

Offline Software system to evaluate Handwriting with pen

Mohammad Sajjad AGHADADI

Master's student of Iran University of Science and Technology – Tehran - Iran

Abstract

Line from the past to the present is of utmost importance, to the extent that most psychologists are aware of the fact that, from the handwriting of individuals can be partly realized their personality and psychological characteristics. Therefore, all people are willing to have a good handwriting. In the field of image processing, valuable work has been done on the field of detection handwriting. However, less attention has been paid to the subject of evaluation handwriting. To the extent that, according to searches that conducted by me, such work has not been done in the Persian line. The purpose of the evaluation in this article is to compare the user's handwriting with reference handwriting in terms of rules and principles and then displays the user's level of matching with the reference line and displays the differences between the user and the reference line. For this purpose, after selecting the reference line, skeleton of the user's handwriting and the reference one is extracted and after framing the picture, comparing this two handwriting by using an innovative method. Then, obtain the user-level and display it. In the end, in order to learn how to properly write, user handwriting and reference line simultaneously displayed in order to distinct and improve it. It should be noted that the project is currently implemented on the Persian letters.

Keywords: Image Processing, Optical Character Recognition, Handwriting, Eskeletonization, Framing

1. Introduction

One of the issues that have many applications is detection of handwritten line, that is use for instance in banks in order to identify the owner of the account by matching signature and other uses are also the subject of research and practical work that has become a subject of fascination. The subject of writing beautifully has attracted the attention of most people. By looking at this fact that I want to do something real and practical that be usable by all people, so I chose pen in my work that is most common tools for people to writing. By two ways I can do my work on handwritten lines: 1- Off-line method: In this method, at first system receive the user's image handwriting; then analysis of the user's handwritten image has been started. 2- The online method: in this method as soon as the user writes, immediately analysis on his handwriting is started and the result displayed real time for him. Our work is offline. It should be noted that by the studies conducted on the articles and related projects, we did not find the same work as this project, so I claim that this project is done for the first time at least in the Persian language. In next steps I explain my proposed approach.

1.1. System architecture

Figure 1 contains overview of the system that is used to evaluate handwriting with pen. This chart is based on optical character recognition systems. In the pre-processing stage, acts such as skeleton, framing, normalization and etc will be done; pre-processing meant to prepare the user's image to perform the comparison and evaluation. In matching level, the user's handwriting image is adjusted by reference handwriting so that the differences between the reference handwriting and the user's handwriting become clear. In valuation step, result and status of user's handwriting is displayed. Finally, to improve the user's handwriting, guidelines will be done.

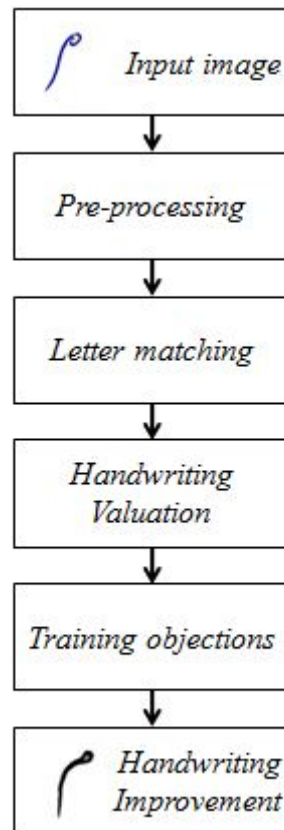


Figure 1: The overall evaluation of handwriting

2. Introduction of evaluation handwriting system by pen

2.1. Persian letters features in the form of Nastaliq

This system needs a reference line to evaluate the user's handwriting. We take Nastaliq as the reference line. For choosing the reference line, we should select a line that has these features, one we can work on them operationally, and have beauty that separates it from the rest of writing styles. The more important things are the reference handwriting is acceptable by people that causes create a desire for users to use the system to improve their handwriting. After consultation with line masters and asking questions of the public about their ideal line, I decided to choose Nastaliq as a reference line.

