Eigen Values Matrix for Recognition English Letters

ISRAA MOHAMMED HASSOON

Department of Computer Science, Collage of Science,
University of Mustansiriyah (UOM), Baghdad, Iraq.
*Corresponding author E-mail; morose262@yahoo.com

ABSTRACT

Reading skills or recognizability of letters and shapes are one of the human potential. The main objective in the design of any recognition system is to reduce the amount of data and get the perfect ratio of recognition. In this work ,eigen values matrix is used to recognize English letters. The process of recognize English letters (small letters , capital letters) in different sizes and different fonts that captured from scanner has been done by extracting topological features that study of qualitative properties of letters , then topological features used to build eigenvalues matrix.

Keywords: Topological features, English letters Recognition, Eigen values, Eigen values Matrix.

INTRODUCTION

Recognition process is an important task because it performed daily by human in the life and because of its important applications in various fields . The human eye can differentiate between letters even if they are largely similar in shape . Different fonts of alphabets and their Sizes cause a major effort in the process of discrimination . The big challenge in designing a system that can differentiate between such similar characters , to achieve this goal, the differences in minor details must be taken into consideration . The process of identify an object or a feature in image named recognition. This concept is used in many applications Optical character recognition , Pattern and gradient matching , Face recognition , License plate matching , Scene change detection. Character recognition is the electronic identification and digital encoding of printed or handwritten characters[1,2] .

The concept of eigen values can be used in wide applications such that : solving systems of differential equations, analyzing population growth models, and calculating powers of matrices , solving systems of differential equations, analyzing population growth models, and calculating powers of matrices , It is sometimes also called the characteristic value[3], in this work we find the eigenvalues of a square matrix .

Topological feature extraction techniques have been highly successful in a wide area such that : merge trees, which encode an ensemble of threshold-based segmentations and graph matching recognition using structural approaches[4]. Thinning is a process of convert binary image to one wide pixel by reduce pixels to one or it is a morphological operation that is used to remove selected foreground pixels from binary images. The resulted line of the thinned image must be connected and infer shape and topology in the original image. The process of thinning is used in the preprocessing step in many applications such as feature detection for computer vision ,diagram understanding, fingerprint analysis, and optical character recognition[5,6] . There are many researches in this field ,some of related work is as follows :[7] used neural networks for developing a system that can discriminate hand written English alphabets ,the results of average accuracy is 82.5%[8] has developed of a offline handwritten English character recognition algorithm by using artificial neural network , the result of applying this algorithm is 92.59% and false rejection rate (FRR) is 0 % . [9]used artificial neural network to recognize hand written numbers by divided the work into 4 stages and resole the digits number . [10] used effective method in order to recognize character by converting that letter to the frequency domain depending on used discrete cosine and used the seven moments . It recognized English capital letters only. transform (DCT) In this study , an approach for English letters recognition has been proposed using eigen values matrix . This paper is organized as follows. In section 1 an introduction to the
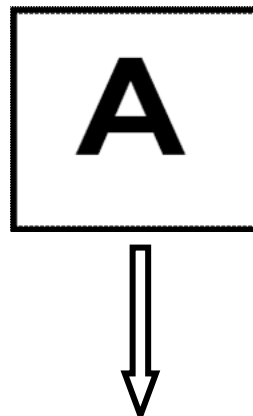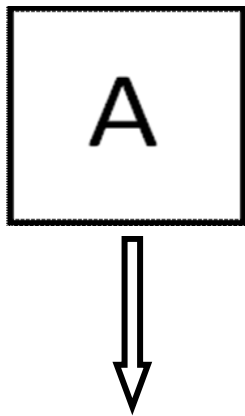
12.4290	9.9852	0
9.9852	4.8990	2.8284
4.8990	3.4641	0
3.4641	2	0
2	0	0
8.8146	7.2915	0
7.2915	4.8990	0
4.8990	2	0
2	0	0
0	0	0
0	0	0
2.8284	2	0
2	0	0
0	0	0

(a) Eigen values matrix

15.3753	12.4290	0
12.4290	9.9852	0
9.9852	4.8990	2.8284
4.8990	3.4641	0
3.4641	2	0
2	0	0
8.8146	7.2915	0
7.2915	5.2263	0
5.2263	2	0
2	0	0
0	0	0
0	0	0
2.8284	2	0
2	0	0

(b) Eigen values matrix

Fig. 1: Similarity between two similar letters

English letter recognition is presented , In section 2 a proposed method is presented , In section 3 a discussion for the experiment results is given , In section 4 a briefly conclusions is presented with a note to future works .

