A Novel Approach to Recognition of the Isolated Persian Characters using Decision Tree

Mir Mohammad Alipour
Department of Computer Engineering, University of Bonab, Bonab 5551761167, Iran

ABSTRACT
Optical Character Recognition (OCR) is an area of research that has attracted the interest of researchers for the past forty years. Although the subject has been the center topic for many researchers for years, it remains one of the most challenging and exciting areas in pattern recognition. Because of the cursive nature of Persian language, recognition of its characters is more difficult than Latin or Chinese language.

In this paper we propose a novel method to recognize the isolated characters of Persian language using decision tree based on structural features of characters. The system has been tested on a database including all letters of Persian language and a recognition rate of 90.56% has been achieved. Our experimental recognition results are encouraging and confirm our expectation that the use of structural features is an interesting issue of Persian character recognition.

Keywords
Cursive Script, Persian, Isolated Character Recognition, Classification, Decision Tree.

1. INTRODUCTION
Optical Character Recognition is the process of converting the image obtained by scanning a text or a document into machine-editable format. OCR is one of the most important fields of pattern recognition and has been the center of attention for researchers in the last forty decades. The goal is to process data that normally is processed only by humans with computers.

Most of the available systems work on Latin scripts, Chinese and Japanese scripts [1-3]. However the machine recognition of Persian/Arabic text has not been fully explored. The difficulty involved in processing Persian/Arabic text is similar to that of cursive Latin. This is primarily due to the connectivity between characters that complicates the segmentation of each character from the word in which it occurs. Furthermore the connectivity of the variant shape of Persian/Arabic characters in different word position creates another problem in recognition.

Like Arabic, Persian or Farsi is a right to left script, but there are some differences like number of alphabets, font styles, vocabulary and signs, which make Persian OCR somehow different from Arabic. In last decades several researchers worked on Arabic OCR [4-5]. In the field of Persian language, there are papers about isolated character/digit recognition [7-10] and printed text recognition [11-14]. It seems that the first paper about Persian printed text recognition is [15].

In this paper, a novel approach to the recognition of isolated Persian characters using decision tree is introduced . Each character has different features that distinguish it from other characters. These features include: number of segments, left-right density ratio, bottom up density ratio, and other features. These features are used as input to the decision tree to recognize the character in question.

This paper is structured as follows: Section 2 contains a general description of Persian text features. Section 3 presents a brief introduction to Persian OCR system. Section 4 describes the feature extraction techniques used in this research. Section 5 uses the extracted features to classify Persian letters. Section 6 includes the experiments carried out and the results. Finally conclusions come in section 7.

2. PERSIAN TEXT FEATURES
Since the characteristics of Persian/Arabic script is different from the Latin one, and some of the readers might be unfamiliar with this script, a brief description of the important aspects of Persian /Arabic will be presented in this section.

Persian writing is very similar to Arabic in terms of strokes and structure. Therefore, a Persian word recognizer can also be used for recognition of Arabic words. The only difference between Persian and Arabic scripts is in the character sets. Persian character set, comprises all of the 28 Arabic characters plus four additional ones.

Persian language features are classified in several items:

1- Persian text is written from right to left.

2- Farsi has 32 characters out of which 18 have 1 to 3 points which may locate below (like "ب"), above (like "ش") or in the middle of the character (like "ج"). Some of them may have some other vowels (like " ِ " and " ُ ").

3- Persian is a cursive script. Characters are connected and make a component. These components are called "sub-words". A single isolated letter is considered as the extreme case of a sub-word. A word may have several sub-words for example "نرگس" is a word and has two sub-words "نر" and "گس".

4- In contrast to English, Persian characters are not divided into upper and lower case categories. Instead, a Farsi character might have several shapes depending on its relative position in a word. The shape of a character should be changed if it is located at the beginning of the word, in the middle of the word, at the end of the word, and in isolation.

Although there are only 32 letters in Persian alphabet, the total number of different classes to be recognized sums up to 127. Most of the letters have dots above, below, or inside them, number of which is variable between one to three. There are letters whose only difference is the number and/or location of their dots. Ignoring the dots reduces the number of classes to 66.

