Available online at www.sciencedirect.com
ScienceDirect
Procedia Computer Science 3 (2011) 1055–1059
Procedia Computer Science
www.elsevier.com/locate/procedia

WCIT 2010

Farsi/Arabic Optical Font Recognition Using SIFT Features

Morteza Zahedi[a] and Saeideh Eslami[a,*]

[a]*School of Computer Engineering and Information Technology, Shahrood University of Technology, Shahrood, Iran*

Abstract

Most optical font recognition (OFR) methods have been designed to recognize the font in non-cursive documents. However, the recognition of cursive font scripts like Farsi/Arabic texts has its own challenges. Thus, most of the currently proposed algorithms fail to exhibit an appropriate recognition rate when facing cursive documents. In this paper, a new method for Farsi/Arabic automatic font recognition is proposed which is based on scale invariant feature transform (SIFT) method. As SIFT features are scale-invariant, the final system is robust against variation of size, scale and rotation. The system does not need a pre-processing stage but in the case of low quality images some noise removal processes can be used. Using a database of 1400 text images, an excellent recognition rate of nearly 100% is obtained.

Selection and/or peer-review under responsibility of the Guest Editor.

Keywords: Scale invariant feature transform (SIFT); Optical font recognition (OFR); Optical character recognition (OCR)

1. Introduction

Font recognition is a process to identify the font style of a text image. It can promote automatic classification of documents and can also be used to improve character recognition in terms of accuracy and performance by selecting a suitable database to use in the optical character recognition (OCR) process [1]. Font recognition is a basic issue in document analysis and is considered to be a complicated and time-consuming task [2]. Only a few researches have been done on the recognition of the typeface named optical font recognition (OFR) compared to the vast researches in the OCR domain. In most approaches a Bayes decision rule is employed as a classifier to find the best match between the unknown font in the given text image and the previously recognized fonts [1-5].

Most of the proposed methods are based on typographical features extracted by means of local attribute analysis but other methods often use global feature analysis [2]. In order to use local features, in the research reported in [3] representative stroke templates from a text image containing the same typeface characters are extracted, classified

* Saeideh Eslami, Tel.: +98-273-3393580; fax: +98-273-3393580.
E-mail address: saeideh_eslami@comp.tus.ac.ir.

1877-0509 © 2010 Published by Elsevier Ltd. Open access under CC BY-NC-ND license.
doi:10.1016/j.procs.2010.12.173

and saved to a database. Also, Zramdini and Ingold suggest using the features extracted from projection profiles of text lines [4]. Considering the capabilities of Gabor filter in the extraction of local characteristics of a text, Kabir and Khosravi have proposed a new enhanced filter called Sobel–Roberts and claim their method is much faster than the Gabor filter. They have reported a recognition rate of 94% on a dataset of 10 popular Farsi fonts [5]. Trying to cope with global deformations, tangent distance has been used by a neural classifier in order to distinguish samples of similar classes where some global changes and deformations have occurred [1]. Another good recognition rate reported for OFR comes from an approach which uses correlation coefficients or multi-scale wavelet analysis to gain a recognition rate of 95% [6].

Some certain characteristics of Persian (Farsi) Alphabets complicate the font recognition process. So traditional OFR methods often fail to accord with these characteristics and they are not applicable in a Farsi/Arabic OFR system. In this paper, first, these characteristics are discussed in the next section. A summary of scale invariant feature transform (SIFT) features which is used by the proposed method is presented in sections 3. While the proposed method is described in section 4, Section 5 gives the experimental results and finally future works and possible promotions of the method are discussed in the conclusion part.

2. Farsi/Arabic OFR

Compared to the vast researches done on English OCR, the Farsi/Arabic OCR is still in its early stages. This is due to the connected nature of the Farsi/Arabic alphabet [7]. Most of the Farsi/Arabic alphabets have two main forms: cursive (connected) form and isolated form. It is also important to mention that some characters have several different cursive forms. For example the character 'ع' has three different cursive shapes depending on its location in a word. Therefore we have:

عـ When it is at the beginning of the word

ـعـ When it is in the middle of the word

ـع When it is at the end of the word

This characteristic of Persian/Arabic fonts leads the Farsi/Arabic alphabets to produce uncountable different possible forms of words and characters. Hence, it is obvious that using such words for the font recognition purpose is complicated, time consuming and yields poor results.

