An Analytical Study of Handwritten Devanagari Character Recognition

Sonal P Ajmire
ME-I (CS & IT)
HVPM, COET, Amravati, India

K. G. Bagade
Dept. of IT
HVPM, COET, Amravati, India

P. L Ramteke
Head, Dept. of IT
HVPM, COET, Amravati, India

Abstract— The concept of character recognition has gained a lot of attention due to its many applications such as in printed postal addressing, filling of various forms, multiple choice questions in certain examination and so on. This paper is an analytical study of handwritten Devanagari character recognition. It describes the statistical techniques used for feature extraction for handwritten character recognition. The character recognition can be of two types namely online character recognition and offline character recognition. This can be further classified as printed character recognition and handwritten character recognition. The handwritten character recognition has more applications as compared to the printed character recognition.

Keywords: Pattern Recognition, Classification, Feature Extraction, Character recognition, Devanagari character.

I. INTRODUCTION

Pattern recognition is the study of how machine can recognize the object or image. Many researchers are working in the field of pattern recognition from last five decades, in spite of this design of a general purpose machine pattern recognizer remains an elusive goal [1]. There are three main approaches for pattern recognition system, statistical, structural and neural. Among these techniques, statistical techniques are always favored in practical application due to their robust characteristics and simple training scheme [2]. Character Recognition is mainly classified as online character recognition and offline character recognition. It is further classified as printed character recognition and handwritten character recognition. This paper deals with the analytical study of handwritten offline Devanagari character recognition. There are around 238 scripts and 6500 different languages. In India 22 languages are officially recognized [3].

Devanagari script is majorly used in various states of India. Many languages use Devanagari script for example Marathi, Hindi, Sanskrit, Nepali, Kokani and Sindhi. The script is used by more than 450 million people in the world [4]. Linguistic data consortium for Indian languages published the information about Indian languages with its spoken percentage and it is given in the following figure 1.

Most people in India use Hindi language. Another language that has been mostly used is Bengali and secondly, Marathi language.
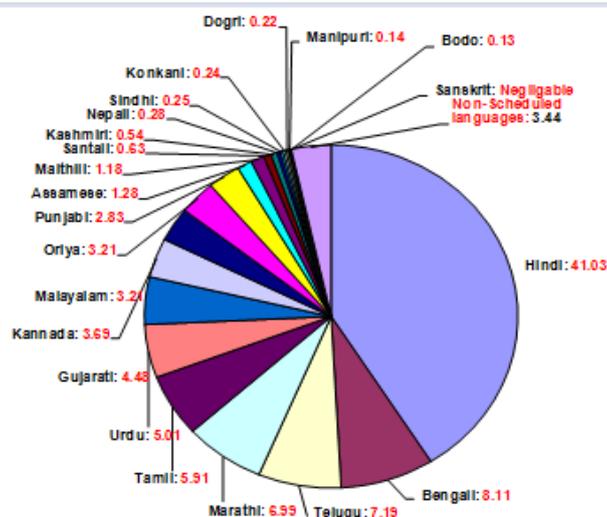

FIG.1: INDIAN LANGUAGES WITH ITS SPOKEN PERCENTAGES.

Devanagari Script is different from Roman Script in several ways.
➢ Devanagari script is two dimensional composition of symbols four character in middle stream, optional modifiers above and/or below core character.

k , ka , ik , kI

➢ A character may have many different forms and the script specified rules for their usage in different context.
➢ Two characters are required to touch in some cases.

sya , tva

Devanagari script has 52 characters which can be written as individual symbols in a word[5]. Most of the characters have half form combine with another character. Such characters are known as compound character [6]. Devanagari Script consists of vowels and consonant. Vowels can be used as individual character as well as it can be combined with consonants. These vowels are also known as modifiers. Every word has header line placed above the core character called as "*Shirorekha*". A single Devanagari character may have lower modifier and upper modifier.
The vowels in devanagari are shown below.

A Aa [[_ { {] P e eo Aao AaO AM A:

The consonants in devanagari are shown below.

k Ka ga xGa D:
ca C ja Ja x~a
T z D Z Na
ta qa d Qa na
pa f ba Ba ma
ya r la va Ya
Sa sa h L Xa क्र

II. LITERATURE SURVEY

Many researchers have been working on the character recognition of pattern matching field. This is because the field of pattern matching has many applications. The matching of Iris of the person becomes helpful for recognizing the person and giving the unique identification number. Thus providing from fraud and giving full security to the identity of a person. The pattern recognition can be used in image processing, segmentation, analysis, computer vision, seismic analysis, radar signal classification/analysis, face recognition, speech recognition/understanding, fingerprint identification, character recognition, handwriting analysis, electrocardiographic signal analysis/understanding, medical diagnosis. In India, people speak hundreds of languages. The first research paper on handwritten devanagari character was published in 1977 by I.K. Sethi and B. Chatterji [7]. This script is a foundation of Indian languages so it plays a very important role in the development of literature and manuscripts.