5- In Persian/Arabic text there is a baseline which usually has more black pixels.

There are other features which are not very important. Some of these features make Persian character recognition very hard and complex. Because of the connectivity between Persian

[1] This work is supported by University of Bonab under Research Projection 100-12.

characters, its recognition is very hard and most of errors occur in the segmentation phase.

3. PERSIAN CHARACTER RECOGNITION

Fig. 1 shows the structure of the proposed Persian OCR system. The system involves 5 stages: Image Acquisition, Preprocessing, Segmentation, Feature Extraction, Classification. A typical OCR system may not include some of these stages.

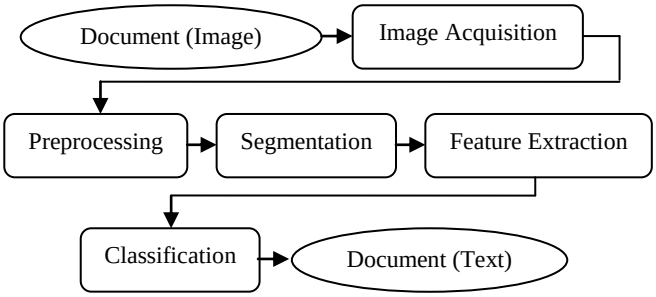

Fig. 1: The structure of the proposed Persian OCR system.

3.1 Image Acquisition

Image acquisition and preprocessing are the two relatively simple stages, which are presented first. Image acquisition is at the image representation level of pattern recognition (PR). It is the process of acquiring a digitized representation of a document or an article to be recognized. A flatbed scanner is used at this stage to acquire 200 dpi, 8-bits gray-level images.

3.2 Preprocessing

All of the processes that improve image quality and prepare it for next stages are called preprocessing. Binarization, filtering, smoothing and thinning are examples of these processes.

3.2.1 Binarization

Binarization is a special case of thresholding, of which there are only two states of outputs in the resulting image, either black or white. It reduces the computational requirements of the system and may enable removal of some noise. A document can be binarized globally or adaptively. Unless the document is printed on an uneven colored paper, global thresholding is good enough to carry out the binarization. Two global thresholding algorithms were studied and implemented. They are Otsu's [16] and Tsai's [17] algorithms. Their results were compared and the Otsu's algorithm was chosen.

3.2.2 Filtering

Noise may appear in the images after scanning or binarization. It influences negatively the system performances. The Gaussian filter is used to remove the noise [18].

3.2.3 Smoothing

A smoothing process was taken. It uses the spatial filter proposed by Amin et al. [19]. It is important to note that this algorithm not only can smooth the image but can also restore missing pixels.

3.2.4 Thinning

Thinning is one of the most important steps in preprocessing stage [20]. It simplifies the texts shapes for segmentation process, feature extraction, and classification. This is resulted in reducing the amount of data that need to be handled. In this paper sequential thinning method based on morphological hit/miss transformation is used [21].

3.3 Segmentation

Segmentation is the most important part of OCR systems especially for Persian and Arabic languages. It directly affects the feature extraction and classification process [22]. When the image is ready to be processed, the first step is to isolate each line of the text from the whole document. A horizontal projection profile technique is used for this purpose. A computer program scans the image horizontally to find the first and last black pixels in a line. Once these pixels are found, the area in between these pixels represents the line that may contain one or more character. Using the same technique, the whole document is scanned and each line is detected and saved in a temporary array for further processing.

Once each line of the text is stored in a separate array, using vertical projection profile, the program scans each array this time vertically to detect and isolate each character within each line. The first and last black pixels that are detected vertically are the borders of the character. It possible that when the characters are segmented, there is a white area above, below, or both above and below the character, except for the tallest character that its height is equal to the height of the line. Since the edges of each character box is needed for the recognition purpose, another horizontal scan is run to detect the top and bottom of the character and isolate the area that only contains the pixels of the character.

At this point, the program has isolated each character in the document and the matrix representation of each character is ready to be processed for recognition purpose. The reminded three stages: Feature Extraction and Classification are described in next sections with more details.