Furthermore, font recognition has other challenges like document size, angle, characters weight (bold or regular), the amount of spaces between lines or words, etc. Some methods need to detect the base line to estimate the pen width [7]. As a result of improper scanning, the document lines may not be completely vertical. The resulting angle intervenes in the line detection process [7]. The OFR methods which use Gabor filter as a feature extractor need the spaces within the lines, words and sub-words to be normalized in order to gain correct results [2]. The font size and its style whether it is bold, italic or simple also should be defined [8]. For correct recognition, the document font in the test image should be of the same size as in the train image otherwise the font size must be normalized.

Thus, almost all OFR methods have a preprocessing stage including noise removal, angle correction, space normalization and font size estimation [9]. This stage takes a long time due to the complexity of the operation but the results are not always optimal due to poor scanning, noise, texture of the paper, etc. On the other hand, features extracted in this way are sensitive to noise, scale, rotation, size and boldness [8]. Hence, using a method which is invariant to scale, rotation and also noise occurrence is almost requested.

3. Scale Invariant Feature Transform

Scale invariant feature transform (SIFT) has been utilized in computer vision domain particularly for object recognition. The main task of SIFT is to detect and describe local features in images to help matching the different views of a certain object [10, 11]. Feature description is formed by interesting points of the object called key points. Extracting this description from the training image can help in recognizing and identifying the object in a test image among many other objects [10]. The important characteristics of these features are their robustness to noise, changes in image scale, illumination, and mild distortions which will lead to a credible recognition [11, 12]. As the features are easy to extract, recognition can be performed at almost real time, especially for small databases. An object is recognized in a new image by comparing each feature of the new image to the database of extracted SIFT key points and finding the best set of matched key points [10]. Promising candidates are those subsets of matched key points that are consistent with the object's general features in the new image. Each group of 3 or more features is verified to check the conflicts with general features of the interesting object. After discarding the outliers, the probability that a particular set of features marks an object in the image is computed [10]. Approved matches can be identified as correct recognitions.

The SIFT feature extraction algorithm consists of several filtering stages [10, 12]. In the first phase named scale-extrema detection, sets of locations and scales which are recognizable in different views of the same object are filtered out with a Gaussian scale-space function. Then difference of Gaussian (DoG) function is calculated by computing the difference between two images; one image with a scale of 'k' times the other one to locate stable key points. Key locations are the extreme points resulting from applying DoG on a set of re-sampled images. Low contrast points and poorly localized points on the edges are disposed to improve confidence in the final key points [12]. Indexing or storing the SIFT keys in the database is done in the feature matching and indexing phase. The best match for each key point in the test image is found by finding its nearest neighbor in the database of training key points simply by calculating the Euclidean distance from the given descriptor vector. Each of the SIFT key points specifies crucial parameters such as location, scale and orientation, and each matched key point in the database keeps a record of the parameters in the matched training image [11,12]. In the model verification stage, a least square-based method is used to verify each identified group of features [10]. In the last phase, the least square solution is again performed on the remaining points in order to discard the outliers. Less than 3 remaining points means the match will be rejected but if a set of 3 or more keys agree on the model`s parameters, this model is recognized in the image with a high probability [10, 12]. The main steps of SIFT algorithm are shown in Fig. 1.

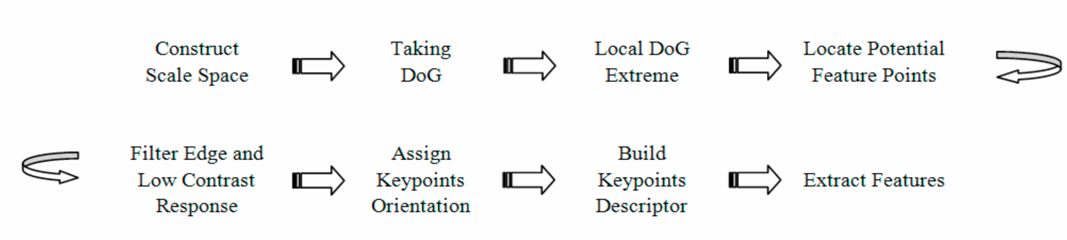


Fig. 1. SIFT feature extraction stages

4. Proposed Algorithm

SIFT has several advantages over other OFR methods. These advantages include recognition independence from font size, font weight, the angle between the lines while maintaining a correct recognition of poorly scanned documents. SIFT features are also very irrepressible against the noise. In the proposed system the main part of computation is done by SIFT. There is no need for any preprocessing stage for normal scanned documents, but in case of noisy documents a simple noise removal algorithm can be used to gain much reliable results. Also there is no need for the unknown given image to have a predetermined size or to be space normalized. Recognition can be done effectively on a single line with a variable amount of spaces between the words.

