

A Review on Offline Handwritten Recognition of Devnagari Script

Snehal S.Patwardhan
M.Tech(CSE) Student,
Dept. Of CS & IT,
Dr.B.A.M.University,
Aurangabad,Maharashtra, India

R.R. Deshmukh, Ph.D
Professor & Head,
Dept.Of CS & IT,
Dr.B.A.M.University,
Aurangabad,Maharashtra, India

ABSTRACT

In country like India, where many languages and scripts exist , Devanagari is third most widely used script, used for several major languages such as Marathi ,Hindi, Sanskrit, and Nepali, and is used by more than 500 million people. Character Recognition is a process by which a computer recognizes letters, numbers, or symbols and turns them into a digital form that a computer can use and it is an active field of research today. It comprises of Pattern Recognition and Image Processing. Character Recognition is broadly categorized into Optical Character Recognition (OCR) and Handwritten Character Recognition (HCR). OCR system is most suitable for the applications like multi choice examinations, printed postal address resolution etc, while application of HCR is wider as compared to OCR. Feature extraction and classification are essential steps of character recognition process affecting the overall accuracy of the recognition system. There is not much work has been done towards off-line handwriting recognition of Devanagari script. This paper gives a detailed overview of different feature extraction and classification techniques for recognition process Devanagari script by the researchers over the past few decades.

General Terms

Pattern Recognition, Indian Scripts, Optical CharacterRecognition.

Keywords

Devanagari Character and Numeral Recognition, Offline Handwritten Recognition, Feature Extraction, Classification

1. INTRODUCTION

OCR is abbreviated as Optical Character Recognition refers to the branch of computer science that involves reading text from paper and translating the images into a form that the computer can manipulate (for example, into ASCII codes). An OCR system enables you to take a book or a magazine article, feed it directly into an electronic computer file, and then edit the file using a word processor.OCR, is the mechanical or electronic conversion of scanned or photographed images of The character recognition is divided into two types as follows: 1) Printed Characters: It is also divided into two types i) Good Quality Printed Documents & ii) Degraded Printed Documents 2) Handwritten Characters. It is divided into follows: i) Offline & ii) Online[1]

Rest of the paper is organized as follows. In Section 2 properties of Indian Scripts are discussed. In Section 3 characteristics of devnagari scripts are given. Process of character recognition is described in section 4. Feature extraction and classification techniques used in Devnagari script is given in Section 5. Result using comparison table are discussed in Section 6. Finally section 7 is based on the Conclusion.

2. PROPERTIES OF INDIAN SCRIPTS

India is a multi-lingual multi-script country and there are twenty three languages, comprising official languages namely, Assamese, Bengali, Dogri, English, Gujarati, Hindi, Kannada, Kashmiri, Konkani, Maithili, Malayalam, Manipuri, Marathi, Nepali, Oriya, Punjabi, Sanskrit, Santhali, Sindhi, Tamil, Telugu and Urdu. There are 14 different scripts in India Assamese, Bangla, Tamil, Devanagari, Gujarati, Gurumukhi, Kannada, Kashmiri, Malayalam, Oriya, Roman, Telugu and Urdu used for writing these languages. Indian scripts are different from non-Indian scripts in several ways. Indian scripts are composition of symbols like: consonants and modifiers. In Indian scripts case sensitivity is absent. Recognition of handwritten Indian scripts is difficult because of the presence of numerals, vowels, consonants, vowel modifiers and compound characters. As compared to non-Indian scripts, the research on OCR of handwritten Indian scripts has not achieved that perfection. Few attempts have been carried out on the recognition of Devanagari, Bangla, Tamil, Oriya, Telugu and Gurumukhi handwritten scripts. There is no complete OCR system is available for commercial use for recognition of handwritten text in any Indian script.[2]

3. CHARACTERISTICS OF DEVNAGARI SCRIPTS

Devnagari script is different from Roman script in several ways. This script has two-dimensional compositions of symbols: core characters in the middle strip, optional modifiers above and/or below core characters. Two characters may be in shadow of each other. While line segments (strokes) are the predominant features for English, most of the characters in Devnagari script is formed by curves, holes, and also strokes. In Devnagari language scripts, the concept of upper-case, the lower-case characters is absent. It consists of 14 vowels and 36 consonants. Vowels occur either in isolation or in combination with consonants. Apart from vowels and consonants characters called basic characters, there are compound characters in Devnagari script, which are formed by combining two or more basic characters. Coupled to this in Devnagari script there is a practice of having twelve forms of modifiers with each for 36 consonants , giving rise to modified shapes which, depending on whether the modifier is placed to the left, right, top or bottom of the character. The net result is that there are several thousand different shapes or patterns, which makes Devnagari OCR more difficult to develop.[3]