Proposed Method

The proposed method is divided into the following steps , image loading ,image preprocessing , features extraction and recognition . First the input image captured by a scanner ,26-capital letters and 26-small letters are captured ,the features of these letters have been extracted and

saved in a database . Number of image processing techniques are applied to the loading image .The loading image is converted to gray image then to binary image . Fast and effective thinning algorithm is used to obtain thinned image .Topological features are extracted that used to construct our eigen values matrix , finally eigen values matrix used in recognition process .
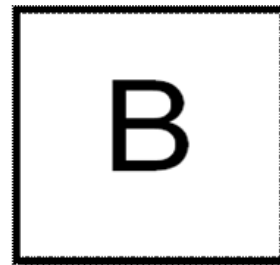
EXPERIMENT

Features for all training set of alphabet images (26-capital letters and 26-small letters)

12.4290	9.9852	0
9.9852	4.8990	2.8284
4.8990	3.4641	0
3.4641	2	0
2	0	0
8.8146	7.2915	0
7.2915	4.8990	0
4.8990	2	0
2	0	0
0	0	0
0	0	0
2.8284	2	0
2	0	0
0	0	0

(a)

23.9507	23.6556	0
23.6556	22.1604	0
22.1604	22.4447	0
22.4447	22.0300	0
22.0300	20.8328	0
20.8328	19.5246	0
19.5246	17.9224	16.4461
17.9224	16.4461	0
15.3959	13.8467	0
13.8467	12.8892	0
12.8892	11.1790	0
11.1790	10.4721	0
10.4721	7.8785	0
7.8785	6.8990	0
5.2263	4.4721	0
4.4721	2	0
2	0	0
0	0	0

(b)

Fig. 2: Similarity between two Different letters A,B

are extracted . Another alphabet images used with different types of fonts and different sizes in order to test the performance of the proposed method . The similarity between two letters are tested by using the value (1) that refers to 100% matching between two letters, maximum value near to (1) refers to large similarity. In experiment 1 two letters have the same font but different size are tested , similarity between them is computed , figure (1) illustrates the similarity between two letters .

The testing for the images above in figure(1) shown that the similarity between them is 9.08 , its near to 1 .The matrices in (a) and (b)

showed the eigen values for each letter no. of rows in matrix refers to no. of parts in thinned image and no. of column refers to no. of neighbors for each part . The operation of portioning thinned image is based on topology features . In next experiment two images for different letters are tested that they have same size and same font , eigen values computed for two images. See below in figure (2) :-

As seen in figure (2) two different letters with same size and same font are tested , two matrices of eigen values for the two letters are computed . The two eigen values matrices have different size , the reason is the two letters have

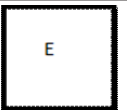**Table 1: Similarity among different font styles for one letter**

A		E	
Font style	**Similarity**	**Font style**	**Similarity**
Arial	0.8996	Arial	0.9196
Time new roman	0.7913	Time new roman	0.7924
Agency	0.9153	Agency	0.8144
Andalus	0.7016	Andalus	0.7011
Calibri	1	Calibri	1
Arial black	0.8050	Arial black	0.9050
Bauhaus	0.4127	Bauhaus	0.5032

Table 2: Similarity among different font styles for one letter

A		E	
Font style	**Similarity**	**Font style**	**Similarity**
12	1	12	1
14	1	14	1
16	1	16	1
18	1	18	1
20	1	20	1
22	1	22	1
24	1	24	1
26	1	26	1
36	1	36	1
48	1	48	1
72	1	72	1

different number of parts and there are different number of neighbors for each part . Letter (A) has (14) parts , letter (B) has (18) , the values in the two matrices were Significantly different . The proportion of similarities between the two letters was too low . The similarity between them was 0.07 , its near to 0 , so the two letters are different .

In last experiment , letters A , E are tested in font size=12 and Calibri as font style ,

seven different font style are used : Calibri , Time new roman , Agency , Andalu , Arial. See table (1) below :-

As seen in table(1) , cells with yellow color indicates to the 100% matching between two letters , cells with green color indicates to low similarity between two letters have the same font size =12 but there have two different style . The first has Calibri as font style and the second has Bauhaus as font
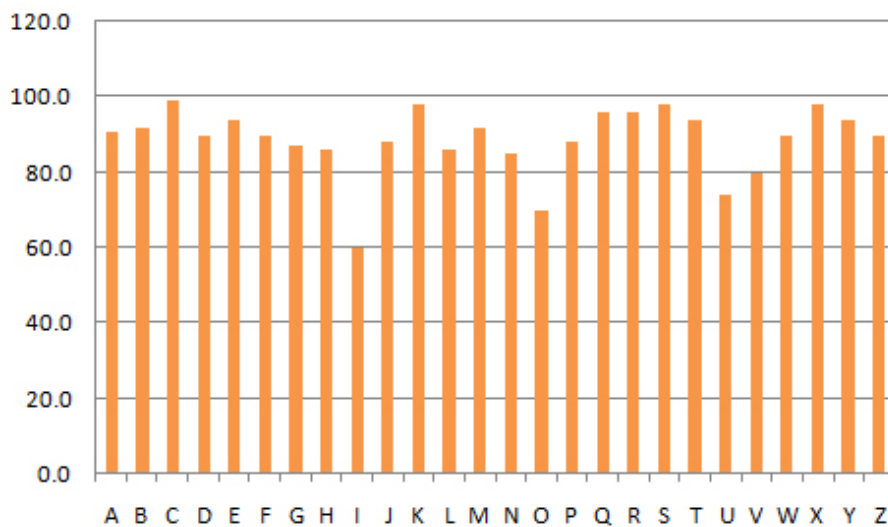