In 2011, S. Arora et al [8] use Support vector machine and artificial neural network for classification of handwritten devanagari character. They achieved 98.16% recognition rate. J. Pradeep et al [9] used diagonal based feature extraction method and neural network for training and classification and they achieved the recognition accuracy of 97.8%.

In 2012, Mrs. V. S. Tapkari et al proposed algorithm based on projection profiles and achieved 98% accuracy.

In 2014, K. V. Kale et al used Zernike moment based technique for feature extraction. They proposed a system using SVM and K-NN classifier and got recognition rate 98.37% and 95.82% respectively.

The table 1 below shows the digit or character, the feature extraction method, classification method, reference number and accuracy of the methods that has been used. The accuracy for every method is above 80%.

III. STAGES IN CHARACTER RECOGNITION

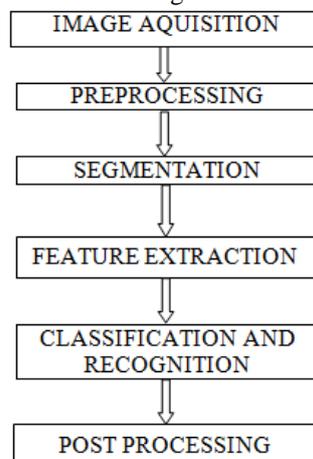The following figure shows the stages in character recognition.

FIG.2: STAGES IN CHARACTER RECOGNITION

In machine learning process every character is to convert in digital form. With the help of scanning process, digital image of the document is captured. Following stages are involved in character recognition.

i. Image Acquisition: Converting paper document in the form of digital image with the help of scanning process is called Image Acquisition.

ii. Preprocessing: The digital image obtained from scanning may contain some amount of noise depending upon the quality of scanner. The removing of this noise from captured image is called preprocessing. Noise might be skewed or broken character. The main advantage of pre-processing a handwritten character is to organize the information so as to make the recognition simpler. Along with the noise reduction, normalization and segmentation of image is also done to make the further recognition process easy. In normalization, resizing of characters is done. Normalization is done for stroke width, slant, slope, height of the characters. The normalization task will reduce each character image to one consisting of vertical letters of uniform height and made up of one pixel-wide strokes[10].

iii. Feature Extraction: Features are the information extracted from the captured or input image. This information must be similar for similar images but it must be distinct for other images. These features plays major role in pattern recognition. Thus the selection of feature extraction technique becomes important factor in achieving high recognition performance. Some feature extraction methods are: projection, zoning, bordered transition, graph matching [11]. The projection method does the compression of the data through a projection. Black pixel counts are taken along parallel lines through the image area to generate marginal distributions. Border transition technique assumes that all the characters are oriented vertically. Each character is partitioned into four equal quadrants. Zoning is a method that involves the division of the character into smaller fragment of areas. The black pixels in each zone are counted and accumulating or averaging the profiles in each zone extracts features[11].

iv. Classification: Classification is carried out on the basis of features of the image. Classification is the process of assigning features data to their corresponding class with respect to groups with homogenous characteristics. In this way, classification divides the feature space into several classes based on the decision rule. Some classification techniques used for recognition of handwritten character are neural network, support vector machine, K-nearest neighbours, Bayseian classification and decision tree classification.

v. Post Processing: Post processing means grouping of symbols into string. The accuracy of Optical Character Recognition can be increased if the output is constrained by a list of words that are allowed to occur in a document. The output stream may be plain text or file of characters [12].

IV. CONCLUSION

The handwritten Devanagari character recognition has developed remarkably in the last decade, though it has been four decade old. In this paper we have specially studied the Devanagari characters and numeral recognition. This analytical study is summarized in the following table.

TABLE 1: ANALYTICAL STUDY OF DEVANAGARI CHARACTERS AND NUMERALS.

Sr. No.	Digit/Character	Feature extraction method	Classification method	Ref	Accuracy
1	Devanagari	Syntactical analysis	Tree classifier	13	90%
2	Devanagari	Gradient features	SVM	14	94%
3	Devanagari	Directional gradient Gaussian filter	Modified Quadratic Classifier	15	94.24%
4	Marathi	Structured based	Rule based	16	85-90%
5	Marathi	Adaptive smoothing	HMM	17	98%
6	Marathi digit	Fourier descriptor	K-NN	18	97.04%
			SVM		97.85%
7	Devanagari digit	Geometric	Combination classifier	19	81.67%

From the above analytical study it is found that statistical features gives better result as compared to structural. Similarly Support Vector Machine is better classifier.