4. Feature Extraction

Character features are characteristics that distinguish one character from another. Human use these features to recognize characters and text. The Persian/Arabic text feature extraction methods can be classified broadly into three main groups, these groups are structural features, statistical features and Global Transformations. The structural features method, in this technique, features are usually extracted based on the text topologies. The structural features of Arabic text may include loops, the intersection points, dots, height, width, number of crossing points, and such [23]. The second is called statistical features methods, these techniques are quick and effective, but may be affected by noise. The statistical features used for Arabic text recognition include: zoning, characteristic loci, crossings and moments [24]. The third is called Global Transformation methods; the Global Transformation aim to shorten the text representation in order to get better results. The global transformations methods used for Persian/Arabic text recognition include: horizontal and vertical projections, coding, Hough transform, Gabor transform [25].

As discussed earlier, at the time of processing, a matrix of pixel values which contains the features of each character image is ready to be used in this stage. Features needed for the recognition process include: number of components, left-right density ratio, bottom-up density ratio and others. Decision trees are then used to classify the characters based on the features that were extracted from the character.

4.1 Secondary parts

More than half of the Persian letters are composed of main body and secondary components. The secondary components

are letter components that are disconnected from the main body. For example, **Beh** (ب) has a dot under its main body, **Teh** (ت) has two dots above its main body, and **Kaf** (ك) has a zigzag enclosed within the main body.

Detecting the secondary components can be done after segmenting the binary image of the letter into its disconnected components using the connected component labeling techniques [26]. Then the main body is easily identified as it is usually the largest component and is closer to the letter's center than the secondary components. The secondary position is then easily found as the position of the secondary components relative to the main body. Finally, the number and position of the secondary components play important role in finding the secondary type. Table 1 shows the classification of Persian characters based on number of parts.

Table 1: Classification of Persian characters based on number of parts (ا and آ are different forms of letter alef and ک and ك are different forms of letter kaf).

Number of parts	Characters
1	ا ح د ر س ص ط ع ک ل م ه و ی
2	آ ب ج خ ذ ز ض ظ غ ف ك گ ن
3	ت ق
4	پ ث چ ژ ش

4.2 Drawing the grid

To create the feature vector, the main body is divided into five equal vertical frames. Then, each frame is divided into five equal cells as proposed in [27]. Drawing a 5*5 grid around the character "ع" is shown in Fig. 2.

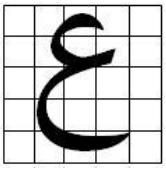

Fig. 2: Drawing a 5*5 grid around the character "ع"

4.3 Bottom-Up (BU) and Left-Right (LR) Density Ratios

Many letters have a noticeable property that distribution of written character on the grid is not equal to the ratio between the pixels of the written letter in the first two rows and the pixels of the written letter the last two rows; or between the first two columns and the last two columns. The first case is called bottom-up ratio while the second one is called left-right ratio. Fig. 3 shows an example of this feature.

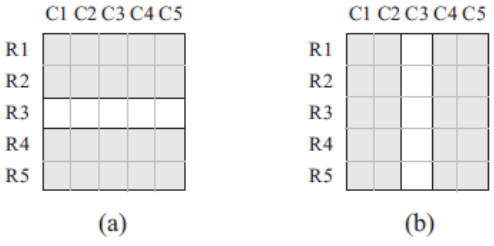

Fig. 3: Density Ratio Calculations (a) BU (b) LR

Every row or column contains five equal size cells; to calculate bottom-up or left-right density ratios we use the following formulae:

$$BU = \#\,pixels\ in\ (R1 + R2) \,/\, \#\,pixels\ in\ (R4 + R5) \quad (1)$$
$$LR = \#\,pixels\ in\ (C1 + C2) \,/\, \#\,pixels\ in\ (C4 + C5)$$

Every pixel corresponds to one element in the lists that we used to store the written character after applying the tracing module.