The training set includes 20 images of a special paragraph written with 20 popular Farsi fonts. The important point about the paragraph lies in its contents. It is previously mentioned that Farsi/Arabic characters can appear in isolated form or rather in a connected form. Thus, almost all possible shapes of each 32 Farsi characters are included in the paragraph. These characters take different shapes in different fonts so the system should be aware of such variations. Although using random paragraphs with known fonts still gives out satisfying results, it is more reliable to train the system with all the possible character forms and provide much more similar key points within future test images. It is also necessary for the system to be familiar with regular and frequently-used Farsi words or sub words like: “با”,”یا”,’ا’,”ا”,”را”,”می”,”است”. Fig 2 shows and illustrates an example of a prototype image.

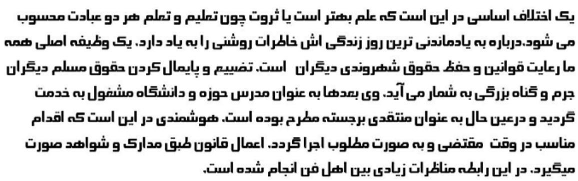


Fig. 2. A sample of training images containing the representative paragraph (consisting of all possible form of characters)

Extracted features or key points are compared and if their similarity is proved then the key points are matched. If the number of matched key points reaches a certain pre-defined minimal point, objects of interest in the test image are declared as a correct recognition in the current train image.

5. Experimental Results

A digital Cannon scanner/printer is used to provide us document images with 300 dpi resolution. A primary set of 75 text images is used to test the performance of the system. Another database including 1400 128×128 text images is used to validate the method. This database has been utilized in [5]. In order to test the system performance, numerous trials were done under various challenging circumstances including document rotation, font weight and size variation, etc. The minimum size for the test image is 128×128 but larger sizes are also possible. Here the experiments conducted on the document images reveal that if the number of matched key points exceeds 30 then the two images contain the same font. On the other hand, if this number is below the defined threshold, then the two fonts are different. It means if the number of matching key points between a test and a train image reaches up to 30 (in our database) their fonts are similar. The proposed algorithm is performed on 75 document images of 20 popular Farsi fonts and a recognition rate of 100% is achieved on this database. The experiments are done on the images with various fonts, bolded words, totally rotated images and noisy images. On the second database the same results are reached. Fig 3 and 4 illustrate some samples of recognition in test images under challenging conditions.

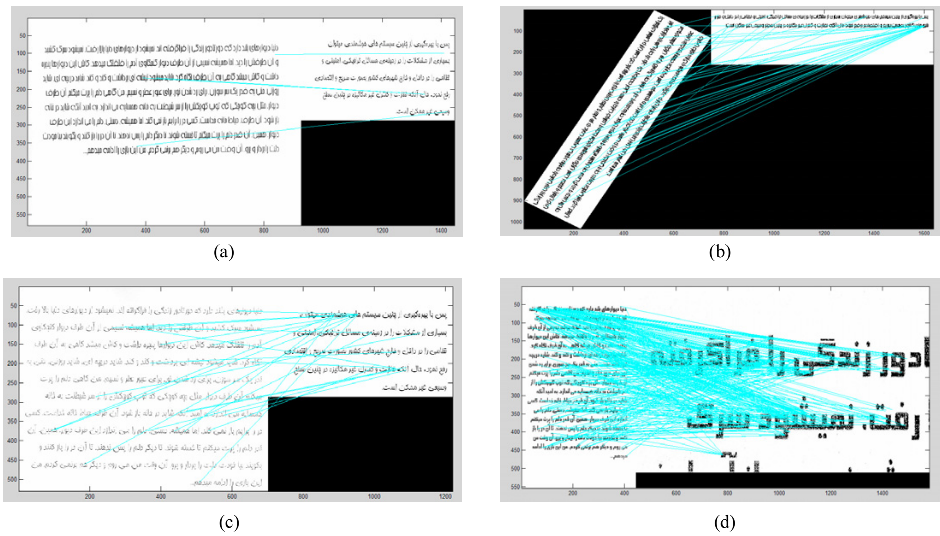
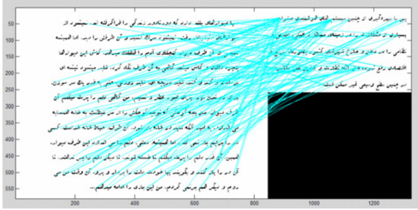


