

# Character Recognition of Sindhi (Arabic) Script

Deepak Garg  
Student

Department of Computer Science  
Punjabi University, Patiala, Punjab (India)

**Abstract**— Optical Character Recognition is a wide area of research and is used in various applications. This paper proposes a new technique of optical character recognition using Recursive Subdivision Method. This method is very effective with the use of Recursive Subdivision Method for feature extraction. The model proposed is trained and validated for Arabic script and the results are found to be very good. The model developed works for the entire character set.

**Key words:** Recursive Subdivision Method, Optical Character Recognition, Recognition

## I. INTRODUCTION

Character recognition helps in data and word processing, automated postal address and ZIP code reading, etc. In the Recent years, optical character recognition (OCR) has gained a speed since the need for converting the scanned images into computer recognizable formats such as text documents has increased applications. OCR is one of the most difficult areas of pattern recognition with various practical applications. The process of character recognition involves extraction of defined characteristics called features to classify an unknown character into one of the known classes. Therefore, OCR involves two processes: 1. feature extraction 2.classification. The process of character recognition becomes very tough in the case of Arabic script. In Arabic script many letters look alike. So classification becomes a big challenge. In image processing, Recursive Subdivision Method is used for feature extraction. In this paper, a method is proposed for character recognition, which uses Recursive Subdivision Method to extract features and statistical method for classification. The proposed method validates training methodology. In training phase, the features extracted using Centroid method is fed to Statistical method.

## II. CHARACTER SET

CONSONANTS							
ڪ	ق	ک	ڱ	گ	ڳ	ڱ	گه
گ	ڳ	ڱ	ڱ	ڱ	ڱ	ڱ	ڱ
ت	ت	د	ڌ	ڌ	ڌ	ڌ	ڌ
ت	ت	د	ڌ	ڌ	ڌ	ڌ	ڌ
پ	پ	پ	م	ي	ر	ل	و
ڻ	ڻ	ڻ					
VOWELS							
آع	آعا	آع	اي عي	آغ	او عو	اي عي	اي عي
او عو	او عو						

Fig. 1: Character Set

## III. FEATURE EXTRACTION

The classification procedure is based on how and what number of features are being extracted from our data set.

Experiment with certain type of features extraction was done but the recursive subdivision method proved to be the best and gave us the good results.

Recursive Subdivision Method takes a binary image and following function is performed,

- 1) The function takes an image and calculates the x-coordinates.
- 2) Then the function subdivides the image into two images by the x-coordinates and step 1 is repeated.
- 3) These coordinates are the output vector which if of length  $2d-1$ , in the example described below the depth is 4.
- 4) This function is also applied on transposed image and so the 126 features are calculated.

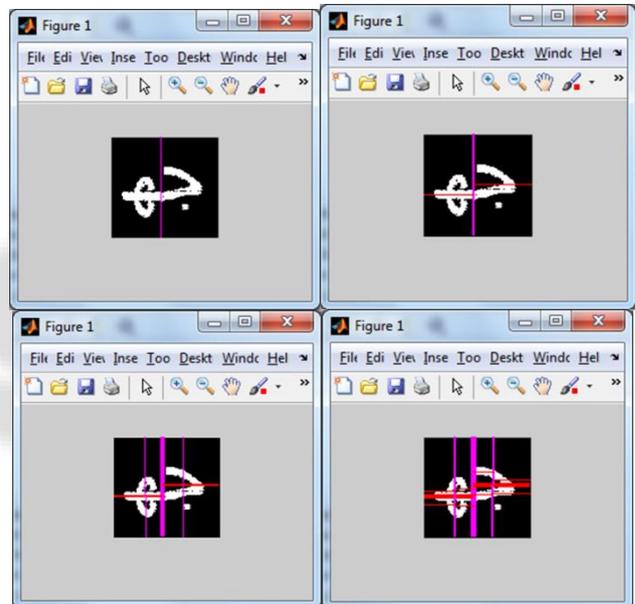


Fig. 2: Screenshot

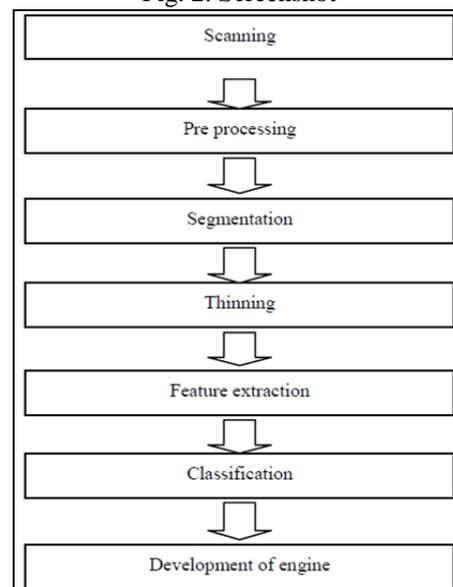


Fig. 1: Flow Chart for Training

#### IV. CLASSIFICATION

For the process of classification we have choose SVM with radial basis function as kernel. SVM is a group of supervised learning method that is used for classification, in SVM takes input data and itself it classifies into one of the classes. SVM classifier is trained by providing a set of data and upon that basis a model is prepared to classify test data. For multiple classes we have to divide the data set into as that many classes. Our data set is also divided into 53 different classes.

#### V. RESULT AND CONCLUSION

In our implementation we have used 10-fold cross validation. The average recognition accuracy is 95.461%. Thus we conclude that we have obtained the maximum recognition rate. We also presented an efficient algorithm for classification of characters using recursive subdivision method. The system was applied for the recognition of scanned images of Sindhi Arabic script characters, which included the entire character set. The work can be extended to increase the results by using some more relevant features.

#### REFERENCES

- [1] S. Singh, A. Aggarwal, R. Dhir, "Use of Gabor Filters for Recognition of Handwritten Gurmukhi Character", Vol. 2, Issue 5, May 2012.
- [2] S. Armon, "Handwriting Recognition and Fast Retrieval for Hebrew Historical Manuscripts", Dec 2011.
- [3] U. Pal, S. Chanda, T. Wakabayashi, F. Kimura "Accuracy Improvement of Devnagari Character Recognition Combining SVM and MQDF"
- [4] Kanimozhi. V.M, Muthumani.I "OCR for English and Tamil Script", IJCSIT, Vol. 5(2), 2014.
- [5] G. S. Lehal, C. Singh, "A Complete Machine printed Gurmukhi OCR", Vivek, 2006.
- [6] B. Anuradha SRINIVAS, A. AGARWAL, and C.Raghavendra RAO, "An Overview of OCR Research in Indian Scripts", IJCSES, Vol. 2, No. 2, April 2008.
- [7] G. S. Lehal, A. Rana, "Recognition of Nastalique Urdu Ligatures", Proceedings of 4th International Workshop of Multilingual OCR, Article No. 6, Washington DC, USA. (2013).