To get control over these features, we defined two constant values as thresholds T1 and T2 (T1 is greater than T2). The values of these thresholds are determined using a trial-error method. For bottom-up density, if the ratio is greater than T1 then we say that the letter is up-oriented and if the ratio is less than T2, we say that the letter is bottom-oriented. If the ratio between T1 and T2 we say that the character has neutral behavior for this feature. Same definitions on left-right density ratio are applied.

Fig. 4 shows some letters that have up-oriented, bottom-oriented, left-oriented, or right-oriented behavior. Some letters can have combination of bottom-up and left-right density ratios.

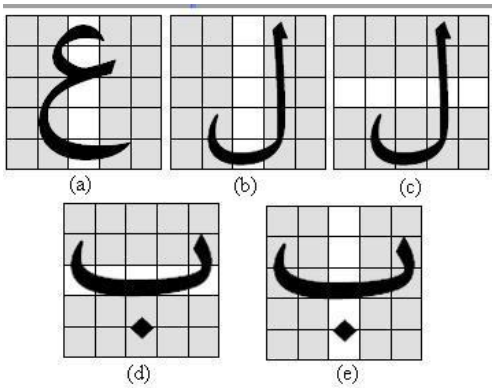

Fig. 4: Density orientation (a) left-oriented, (b) right-oriented, (c) bottom-oriented, (d) up-oriented, (e) neutral left-right orientation.

4.4 Horizontal–Vertical Orientation (HV)

Another helpful feature is the horizontal-vertical orientation. This feature depends on the range of x and y coordinates, HV ratio can be defined as follows:

$$HV = (y_{max} - y_{min})/(x_{max} - x_{min}) \quad (2)$$

Because of different writing styles we define two threshold values S1 and S2 (S1 is greater than S2). The following production rules are used for the decision:

$$If\ \ HV > S1\ then\ the\ letter\ is\ horizontal - oriented \quad (3)$$
$$If\ HV < S2\ then\ the\ letter\ is\ vertical - oriented$$
$$If\ S1 > HV > S2\ then\ the\ letter\ has\ neutral\ orientation$$

In other words, this feature gives us a hint of the grid shape. The grid shape may be a square, vertical rectangle, or a horizontal rectangle. Feature effect can be noticed as it appears Fig. 4. Fig. 4 (a-c) shows horizontal rectangle and Fig. 4 (d-e) shows vertical rectangle.

4.5 Loop detection

There are nine Persian characters (ص , ض , ط , ظ , ف , ق , م, ه and و) that have closed loops. Many algorithms are used for detecting shapes, curves and motions in the field of image processing and computer vision such as hough transform, however hough transform has several shortcomings, including high computational cost, low detection accuracy and possibility of missing objects. In this paper we used the algorithm for Loop detection in a character which have developed by Surhone *et al [28]*.

4.6 Sharp Edges (SE)

Sharp edge detection is the most difficult feature to be extracted. Sharp edge is similar to 20-40 degree angle. To illustrate this feature see Fig. 5 that shows some letters that have sharp edges. There are two types of sharp edges with regard to the direction of the edge. In Fig. 5 (a) the letter ”AYN” is a y-direction sharp edge type while in Fig. 5 (b) the letter ”SAAD” is an x-direction type. Y-direction sharp edge is detected during the movement of the pen from upward to downward, and the x-direction is detected when a sharp turning point exists with the movement from right to Left.

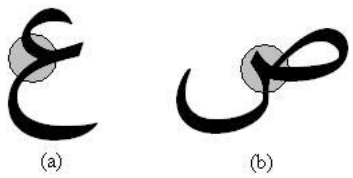

Fig. 5: Letters containing sharp edge: (a) y-direction, (b) x-direction sharp edges.

4.7 Characteristics of Persian characters

This section presents characteristics of Persian characters and some observations. These characteristics are found by analyzing the features extracted from the characters. According to the number of components, we have classified the Persian characters into four classes:

1. One - part characters.

2. Two - part characters.

3. Three - part characters.

4. Four - part characters.

To distinguish between these classes, one should verify the existence or absence of a sharp edge (SE).