4. PROCESS OF CHARACTER RECOGNITION

The entire process of character recognition consists of various phases as described in the following Figure. These phases can be applied to the image that is obtained from either the scanned document or photograph. Optical Character recognition system comprises of following 5 steps.[4]

4.1 Image acquisition:

The image acquisition takes place with the help of scanning process digital image of the document is captured. Digitization is the process where a document is scanned and an electronic representation of the original, in the form of a bitmap image, is produced. Digital imaging or digital image acquisition is the creation of digital images. Digitization produces the digital image, which is fed to the pre-processing phase.[5]

4.2 Pre-processing:

Preprocessing is used for skew detection/correction, skeletonization, and noise reduction/removal. Skewness refers to the tilt in the bit mapped image of the scanned paper for OCR. It is usually caused if the paper is not fed straight into the scanner. Skeletonization is used for decreasing the line width of text from many pixels to single pixel. Noise removal is used to remove unwanted bit pattern which does not play any significant role in document. Thus, pre-processing is required which involves elimination of noise, Binarization of the image and segmentation (line, word and character level).[6]

4.3 Feature Extraction:

The objective of feature extraction is to capture the essential characteristics of the symbols. Being the most important and crucial step of the recognition process, selection of the feature extraction technique becomes important factor in achieving the high recognition performance. It can be said to be one of the most difficult problems of pattern recognition. Some feature extraction methods are Template matching, Deformable Templates, Zoning, Projection Histogram, Contour Profile. Table I, shows the all feature extraction methods for the various representation forms like gray level, binary, vector.[7]

Table I. Overview of feature extraction methods for the various representation forms (gray level, binary, vector)

Gray scale subimage	Binary		Vector (skeleton)
	Solid symbol	Outer contour	
Template matching	Template matching		Template matching
Deformable templates			Deformable templates
Unitary transforms	Unitary transforms		Graph description
	Projection histograms	Contour profiles	Discrete features
Zoning	Zoning	Zoning	Zoning
Geometric moments	Geometric moments	Spline curve	
Zernike moments	Zernike moments	Fourier descriptors	Fourier descriptors

4.4 Classification:

Classification is the process of assigning the sensed data to their corresponding class with respect to groups with homogeneous characteristics, with the aim of discriminating multiple objects from each other within the image. Classification is carried out on the basis of stored features in the feature space, such as structural features, global features etc. It can be said that classification divides the feature space into several classes based on the decision rule. Some classification techniques used in Optical character recognition systems are Classifier like Projection distance (PD), Subspace method (SM), Linear discriminant function (LDF), Support vector machines (SVM), Modified quadratic discriminant function (MQDF), Mirror image learning (MIL), Euclidean distance (ED), Nearest neighbour, k-Nearest neighbour (k-NN), Modified Projection distance (MPD), Compound

projection distance (CPD), and Compound modified quadratic discriminant function (CMQDF) are considered.[8]

4.5 Post-processing:

Post-processing step involves grouping of symbols. The process of performing the association of symbols into strings is referred to as grouping.[9]

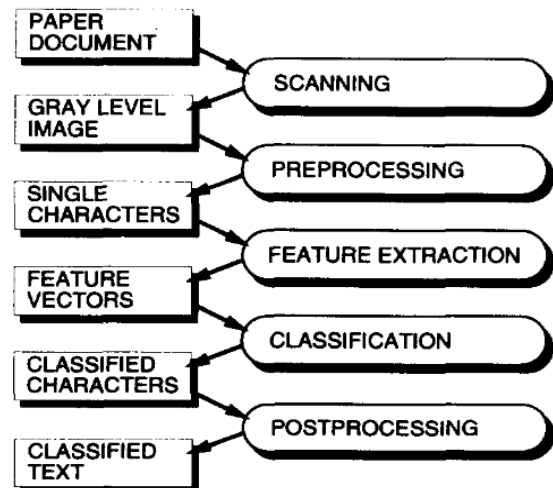


Fig 1: Steps in Character Recognition System

5. FEATURE EXTRACTION AND CLASSIFICATION TECHNIQUES USED FOR DEVNAGARI SCRIPTS

N. Sharma, U. Pal, F. Kimura, and S. Pal [10] designed quadratic classifier based system for Devanagari script. Histograms of direction chain code of the contour points of the characters are used as feature for recognition. Here they use 64 dimensional features and Modified Quadratic Discriminant function (MQDF) is used by the quadratic classifier. They achieved 94% accuracy.

Veena Bansal and R.M.K Sinha [11] presented a complete OCR for printed Hindi text written in Devanagari script. The system used following features: Coverage of the region of the core strip, Vertical bar feature, Horizontal zero crossings, Number of positions of the vertex points, Moments, Structural descriptors of the characters for classification, Tree classifiers are used. Overall accuracy obtained at the character level is 93%.