Persian letters in Nastaliq format have unique characteristics in terms of size and manner of writing. In order to evaluate and recognize letters we have learned from them. One of the features that Persian letters have and are more in Nastaliq is that letters can be divided into

separate classes, this work done with respect to the structure and shape of letters. Class letters and also features in the form of Nastaliq according to my experience in this field are collected in Table 1.

Table 1: Classification and feature of Persian letters in Nastaliq format

Letters	Features
ا	The length is fourth point which is equal to 4 mm
ب پ ت ث	This four letters are in a same family by length of 5 mm and length of each dent is equal to 1 mm
ج چ ح خ	This four letters are in a same family and have a circular arc by radius 1 mm
د ذ	The length of this two letters are 2 mm
ر ز ژ	The length of this two letters are 2 mm
س ش	The circular shape of the two words form a circle with a radius of 1 mm
ص ض	The circular shape of the two words form a circle with a radius of 1 mm
ط ظ	The length of this two letters are 4 mm
ع غ	The arc such as the circular by radius of 1 mm
ف	The length of this is 4 mm
ق	It's a round circle with a radius up to 1 mm
ک گ	This two letters have 4 mm height and 4 mm length
ل	Circular arc with radius of 1 mm by 4 mm line
م	The length of this is 5 mm
ن	Circular arc with radius of 1 mm
و	The length of this is 2 mm
ه	The length of this is 1 mm
ی	The first part have 2 mm longitudinal and circular arc by radius of 1 mm