Tables 2-5 show the characters according to this classification. The following abbreviations are used in these tables: *B: Bottom, U: Up, L: Left, R: Right, V: Vertical, H: Horizontal, DC: Don't care, X: x-direction, Y: y-direction.*

Table 2: Characteristics of One-part characters

Character	Loop	H/V	BD	LR	SE
ا	No	H	U/DC	DC	No
ح	No	V	U	DC	Yes-Y
د	No	V	B	R	No
ر	No	V	U	L	No
س	No	V	U	DC	Yes-X
ص	Yes	V	U	L	Yes-X
ط	Yes	V	B	L	No
ع	No	H	U	DC	Yes-Y
ک	No	V	U	R	No
ل	No	H	B	R	No
م	Yes	H	U	DC	No
ه	Yes	V	U/DC	DC	No
و	Yes	V	U	R	No
ی	No	V	B	DC	No

Table 3: Characteristics of Two-part characters

Character	Loop	H/V	BD	LR	SE	SP Location
آ	No	H	U	DC	No	-
ب	No	V	U	DC	No	Bottom
ج	No	V	U	L	Yes-Y	Middle
خ	No	H	U	L	Yes-Y	Up
ذ	No	H	U	R	No	Up
ز	No	H	B	R	No	Up
ض	Yes	V	DC	R	Yes-X	Up
ظ	Yes	V	B	R	No	Up
غ	No	V	B	DC	Yes-Y	Up
ف	Yes	V	DC	R	No	Up
ك	No	V	B	R	No	-
گ	No	V	U	R	No	-
ن	No	V	B	DC	No	Up

Table 4: Characteristics of Three-part characters

Character	Loop
ت	No
ق	Yes

Table 5: Characteristics of Four-part characters

Character	BD	LR	SE	SP Location
ب	B	DC	No	Bottom
ث	B	DC	No	UP
چ	DC	DC	Yes-Y	Middle
ژ	U	R	No	Up
ش	U	R	Yes-X	Up

5. CLASSIFICATION

The classification stage is the decision making part of a recognition system. Persian OCR systems can recognize the text by either the Holistic (Global) or Analytic strategies. The Holistic (Global) strategy recognizes the whole words or sub-words, as well as it does not require segmentation, and it works on limited number of vocabularies [29]. On the other hand the Analytic Strategy recognizes the segmented features, as well it requires segmentation, and can be applied on unlimited vocabularies.

The features extracted in the previous phase are used in the classification stage. The decision tree is used to recognize the Persian letters. To simplify the decision tree, for each class of letters, one-part, two-part, three-part and four-part, different tree is drawn. So, each character is sent to appropriate tree according to its number of parts. In each decision tree, rectangles show group of letters with some common attributes. Diamonds are used to separate letters in the above rectangle based on the value of attribute specified as a condition (for example have a loop or haven't?). In other words, the values of the attributes determine the branch that should be selected. Finally circle is used to show identified letter. The decision tree used for one-part letters is shown in Fig. 6. The abbreviations are used in these trees are identical with tables 2-5 along with *X/Y:* types of sharp edges.