Fig. 3. (a) The results on the documents of two different fonts with the same font size. (b) Different texts, same font, rotated by 60 degrees. (c) Different texts, same font, different sizes. (d) key point matching for the right text which is a small part of the scaled image of the left one

(a)



(b)

Fig. 4. (a) Same font: one bold and the other one normal. (b) High key point matching rate on different texts with the same font and font size

6. Conclusion and Outlook

In this paper, SIFT method which extracts a set of robust and reliable features is employed for font recognition especially for Farsi and Arabic languages including cursive characters. According to the characteristics of the SIFT features that is robustness against scale, rotation, translation and also noise occurrence, a great benefit in terms of efficiency and accuracy is expected. Thus, not only a very good result with a recognition rate of 100% is obtained, but also it recognizes special fonts (like Tabassom) which are known to be weak points for other OFR methods.

Based on the analysis of the results after execution on a PC computer, the process time can be a problem when using huge databases. So, we suggest employing the speeded up robust features (SURF) method which is introduced by Herbert Bay [10]. This method is inspired by SIFT, but although having the same accuracy, SURF can work with less computational time.

Acknowledgements

Here we thank Hossein Khosravi for providing us with the used database of text images.

References

1. S. La Manna, A. Colla and A. Sperduti, Optical Font Recognition for Multi-Font OCR and Document Processing , 10th International Workshop on Database and Expert Systems Applications, Florence, Italy, September 1999, pp. 549-553.
2. Y. Zhu, T. Tan, and Y. Wang, Font Recognition Based on Global Texture Analysis, IEEE Trans. Pattern Analysis and Machine Intelligence, VOL. 23, NO. 10, October 2001, pp. 1192-1200.
3. H.M. Sun, Multi-Linguistic Optical Font Recognition Using Stroke Templates, 18th International Conference on Pattern Recognition (ICPR), VOL 2, 2006, pp.889-892
4. A. Zramdini. and R. Ingold, Optical Font Recognition Using Typographical Features, IEEE Transactions On Pattern Recognition and Machine Intelligence, VOL. 20, NO. 8, August 1998
5. E. Kabir. and H. Khosravi, Farsi font recognition based on Sobel–Roberts features, Elsevier B.V. 2009
6. E. Rashedi, E. Nezam abadipour, and S. Saryazdi, Farsi font recognition using correlation coefficients (in Farsi), 4th Conf. on Machine Vision and Image Processing, Ferdosi Mashhad, 2007.
7. M.H. Shirali-Shahreza. and S. Shirali-Shahreza, Persian/Arabic Text Font Estimation Using Dots, IEEE International Symposium on Signal Processingand Information Technology, 2006
8. B. Allier, E. Hubert, Font Type Extraction and Character Prototyping Using Gabor Filters, Proceedings of the Seventh International Conference on Document Analysis and Recognition, August 03-06, 2003, p.799.
9. A. Zramdini, Study of Optical Font Recognition Based on Global Typographical Features, Thesis of the University of Fribourg, Fribourg (Switzerland), 1995.
10. D. G., Lowe, Object Recognition from local scale-invariant features, Proceedings of the International Conference On Comuter Vision, 1999, pp. 1150–1157.
11. V.Pham, T. Miyaki, T.Yamasaki, K. Aizawa, Geometrically invariant object-based watermarking using SIFT feature, Proceedings of IEEE International Conference on Image Processing, VOL. 5, 2007, pp. 473-476.
12. D. G., Lowe, Distinctive Image Features from Scale-Invariant Keypoints, International Journal of Computer Vision, VOL. 2, NO. 60, 2004, pp. 91-110.