Sinha and Mahabala [12] designed a syntactic pattern analysis system for Devanagari script recognition. The system stores structural descriptors for each symbol of the script. They achieved 90% accuracy. Sandhya Arora [13] have used Intersection features with Neural Network for Devanagari script and they achieved 89.12% accuracy.

A.Bharat & Sriganesh Madhavath [14] presented HMM based Lexicon driven & free word recognition by using angle features with the help of HMM classification. They got 85.56% accuracy. Holambe and Thool, Jagade[15]designed a support vector machine classifier based system for both Printed and Handwritten Character & Number Recognition using Gradient features and they will achieved 94% accuracy for both systems.

S.B.Patil, G.R.Sinha, V.S.Patil [16] designed an Isolated handwritten devanagari numerals recognition system Using HMM classifier.They achieved 99% accuracy. G.Sinha,

R.Rani,R.Dhir [17] designed a handwritten devanagari numeral recognition using zonal based feature extraction method and SVM classifier. They achieved 99.11% accuracy.

A comparative study was done by Pal et al. [18] on Devanagari handwritten character recognition using 12 different classifiers like PD, subspace method (SM), linear discriminant function (LDF), SVM, MQDF, mirror image learning (MIL), Euclidean distance (ED), nearest neighbor, k-NN, modified PD (MPD), compound PD (CPD), and compound MQDF (CMQDF). The method proposed by Pal et al. has the highest accuracy, as shown in table. Chain code and gradient-based features are used for Devanagari numeral recognition. Even for recognizing Devanagari handwritten characters, the method proposed by Pal et al. [19] has the highest accuracy.

A continuous density HMM is also proposed by Shaw et al. [20] to recognize a handwritten word images. The states of the HMM are not determined a priori, but are determined automatically based on a database of handwritten word images. An HMM is constructed for each word. To classify an unknown word image, its class conditional probability for each HMM is computed. The class that gives highest such probability is finally selected.

Kumar [21] compared the performances of SVM and MLP classifiers with six different features on handwritten characters and found that the performance of SVM classifier was superior to MLP in all the six cases. But the classification time required for SVM was greater than that of MLP.

6. RESULT USING COMPARISON TABLES

Many researchers have proposed various techniques for offline handwritten Devanagari Numerals and Characters. The results are given below:

Table II: Comparison of Feature Extraction ClassificationMethods for Devnagari Characters

Sr No.	Feature Extraction Methods	Classification Methods	Recognition Rate(in %)
1.	Chain code	Quadratic classifier	94.28
2	Vertical Feature Bar,Horizontal Zero crossing, Number of positions on the vertex point, Structural description	Tree Classifier	93
3	Syntactical Analysis	Tree Classifier	90
4	Intersection Feature	Neural Network	89.12
5	Angle feature	HMM	85.56
6	Gradient Feature	SVM	94

Table III:Comparison table of Feature Extraction &Classification Methods for Devnagari Numerals

Sr No.	Feature Extraction Methods	Classification Methods	Recognition Rate(in %)
1	Gradient Feature	SVM	94
2	Moment And Density feature	SVM&KNN Classifier	97.89 & 96.53
3	MultiLayerPerceptron	HMMClassifier	99
4	Zonal based feature	SVM classifier	99.11

6. CONCLUSION

In this paper, we presented work done on handwritten Indian scripts. Firstly we gives properties of Indian script as well as properties of devnagari script. Then we discussed process of complete character recognition system with the help of diagram. After that we explore the techniques and methods, developed for recognition of particular Indian script ie. Devanagari. A lot of work has been done on Devanagari and Bangla handwritten characters recognition. Little work has been done to recognize the Oriya, Kannada, Tamil, Malayalam and Gurumukhi handwritten script. But till now there is no complete recognition system is available for recognition of Indian scripts. As such, there is a need for a handwriting OCR for Indian script that can help the people for converting the handwritten text ie. characters and numerals to computer processable format. To get idea of the recognition results of different classifiers and to provide new benchmark for future research, in this paper a comparative study of Devnagari handwritten character and numeral recognition results is reported in the form of comparison table. In future work, these techniques used for other offline handwritten Indian scripts so that accuracy of recognition can be prefect as Devnagari character.

7. ACKNOWLEDGMENT

The review has been partially supported by Department of Computer Science & IT, Dr. Babasaheb Ambedkar Marathwada University, and Aurangabad. The views expressed here are those of the authors only.