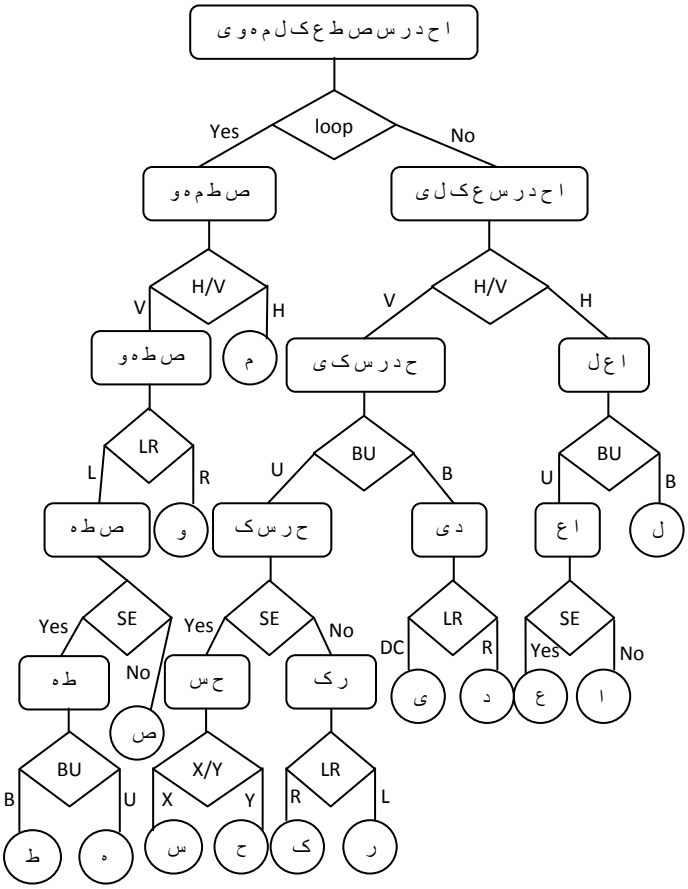

Fig. 6: The decision tree for one-part letters.

As you can see from the figure, every letters of one-part class is identified successfully according to attributes determined in table 2. The decision tree used for two-part letters is shown in Fig. 7

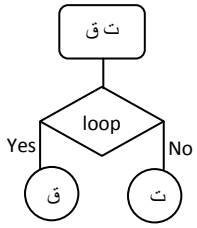

Fig. 7: The decision tree for two-part letters.

Also every letters of two-part class is identified successfully according to attributes determined in table 3. The decision trees used for three-part and four-part letters are shown in Fig. 8,9.

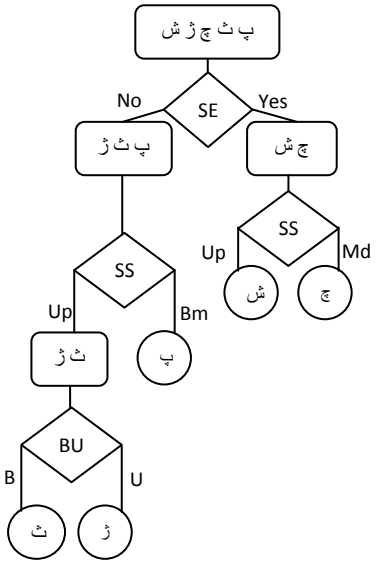

Fig. 8: The decision tree for three-part letters.

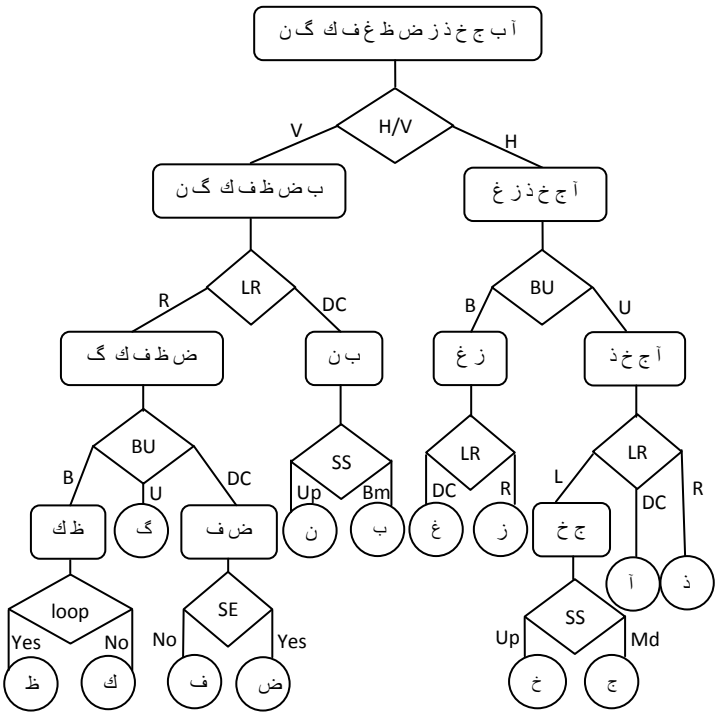

Fig. 9: The decision tree for four-part letters.