Fig. 3: Recognition rate for each capital letter in English alphabetic
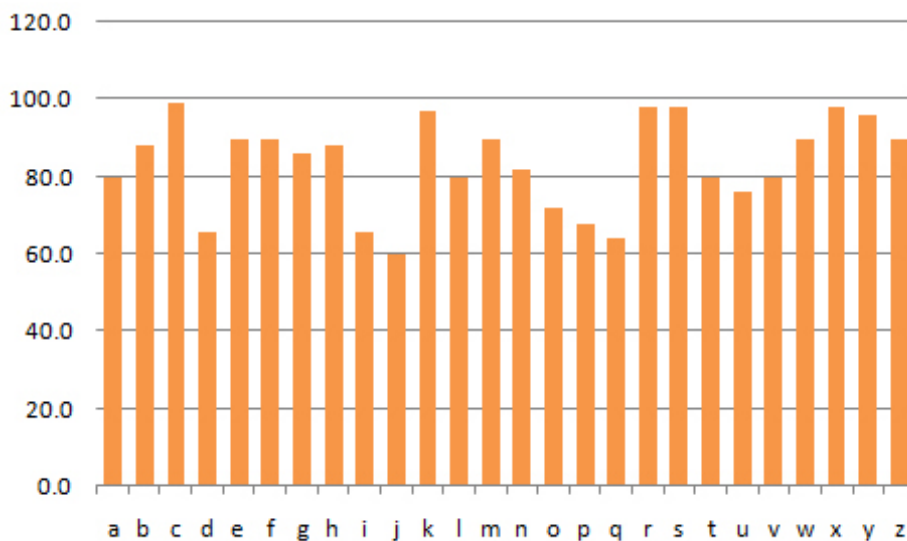

Fig. 4: Recognition rate for each small letter in English alphabetic

style .In table (2) below eleven different font size are used : 12,14 ,16, 18 , 20 , 22 , 24 , 26 , 36 , 48 ,72 . Below table (2) describes the similarity among different font size for one letter .

Cells with yellow color indicates to the 100% matching between two letters have different size , as seen all cells in similarity column have the value =1 in spite of different in size , therefore we conclude that the difference in size does not affect the process of recognition .

In this work (25) different size are used and (25) different font style also are used in testing process , (50) samples for capital letter and (50) samples for small letters the recognition rate of capital and small letters using different sizes = 100% . Recognition rate of capital and small letters using different font style was excellent and the error rate was low .

The results that obtained were close , highest recognition rate, whether large or small letters were in letters : c , k , s , x , It ranged from 96% to 98% . Lower recognition rate appears in small letters in letters : j , p, q , It ranged from 60% to 65% .

CONCLUSIONS

Eigen values matrix is used to recognize capital and small English letters which have different font style and size . Extraction topological features has excellent effect on the recognition process , the results were excellent . Experimental results showed that the difference in size for fonts does not affect on the process of recognition, the result of recognition using (25) different size for fonts = 100 % . Different (100) samples are tested including different font style and different size and the recognition rate was 96. 4%. For future work , this approach can be applied for cursive English letters .

REFERENCES

1. M. Cheriet, N. Kharma, C. Liu and C.Y. Suen, " *Character Recognition Systems",* (2007).
2. R. C. Gonzalez, and R.E. Woods, "Digital Image Processing", Addison-Wesley, Newyork, (2000).
3. Jacqueline Fleckinge , Michel L. Lapidus , " Eigenvalues Of Elliptic Boundary Value Problems With An Indefinite Weight Function " , *Transactions Of The American Mathematical Society,* **295**(1); (1986).
4. Daniel Archambault , Tamara Munzner , David Auber, " TopoLayout: Graph Layout by Topological Features ", (2006).
5. H. Zhou , J. Wu and J. Zhang, " Digital Image Processing: Part II" , ISBN, (2010).
6. D. Vernon *,* "Machine Vision ", Prentice-Hall , (1991).
7. Yusuf Perwej , Ashish Chaturvedi ," Neural Networks for Handwritten English Alphabet Recognition " , *International Journal of Computer Applications* (0975 – 8887) **20**(7); (2011).
8. Tirtharaj Dash , Tanistha Nayak , " English Character Recognition using Artificial Neural Network ", *Proceedings of National Conference* on AIRES-2012, Andhra University .
9. Ahmad Saeed Mohammad , Ahmed Khalaf Hamoudi , Yasmin Abdul Ghani Abdul Kareem , " Hand Writing Numbers detection using Artificial Neural Networks ", *IJCCCE,* **13**(2); (2013).
10. Abdul Monem S. Rahma , Ikhlas F.Nassir , " English Capital Letters Recognition Depends On Computing The Seven Moments ", *Al-Mansour Journal,* **16**: (2011).