2.2. System introduction

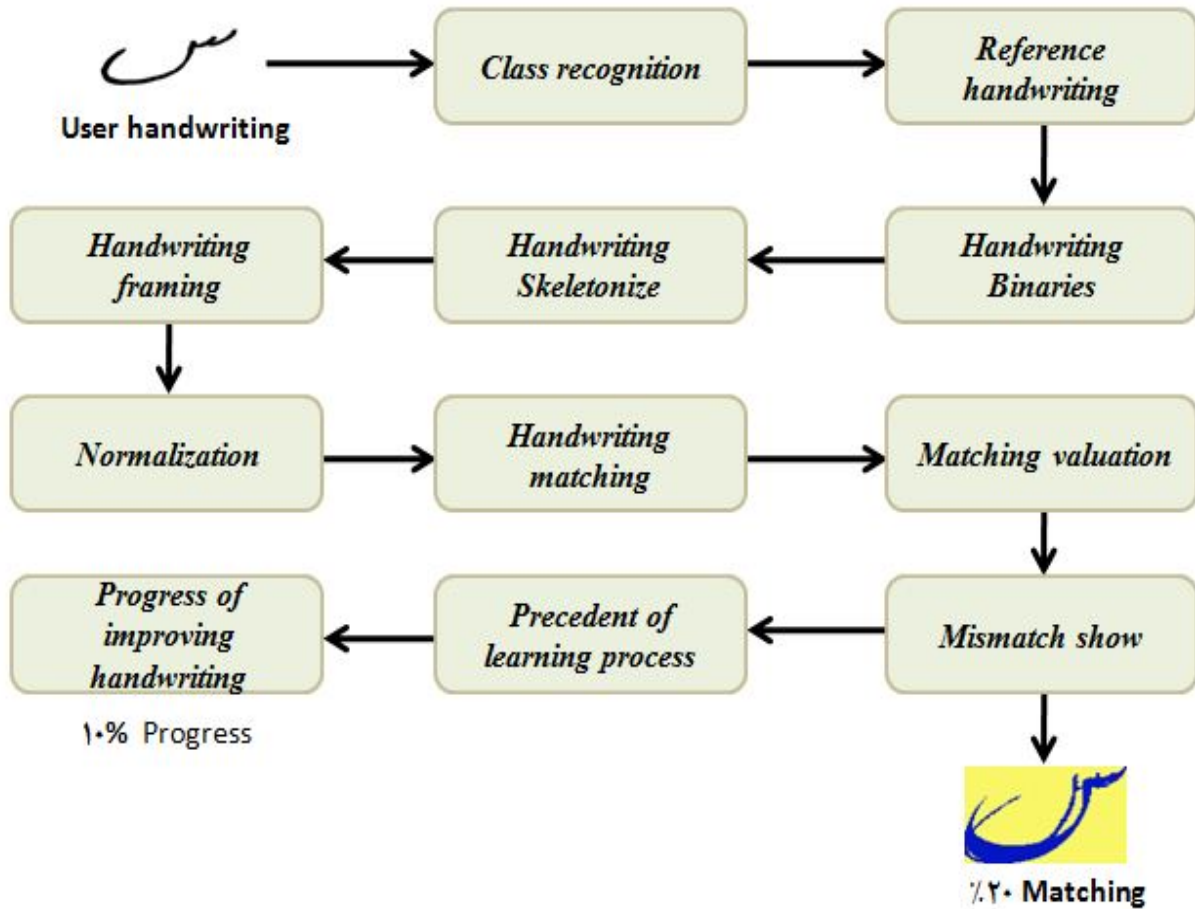


Figure 2: Algorithm for evaluate handwriting with pen

At the beginning, user enters handwriting image to the system, the user's handwriting must be written by pen and a photograph that was taken from it in order to apply the required processing should be have good quality. This system can operate the evaluation with any quality but if the user needs to see good quality output, that is required to enter an image with good quality. In order to start the evaluation process, it is necessary to understand what is the letter that the user entered, the following method is used: In the first approach, commensurate with the characteristics and shape of the letters that was mentioned in the previous section, we have divided the letters into separate classes and according to do the evaluation we only need to know the letter's class. At the beginning, a question will be asked that you plan to use to evaluate which of the classes of letters, and user should be chosen the classes' number that is required to start evaluation process. By entering the number of letter's class, system enters the related reference handwriting and by doing some works the evaluation handwriting will be done.

2.2.1. Convert to binary image

In order to perform image processing on the images of handwriting it is necessary to convert color images to binary images. Therefore, the first step is to convert the images into binary image. To obtain a binary image we use `im2bw` recipe. It is necessary to enter the image name as a parameter to the function.

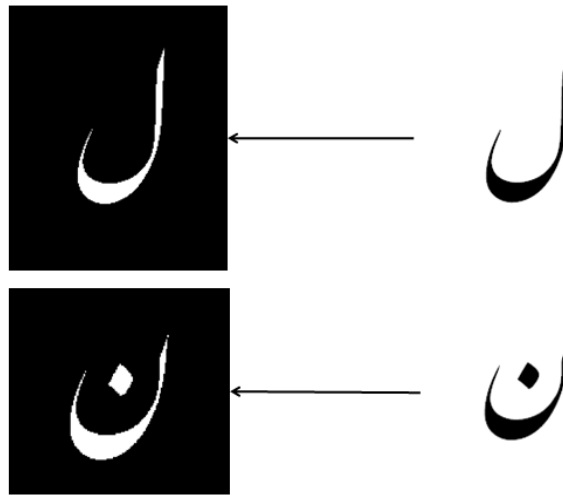


Figure 3: convert to binary image

2.2.2. Skeleton image of handwriting

The first step after the binary image of handwriting image is accessible; get a skeleton from the user's letter. The purpose of acquiring the skeleton of handwriting is that we are looking for a way that not only have this possible to comparable but also have the general style of the manuscripts, so we decided to obtain skeleton of the letters. To obtain the skeleton of handwriting I check a lot of algorithms and compare the strengths and weaknesses of them. I decided to use the skeletons algorithmic that was written by Mr. Hown [1]. As is clear from the figure the skeleton image has the overall structure of the real image. This method is prepared for use in hand-written images that are more efficient than the tools exist in MATLAB.