ACKNOWLEDGEMENT

REFERENCES

[1] A.K.Jain, Robert P.W.Duin and J.Mao, "Statistical Pattern Recognition: A Review", IEEE transactions on Pattern Analysis and machine intelligence, Vol. 22, no. 1, PP: 4-37, January 2000.

[2] P. E. Ajmire, R. V. Dharaskar, V. M. Thakare "Statistical Techniques for feature extraction for handwritten character recognition"- A Survey, National conference on recent advances in electronics and computer science, Science college, Nagpur.

[3] A Raj, R Shrivastav, T Patnaik, B Kumar, "A Survey of feature extraction and classification techniques used in character recognition for Indian script", International journal of engineering and advanced technology(IJEAT), Vol.2, Issue 3, PP:566-568, February 2013.

[4]   Veena Bansal and R. M. K Sinha, "Integrating knowledge sources in devanagari text recognition system", IEEE transactions on systems, man and cybernetics. Vol. 30 no.4, PP 500-505, July 2000.

[5]   Veena Bansal and R. M. K Sinha, "Designing a front end OCR system for machine translation- A Case Study for Devnagari.", IIT Kanpur.

[6]   P. E. Ajmire, R. V. Dharaskar, V. M. Thakare, "Handwritten Devanagari (Marathi) Compound Character Recognition Using Seventh Central Moment", International Journal of innovative Research in Computer and Communication Engineering Vol.3, Issue 6, PP 5312-5319, June 2015.

[7]   I.K. Sethi, B. Chatterjee, "Machine recognition of constrained hand printed devanagari", Pattern recognition, The journal of pattern recognition society, Vol. 9 , Issue. 2, PP: 69-75, July 1977.

[8]   S. Arora, D. Bhattacharyaji, M. Nasipuri, D. K. Basu and M. Kundra, "Complementary features combined in MLP- Base system to recognize handwritten Devanagari character", Journal of information hiding and multimedia signal processing. Vol. 2, No. 1, January 2011.

[9]   J. Pradeep, E. Shrinivasan and S. Himavathi, "Diagonal based feature extraction for handwritten alphabets recognition system using neural network", International journal of computer science and information technology, Vol. 3, no. 1, February 2011.

[10]   P. E. Ajmire, R. V. Dharaskar, V. M. Thakare, "Handwritten Devnagari (Marathi) Compound Character Recognition Using Seventh Central Moment", International Journal of innovative Research in Computer and Communication Engineering, Vol.3, Issue 6, PP 5312-5319, June 2015.

[11]   www.ijetae .com/files/Volume2Issue9/IJETAE_0912_07.pdf

[12]   S. G. Dedgaonkar, A. A. Chandavale, A. M. Sapkal, "Survey of Methods for character recognition", International Journal of Engineering and Innovative Technology (IJEIT), Vol. 1, Issue 5, PP: 180-189, May 2012.

[13]   R. M. K. Sinha and Mahabala, "Machine recognition of devanagari script", IEEE transactions on systems, man and cybernetics. Vol. 9, PP: 435-449, 1979.

[14]   Ashutosh Agrawal, Rajnish Rani and Renu Dhir, "Handwritten devanagari character recognition using gradient features", International Journal of advanced research in computer science and software engineering, Vol. 2, Issue. 5, PP: 85-90, May 2012.

[15]   U. Pal, N. Sharma, T. Wakabayashi and F. Kimura, "Offline handwritten character recognition of devanagari script", IEEE , International conference on document analysis and recognition, ICDAR, 2007.

[16]   C. Namarata Mahender and K. V. Kale, "Structural based feature extraction of handwritten Marathi word", International Journal of computer application, Vol.16, no.6, PP : 42-47, February 2011.

[17]   Mrs. A. P. Jane, M. A. Pound, "Recognition of similar shaped handwritten Marathi characters using artificial neural network", Global journal of computer science and technology neural and artificial intelligence, Vol. 12, Issue 11, 2012.

[18]   G. G. Rajput and S. M. Mali, "Fourier Descripter based isolated Marathi handwritten numeral recognition", International journal of computer applications, Vol. 3, No. 4, June 2010.

[19]   V. J. Dongre, V. H. Mankar, "Devanagari Handwritten Numeral Recognition using Geometric Features and Statistical combination Classifier", International Journal on Computer Science and Engineering, Vol. 5, no. 10, October 2013.