6. EXPERIMENTAL RESULTS

To evaluate the proposed system, we constructed a database including all letters of Persian language. For each letter, 10 images (various fonts and sizes) were stored in the database. The system shows the performance of itself. Here the average recognition rate of 90.56% was achieved. Table 6 presents the recognition accuracy for each Persian letter.

Table 6: Recognition rate for each Persian letter.

Letter	Rec. Rate (%)	Letter	Rec. Rate (%)
آ	98	ص	90
ا	98	ض	88
ب	96	ط	95
پ	92	ظ	93
ت	94	ع	78
ث	94	غ	74
ج	74	ف	94
چ	74	ق	90
ح	76	ک	95
خ	72	ك	91
د	92	گ	94
ذ	92	ل	98
ر	97	م	93
ز	96	ن	90
ژ	94	و	96
س	96	ه	96
ش	95	ی	94
Average Rate (%)		90.56	

These results show that the system performance on typical documents of familiar fonts with resolution of 300 dpi is quite well. However, for the case of low quality documents like Fax pages, old documents, camera images and low-resolution documents, the problem is still open. In this work, we did not consider these kinds of documents. In these cases there are several broken characters, nonlinear distortions and noise so that we need some other algorithm to be designed to improve the system accuracy. For example a preprocessor module may be required to enhance the quality of the image.

Further more, testing results show that the system had some trouble identifying letters Jim "ج", Che "چ", Ha "ح" and khe "خ". This is may be caused by the fact that these letters have a little bit similarity with letters Ain "ع" and Ghein "غ".

7. CONCLUSION

The recognition of Persian Letters is hard because of the similarity between letters. Classification stage introduces one of the most serious problems in the development of cursive script OCR system including Persian language scripts. In order to overcome this problem, we use a set of features to be extracted. That is, letters are divided into body and secondary parts then a set of features were detected such as the number and type of secondary parts, the position of the secondary parts whether above or under the body of the character, the existence of loops and other structural features of the Persian characters.

The proposed system can recognize several popular Farsi fonts with different sizes. The proposed system was tested on a database including all letters of Persian language which for each letter, 10 images (various fonts and sizes) were stored in the database. The average recognition rates of 90.56% were achieved. From testing results, it was concluded that the overall system is proved to be stable and successful.

8. REFERENCES

[1] J. Mantas, "An Overview of Character Recognition Methodologies", Pattern Recognition 19, 1986, pp.425-430.

[2] R. M. Bozinovic and S. N. Shihari, "Off Line Cursive Script Word Recognition", IEEE Trans. Pattern Anal. Mach. Intell. PAMI 11, 1989, pp. 68-83.

[3] R. Casey and G. Nagy, "Automatic Reading Machine", IEE Trans. Comput. 17, 1968, pp. 492-503.

[4] Amin, A.: Off-line Arabic character recognition: the state of the art. Pattern Recognition. 1998, 31(5), 517–530

[5] Gouda, A.M., Rashwan, M.A.: Segmentation of connected Arabic characters using hidden Markov models. IEEE International Conference on Computational Intelligence for Measurement Systems and Applications, USA 2004, pp. 115–119

[6] Kurdy, B., AlSabbagh, M.: Omnifont Arabic optical character recognition system. In: Proceedings of International Conference on Information and Communication Technologies: From Theory to Applications, pp. 2004, 469–470

[7] Khosravi, H., Kabir, E.: Introducing a very large dataset of handwritten Farsi digits and a study on their varieties. Pattern Recognit. Lett. 2007, 28(10), 1133–1141

[8] Mansoory, S., Hassibi, H., Rajabi, F.: A heuristic Persian handwritten digit recognition with neural network. In: The 6th Iranian Conference on Electrical Engineering, 1998, pp. 131–135

[9] Soltanzadeh, H., Rahmati, M.: Recognition of Persian handwritten digits using image profiles of multiple orientations. Pattern Recognit. Lett. 2004, 25(14), 1569–1576

[10] Mozaffari, S. and H. Soltanizadeh, 2009. ICDAR 2009. handwritten Farsi/Arabic character recognition competition. Proceedings of the 10th International Conference on Document Analysis and Recognition, July 26-29, IEEE Xplore, Barcelona, 2009, pp: 1413-1417. DOI: 10.1109/ICDAR.283

[11] M. Alipour, "A New Approach to Segmentation of Persian Cursive Script based on Adjustment the Fragments," International Journal of Computer Applications 2013, Vol. 64, No 11, pp. 21–26.