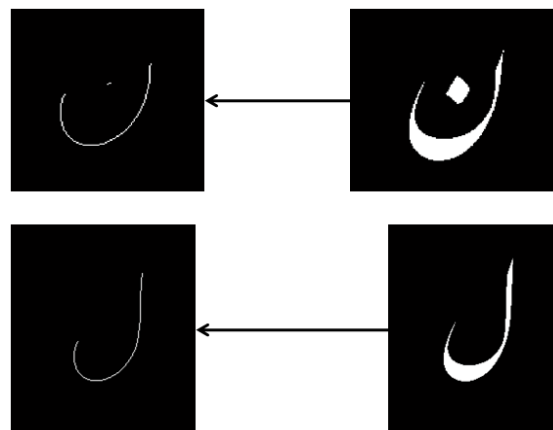


Figure 4: Image skeletonize

2.2.3. Framing

Framing is the process of extracting the specific part of real picture in order to view more detail. In this step because the skeleton image size is equal to the original image and because we want to evaluate the handwriting by skeleton of reference image, it is necessary to remove extra unused area and extract just skeleton image. Also in order to have high accuracy

in evaluation process we need to extract the letter from whole image. For this purpose, we use an innovative method that just extract skeleton from within the overall picture. Figure 5 is an example of extracted the skeleton image from the reference image.

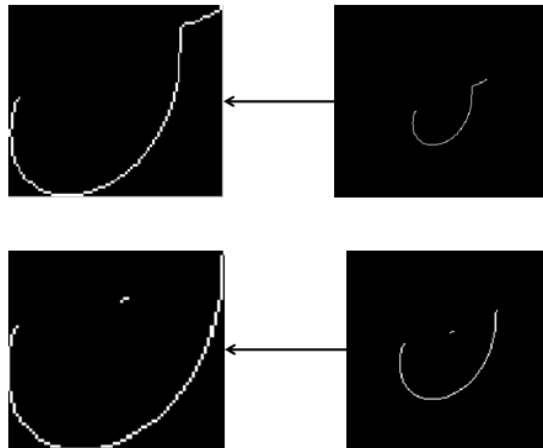


Figure 5: Framing

The method used to extract letters from the picture is: In order to find the letters in the image and extract it, it was necessary to find the pixels that are forming the letters. For this purpose, I used the MATLAB tool and `bwtraceboundary` command. This function gives us the output that consists of a matrix of values that have position of pixels that has one value. Therefore, at this stage we have the amount of pixels in the image that create the letters. We have to get out the letter from the image, for this I use cropping tools in MATLAB.

2.2.4. Normalization

Normalization in this project is to make size of user image same as the reference image, which this stage is the last step before the handwriting evaluation; Frame size of user should be same as reference image. The reason for this is, to get high accuracy and quality of evaluation we need to have two images by same size. It should be noted that the user's handwriting image and the frame of reference image has slight differences in the length and width with a few pixels. We use the following way to normalize image:

System use MATLAB tools to make size of user's handwriting image same as the size of reference image. To do this we use the command `imresize`. In this command for the first parameter, enter the name of the image that we want to change its size, and we enter the second parameter, consider two values as a vertical and horizontal component. Here the parameters that we must measure are the vertical and horizontal components of the reference image.

2.2.5. User's handwriting matching with reference handwriting

The way of compare and matching of the user and the reference image is shown below:

After the skeleton image of user and reference are reachable and framing and normalization have be done we start comparison step. By given that the original reason that makes Nastaliq become the most beautiful style is how the letters are formed and how to write it, in other words skeletons and body of letters are very important, this is reason that we want to compare the skeleton of handwriting. By examining the skeletons and comparing with the original letters, skeleton form not eliminates the characters of letters.

Another point that should mention it is, if the positions of pixel that exist in skeleton of user's handwriting are match to the pixel of reference skeleton, it means that user write accordance of principle and rules of Nastaliq format. The more you appreciate this; the adaptation is more and means you follow the rules of Nastaliq style. The method of matching handwriting with a reference image according to the items listed above is that, we check the percentage of pixel location that has same position from user handwriting to reference handwriting.