8. REFERENCES

- [1] U. Pal, and B. B. Chaudhari, "Indian Script Character Recognition: a Survey", IJSER, vol. 37, Issue 9, pp. 1887-1899, 2004.
- [2] Ghosh D., Tulika Dube, and Adamane P. Shivaprasad, "Script Recognition—A Review", IEEE Transactions on pattern analysis and machine intelligence, vol. 32, no. 12, december 2010.
- [3] U. Pal, N. Sharma, T. Wakabayashi and F. Kimura, "Off-Line Handwritten Character Recognition of Devnagari Script", In Proc. 9th ICDAR, pp.496-500, 2007.
- [4] R. Plamondon and S. N. Srihari, "On-Line and off-line handwritten recognition: A comprehensive survey", IEEE Trans on PAMI, Vol.22, pp.62-84, 2000.

- [5] U. Pal, T. Wakabayashi, and F. Kimura, “Comparative study of Devanagari handwritten character recognition using different features and classifiers,” in Proc. 10th Conf. Document Anal. Recognit., pp. 1111–1115, 2009.
- [6] P. M. Kakde, S.M.Gulhane, “Handwritten Devnagari Script Recognition :A Review”, IJETAE,ISSN 2250-2459, ISO 9001:2008 Certified Journal, Volume 4, Issue 9, September 2014.
- [7] A.S.Ramteke, M.E. Rane, “A Survey on Offline Recognition of Handwritten Devanagari Script” International Journal of Scientific & Engineering Research Volume 3, Issue 5, ISSN 2229-5518, May-2012.
- [8] R. Jayadevan, S. R. Kolhe, P.M. Patil, & U.Pal, “Offline Recognition of Devanagari Script: A Survey”, IEEE transactions on systems, man, and cybernetics—part c: applications and reviews. Nov 17, 2010.
- [9] S. M. Mali , “Moment And Density Based Hadwritten Marathi Numeral Recognition”, ISSN : 0976-5166 Vol. 3 No.5 Oct-Nov 2012
- [10] N. Sharma, U. Pal, F. Kimura, and S. Pal, “ Recognition of Off-Line Handwritten Devnagari Characters Using Quadratic Classifier”, P. Kalra and S. Peleg (Eds.): ICVGIP 2006, LNCS 4338, pp. 805 – 816, 2006.© Springer-Verlag Berlin Heidelberg 2006.
- [11] Veena Bansal, M K Sinha, “A Complete OCR for printed Hindi Text in Devanagari Script”, IEEE 800 – 804, 2001 .
- [12] Sinha. M. K., Mahabala., “Machine Recognition of Devnagari Script”, IEEE T. SYST. MAN Cyb., vol.. 9, pp.435-449, 1979.
- [13] Sandhya Arora , “A Two Stage Classification Approach for Handwritten Devanagari Characters”, IEEE 399 - 403 vol 2.
- [14] A.Bharat, S. Madhvanath , “HMM Based lexicon-driven & lexicon-free Word Recognition for Online Handwritten Indic Scripts”, IEEE transc vol 34, no.4, April 2012.
- [15] A.N. Holambe, R.C.Thool, S.M.Jagade, “Printed and Handwritten Character & Number Recognition of Devanagari Script using Gradient Features”, International Journal of Computer Applications (0975 – 8887) Volume 2 – No.9, June 2010
- [16] S.B. Patil, G.R.Sinha, V.S.Patil, “Isolated Handwritten Devnagari Numerals Recognition Using HMM” 978-0-7695-4329-1 © 2011 IEEE DOI 10.1109/EAIT, 2011.
- [17] G.Sinha, R.Rani, R.Dhir, “Handwritten Devanagari Numeral Recognition Using Zonal Based Feature Extraction Method and SVM Classifier”, IJARCSSER Volume 2, Issue 6, ISSN: 2277 128, June 2012.
- [18] U. Pal, T. Wakabayashi, and F. Kimura, “Comparative study of Devanagari handwritten character recognition using different features and classifiers,” in Proc. 10th Conf. Document Anal. Recognit., pp. 1111–1115, 2009.
- [19] U. Pal, T. Wakabayashi, N. Sharma, and F. Kimura, “Handwritten numeral recognition of six popular Indian scripts,” in Proc. 9th Conf. Document Anal. Recognition, pp. 749–753, 2007.
- [20] B. Shaw, S. K. Parui, and M. Shridhar, “Off-line handwritten Devanagari word recognition: A holistic approach based on directional chain code feature and HMM,” in Proc. Int. Conf. Inf. Technol., pp. 203–208, 2008.
- [21] S. Kumar, “Performance comparison of features on Devanagari handprinted dataset,” Int. J. Recent Trends, vol. 1, no. 2, pp. 33–37, 2009.
- [22] Kumar, D., “AI approach to hand written Devanagri script recognition”, The Proceedings of IEEE Region 10th International Conference on EC3-Energy, Computer, Communication and Control Systems, vol. 2, pp. 229–237, 2008.