[12] Azmi, R., Kabir, E.: A new segmentation technique for omnifont Farsi text. Pattern Recognit. Lett. 2001, 22, 97–104

[13] Ebrahimi, A., Kabir, E.: A pictorial dictionary for printed Farsi subwords. Pattern Recognit. Lett. 2008, 29(5), 656–663

[14] Mehran, R., Pirsiavash, H., Razzaziy, F.: A front-end OCR for omni-font Persian/Arabic cursive printed documents. Digital Imaging Computing: Techniques and Applications, 2005, pp. 385–392

[15] Parhami, B., Taraghi, M.: Automatic recognition of printed Farsi texts. Pattern Recognit. Lett. 1981, 14, 395–403

[16] N. Otsu, A threshold selection method from Gray-level histogram, IEEE Trans. Systems Man Cybernet. 9 (1) 1979, 62-66.

[17] W.H. Tsai, Moment-preserving thresholding: a new approach, Comput. Vision Graphics Image Process. 29, 1985, 377-393.

[18] C. Gonzales. Rafael and E. Richard,Woods., Digital Image Processing. 2nd ed. Englewood Cliffs, NJ: Prentice-Hall, 2002.

[19] A. Amin, W.H. Wilson, Hand-printed character recognition system using artificial neural network, Proceeding of second International Conference on Document Analysis and Recognition, 1993, pp. 943-946.

[20] B. K. Jang and R. T. Chin, "Analysis of thinning algorithms using mathematical morphology,"IEEE Trans. Patt. Anal. Machine Intell., 1990, vol. PAMI-12, no. 6, pp. 541-551.

[21] B. Timsari, Character recognition in typed Persian words: a morphological approach, M.S. thesis, Isfahan Univ. of Tech. 1992, Iran.

[22] R. Safabakhsh, and P. Adibi. Nastaaligh Handwritten Word Recognition Using a Continuous-Density Variable-Duration HMM. The Arabian Journal for Science and Engineering. 2005, 30: 95-118. April.

[23] H. Goraine, M. Usher, and S. Al-Emami. Off-Line Arabic Character Recognition,‖ Computer, 1992, vol. 25, pp. 71-74.

[24] B. AL -Badr and S. Mahmoud. Survey and bibliography of Arabic optical text recognition. Signal Processing, 1995, 41(1): 49-77.

[25] F. Zaki, S. Elkonyaly, A. Elfattah, and Y. Enab. A new technique for arabic handwriting recognition. Proceedings of the 11th International Conference for Statistics and Computer Science, Cairo, Egypt, 1986, pp; 171–180.

[26] A. Rosenfeld and A. Kak, Digital Picture Processing, Academic Press, New York, 1976.

[27] R. El-Hajj , L. Likforman-Sulem, C. Mokbel, "Arabic Handwriting Recognition Using Baseline Dependant Features and Hidden Markov Modeling", Proceedings of the 8th International Conference on Document Analysis and Recognition (ICDAR'05), Seoul, Korea, 2005

[28] Surhone, L.M., M.T. Tennoe and S.F. Henssonow. Randomized Hough Transform. 1st Edn., VDM Verlag Dr. Mueller AG and Co. Kg, Germany, ISBN-10: 6134695823, 2010, pp: 92.

[29] A. Dehghani, F .Shabani and P. Nava. Off-Line Recognition of Isolated Persian Handwritten Characters Using Multiple HiddenMarkov Models, Proc. Int'l Conf. Information Technology: Coding and Computing, 2001, pp. 506-510.