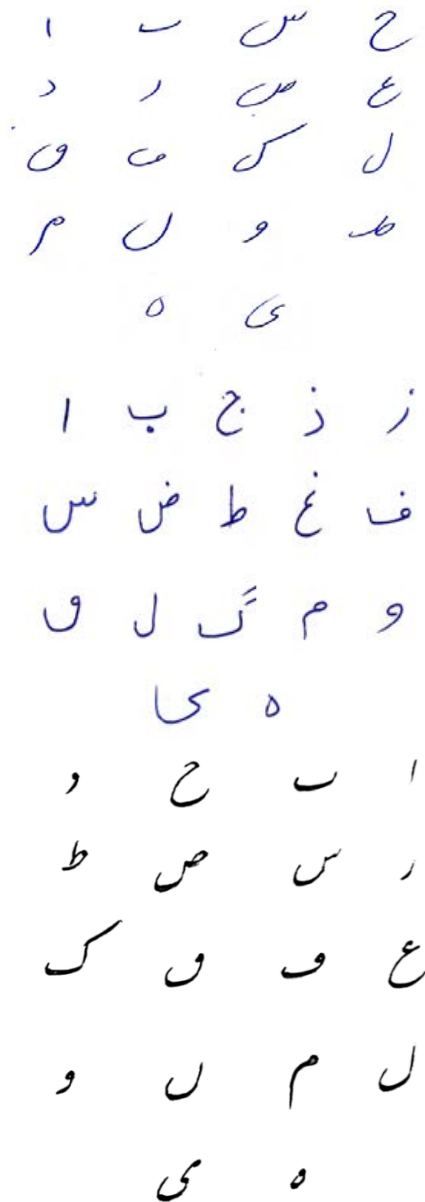


Figure 6: Sample of user's handwriting

2.3. Handwriting valuation

According to our goal that it was evaluate the user's handwriting, so we must set criteria to evaluate handwriting. At this stage, by reference to the work that we did above for handwriting we define level of user in writing. For this purpose we have three level, good (A), intermediate (B) and basic (C). The method of handwriting ranking is as follow:

At first, system count the total number of pixels by one value in the reference frame, then count the total number of pixels by one value in user frame that have a same position of the reference frame. Then by divide these two values together, if the number is less than 20, level of user is C, if the resulting number is between 20 and 60 stand in level B, if the number is between 60 and 100 stand in level A. Below the range for valuation handwriting is shown.

$$\begin{aligned} \text{A: } 60 < & \frac{\text{Number of match pixel}}{\text{Number of whole pixel of reference picture}} * 100 < 100 \quad (1) \\ \text{B: } 20 < & \frac{\text{Number of match pixel}}{\text{Number of whole pixel of reference picture}} * 100 < 60 \quad (2) \\ \text{C: } 0 < & \frac{\text{Number of match pixel}}{\text{Number of whole pixel of reference picture}} * 100 < 20 \quad (3) \end{aligned}$$

2.4. Methods of measurement

2.4.1. Show the mismatch

After finishing the comparing and valuation of handwriting and reference image, in order to confirm the user to know their bugs and also aware that in which part of letters have fault, system shows the Handwriting image and the reference image with each other by different colors. By seeing these two pictures, user can learn which part has wrong and try to improve their handwriting in next times. This will essentially teach the user to figure out what to write. In Figure 7, we show the mismatch.

As mentioned in the beginning, the purpose of this section is to show the user's fault to write Persian letters base on rule of the Nastaliq format and learning aspect for users, is an important part of this system. In this section we have tried to have color images because the learning aspects is very important to be used for the user.

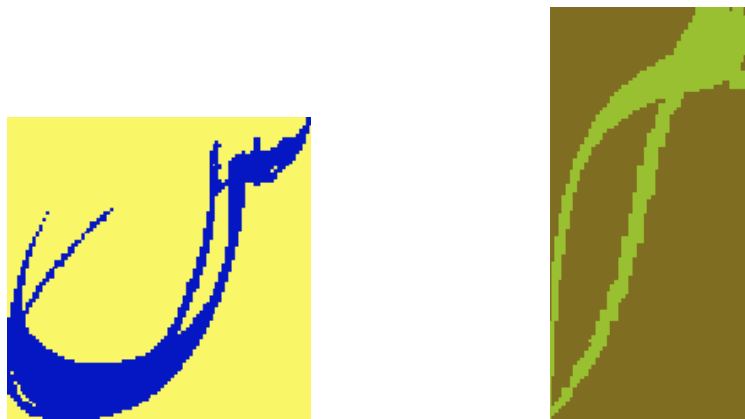


Figure 7: Match and mismatch

2.4.2. User experience of learning

Users may want to evaluate their handwriting by write letters for several times, In this case the result store in system and if the user want to aware of the process of learning, system show the score that user earn. This score consist of number up to 100 that is display to the user.

