	0.2200	2.3333
22.5°	0	0.1176	0.2632	0.1875	0.2273	0.0667	0.1765
45°	0.0345	0.4615	0.03500	0.3214	0.4211	0.2727	0.2941
67.5°	0.3182	0.0741	0.1500	0.0556	0.1591	0.2468	0.0370
90°	0	2	0.6923	0.7500	0.7500	0.9524	0.4400
112.5°	0	0.3636	0.7273	0.1282	0.2564	0.2800	0.1220
135°	0	0.2500	0.2308	0.2857	0.3611	0.3333	0.3333
157.5°	0	0.1429	0.1111	0.4194	0.1064	0.1429	0.3103
Edges regularity							
0°	0	0.0704	0.0870	0.1584	0.1927	0.1170	0.1694
22.5°	0	0.0563	0.0725	0.0594	0.0917	0.0745	0.0242
45°	0.0192	0.1690	0.1014	0.0891	0.1468	0.2553	0.0403
67.5°	0.1346	0.0282	0.0435	0.0198	0.0642	0.2021	0.0081
90°	0.4231	0.2817	0.1304	0.2079	0.1927	0.6596	0.0887
112.5°	0	0.0563	0.1159	0.0495	0.0917	0.0745	0.0403
135°	0	0.0423	0.0435	0.0990	0.1193	0.0851	0.0887
157.5°	0	0.0423	0.0290	0.1287	0.0459	0.0638	0.0726

Table 3. The Experiments Results of the EDMs on the Selected Characters

Characters	ا	أ	إ	ب	ت	س	ح
Weight							
	19.2115	14.0704	14.4783	9.8911	9.1651	10.6277	8.0565
Edge Direction							
	7	8	5	13	14	40	13
pixel regularity							
0°	0	0.0030	0.0030	0.0050	0.0140	0.0090	0.0130
45°	0	0.0010	0.0020	0.0060	0.0060	0.0050	0.0090
90°	0.0070	0.0080	0.0050	0.0130	0.0100	0.0400	0.0100
135°	0.0010	0.0080	0.0050	0.0060	0.0120	0.0160	0.0020
Homogeneity							
0°	0	0.1765	0.2500	0.2000	0.5000	0.1475	0.6190
45°	0	0.0526	0.1538	0.2500	0.1667	0.0769	0.3600
90°	7	0.6667	0.5000	0.7647	0.3125	1.3333	0.4167
135°	0.1429	0.6667	0.5000	0.2500	0.4000	0.2963	0.0625
Edges regularity							
0°	0	0.0423	0.0435	0.0495	0.1284	0.0957	0.1048
45°	0	0.0141	0.0290	0.0594	0.0550	0.0532	0.0726
90°	0.1346	0.1127	0.0725	0.1287	0.0917	0.4255	0.0806
135°	0.0192	0.1127	0.0725	0.0594	0.1101	0.1702	0.0161

The variation in **correlation/regress** in the comparison tests is shown in Table 4 which clearly shows that our EEDMs technique is better because it gives correlation rates that are much lower than the EDMs technique. That will positively affect the final decision of classification stage and improve the recognition capability which is a main goal behind this work.

Table 8. Comparison between EDM and EEDM in the Results Correlation Rates

	Characters In test	**EDMs**	**EEDMs**
Test 1	ا / أ	0.671432	-0.01322
Test 2	ا /س	0.947714	0.415714
Test 3	س / ح	0.42315	0.213679

CONCLUSION

This paper has investigated the EDMs as a feature extraction technique for the isolated characters of online Arabic handwriting recognition system. EDMs are extended to EEDMs.

EEDMs algorithm extracts the feature values through occurrence calculation. We compared them with features that we have pre-calculated from application of EDMs technique.

These results of comparison prove that the EEDMs method has resulted in better performance than EDMs. The technique is effective and efficient for the isolated characters.

REFERENCES

[1] M. A. Abuzaraida, A. M. Zeki and A. M. Zeki, "Difficulties and Challenges of Recognizing Arabic Text," in Computer Applications: Theories and Applications, ed Kuala Lumpur: IIUM Press Malaysia, 2011.

[2] Bataineh, B., Abdullah, S. and K. Omar. "Astatistical global feature extraction method for optical font recognition." Intelligent Information and Database Systems, 257-267, 2011.

[3] Amin, A. and Singh, S. "Recognition of handprinted Chinese characters using decision trees/machine learning C4. 5 system."Pattern Analysis & Applications, 1(2): 130-141, 1998.

[4] Amin, A. "Recognition of printed Arabic text based on global features and decision tree learning techniques." Pattern recognition, 1309-1323, 2000.

[5] Guo, H. & Zhao J. 2011. Research on Feature Extraction for Character Recognition of NaXi Pictograph. Journal of Computers 6(5): 947-954.

[6] Al-Taani, A. T. and Al-Haj, S. "Recognition of On-line Arabic Handwritten Characters using Structural Features." Journal of Pattern Recognigtion Research 5(1), 23-37, 2010.

[7] El-Sheikh, T. S. and Guindi, R. M. "Computer recognition of Arabic cursive scripts." Pattern recognition , 293-302, 1988.

[8] Naeimizaghiani, M., Abdullah, S. N. H. S., Bataineh, B. & Pirahansiah, F. 2011. Character Recognition Based on Global Feature Extraction. 1-4.

[9] Mohammmed a.talab, Siti norulhuda sheikh abduliah, Bilal bataineh, "edge direction matrixes-based local binary patterns descriptor for shape pattern recognition, "in IEEE, 2013, pp. 13-18.

[10] Huzeifa A. Alshafy.,Mohamed E .Mustafa , " Datasets for Online Arabic Handwriting " ,In Proceeding of the second International Conference on E-Applications in developing countries (ICEA 2011), Khartoum , Sudan, 27-29 Nov 2011.

[11] Huzeifa A. Alshafy.,Mohamed E .Mustafa , "HMM Based Approach for Online Arabic Handwriting Recognition " ,In Intelligent Systems Design and Applications (ISDA), 2014 14th International Conference on IEEE, Khartoum , Sudan, 27-29 Nov 2014.