By using these numbers user can know their change and growth and decline the process of writing. Table 2 shows an example of the history of a user who proceeded to write the letter ω for 10 times.

Table 2: Score history of user

History of user to write ω
48.6667
35.6667
25
60.6667
52.3333
49.6667
39.256
34.8776
20.788
50.4464

2.4.3. Show the improvement of handwriting

In this section, according to score that user obtained from writing a letter, at this step the user's handwriting improvement was shown in graph in order to see the process of progress. In this diagram the average user's score is displayed so the user can evaluate their work. Figure 8 shows a diagram of process flow.

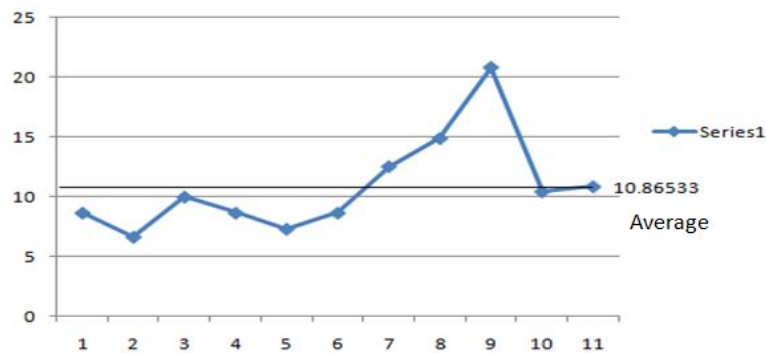


Figure 8: user's progress 10 samples

Due to interest and ability of individuals to write in a Nastaliq format some changes would be there in diagram. For example, people who are interested to improve their handwriting, in each iteration, trying to write better than before, they climbed nearly progress, and if they have decline the amount of it is little. Some people who are not very interested in improve their handwriting may have positive and negative progress and have long ascent and descent in their chart, like the previous figure. The Enthusiasm of people to improve their handwriting can be predicted on the graph of their progress.

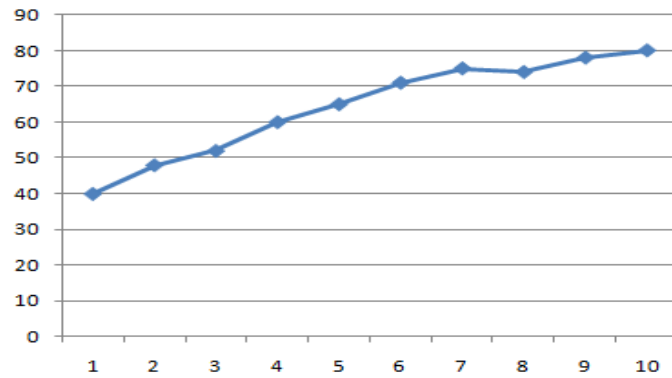


Figure 9: Chart of person handwriting who has learned well Nastaligh

Figure 9 is a graph of person who has learned Nastaligh format well and this progress is obvious from their graph.

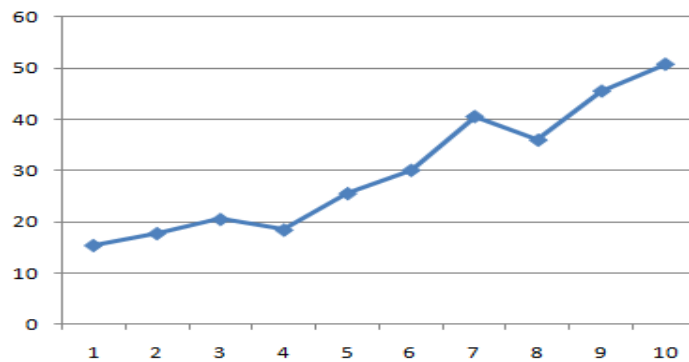


Figure 10: Chart of person who is interested in learning Nastaligh

Figure 10 also show the user that makes efforts to improve their handwriting and are among the interested people in learning Nastaligh.

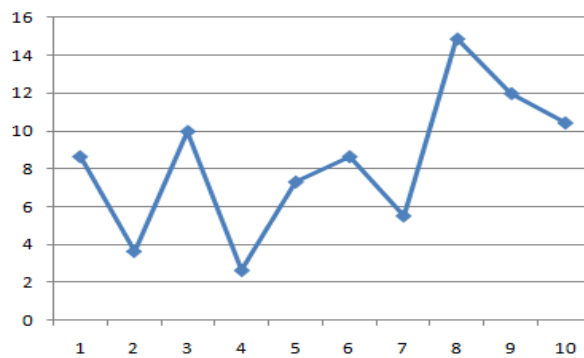


Figure 11: Chart of who has no interest in learning writing

According to Figure 11 realize that a user that has such progress chart who does not want to improve their status for write beautifully.

3. Conclusion

In this system the color of pen is effect on how the system works. So system has some threshold parameters that is require for each color, these parameters can be changed, but if we

want to use fixed threshold it is necessary to use black pen. The default system is intended for black pen and in some cases that was written with different color it is necessary to change parameters.

The system is sensitive on pressure of pen. Given that more and more pressure when writing, text color spontaneously becomes bold and by considering the above-mentioned point, in most cases we should check the parameters and if necessary change it according to the text color. The pressure of pen in the system will be an important issue.

Given that the purpose of this system is evaluate handwriting. The user's writing does not depend on dots so in this system ignored the evaluation of dots and we focus more attention on the structure and context of letters, and if necessary can be taught the user how to punctuation separately.

As mentioned in the previous section we divide users into different groups. One analysis that can we do is that, base on the chart progress that users have; we guess what is the group of this user that use the system. Therefore, another feature that this system has is the ability to recognize the characteristic of people to learn about Nastaliq and improving the handwriting.

3.1. Innovation

As mentioned at the outset that there is more research on the character recognition and handwriting recognition is one of the issues. During a search among various articles in this discussion we came to the conclusion that handwriting recognition has been widely used in the English language and while the number of similar articles in Persian handwriting recognition is lower than English. For example, issues such as numbers recognizing in Persian language exist. In evaluating handwritten after a lot of searching among the papers in English and Farsi to the best of our knowledge we can say, at least in the Persian handwriting did not find anything similar to my work. So we had do from the beginning to the end by myself, which involves defining the problem, define evaluation criteria and all of the work that required for this project and this led to some problems encountered in the process of working.

3.2. Experimental settings

As mentioned above, the system requires the user's handwriting and reference handwriting. As mentioned reference handwriting is the Persian letters in the form of Nastaliq. To get the letters in the form of Nastaliq and enter those to system as a reference, have few ways: First, using pen and by helping masters in nastaliq give letters and store in the system. In order to enhance the accuracy of the skeleton of the reference handwriting, after a review we concluded that use Nastaliq fonts in Microsoft Word. This font has the principles and rules of Nastaliq format. Persian-language letters in form of Nastaliq shown in the image below.

ابپت شجچ حخ دذرز ژس شس ص

ض ط ط ع غ ف ق ك ك ل م ن و ه ی

Figure 12: Persian letters in Nastaliq format

3.4. Future plan

Due to the large enterprises and the possibility of continuing this work at higher levels my future plans is to complete my project as well as to add some items such as full training by the system, able to write at different levels of pressure and it can be analyzable by the system, take advantage of online and view the results of evaluation in online mode and also analysis handwriting with different colors are my future work that will be done to improve the system. Try to strengthen the aspects of learning in the system until the user can improve their handwriting without having to go to class and practice only with this system. Also part of my future work is working on words and sentences of Persian language.

Due to limitations in some cases exist in MATLAB software, in order to upgrade the system at higher levels, in future work this system will be writing by C-language.

Reference

[1].<http://www.mathworks.com/matlabcentral/fileexchange/11123-better-skeletonization>