

An Analytical Study of Devanagari Script Recognition

S E Warkhede

V M Thakare

P E Ajmire

Abstract- Devanagari is an ancient Indian script. This script is used to write more than 100 languages spoken in India and Nepal including Sanskrit, Hindi, Marathi, Bhojpuri, Nepali, Kokani, Sindhi, Marwari, Maithili, Pali and many more. The word "Devanagari" is a combination of two words-"Deva" which means God and "Nagari" which means urban establishment. Thus, "Devanagari" words mean "Script of the God". It comprises of 49 primary alphabets, 13 vowels, and 36 consonants, and 10 digits. In addition, the alphabets are modified when a vowel is added to a consonant. This script also supports various compound characters. Due to this Devanagari script recognition is a challenging task and it has been an active area of research till now. The researcher used various recognition techniques like artificial neural network, support vector machine, particle Swarm optimization technique and genetic algorithm. This paper deals with the analytical study of different classification techniques used for Devanagari character recognition.

Keywords- Alphabets, Devanagari, Character recognition, optimization technique.

I. INTRODUCTION

The handwritten character recognition is an online handwritten character recognition and offline character recognition. The task of handwritten character recognition includes reading handwriting and recognition of character that closely resembles human performance. For machine character recognition is a similar task to human intelligence, which is real challenge and hence handwritten character is still an open problem. There are thousands of languages through the world and the complexity of character recognition differs from language to language. The handwritten Devanagari character recognition is complex task because it presents multiple modifiers (upper and lower) compound characters and presence of *Shirorekha*. The use of *Shirorekha* is user dependent. Some user may not prepare the use of *Shirorekha*. In offline character recognition the written text is captured optically by using the camera or scanner and this is converted into an image form. This image is used as input for offline recognition system. The online character recognition is much more complex as compared to offline because it is a real time processing of stroke detection. The research of online

character recognition started in 1960's and has been receiving intensive interest from 1980's. An online handwritten character written on a digitizing tablet is represented as a sequence of stroke, which are the loci of the pen tip from its pen-down to pen-up positions [1].

Historically, India is multilingual country and Indian constitution recognizes 22 languages which are written with 11 different script and nearly 6000 dialects used in different states of the country. Hindi, Sanskrit, Marathi, Nepali and Maithili used the Devanagari script for written and oral communication [2].

Devanagari script has 13 vowels and 36 consonants and they are shown in following figures

अ आ इ ई उ ऊ ऋ
ए ऐ ओ औ अं अः

Figure 1: 13 Vowels in Marathi Language

क	ख	ग	घ	ङ	च	छ	ज	झ
ञ	ट	ठ	ड	ढ	ण	त	थ	द
ध	न	प	फ	ब	भ	म	य	र
ल	व	श	ष	स	ह	क्ष	त्र	ज्ञ

Figure2: 36 Consonants in Marathi Language

II. PATTERN RECOGNITION MODEL

The pattern recognition model has two tasks; the first and most important task is to extract the features from the input data set / image. The second most vital part is pattern classification stage. This stage will allocate an unknown character sample to one of feasible classes by using the data of feature extraction stage. The distinct types of classifiers can be used to identify the extracted features. Any character recognition system goes under following steps, i.e. Image acquisition, Preprocessing, Segmentation, Feature extraction, classification and post processing [3]. The detail diagrammatic representation of this pattern recognition model especially for

handwritten character recognition as shown in figure 3. Since last four decades, significant research activities have been devoted to character recognition. The handwritten character recognition means making and an interpretation of human understandable characters to machine-understandable codes. Immense activities has been made on character recognition, as it gives an answer for handling substantial volumes of information naturally in an expansive variety of scientific and business applications. Optical character recognition (OCR) manages the recognition of optically handled characters instead of attractively prepared ones. OCR is a procedure of automatic recognition of characters by PCs in optically examined and digitized pages of text [4]. OCR is the most interesting and challenging area of pattern recognition with different valid applications. It can contribute extremely to the progression of a computerization procedure and can enhance the interface amongst man and machine in various applications [5].

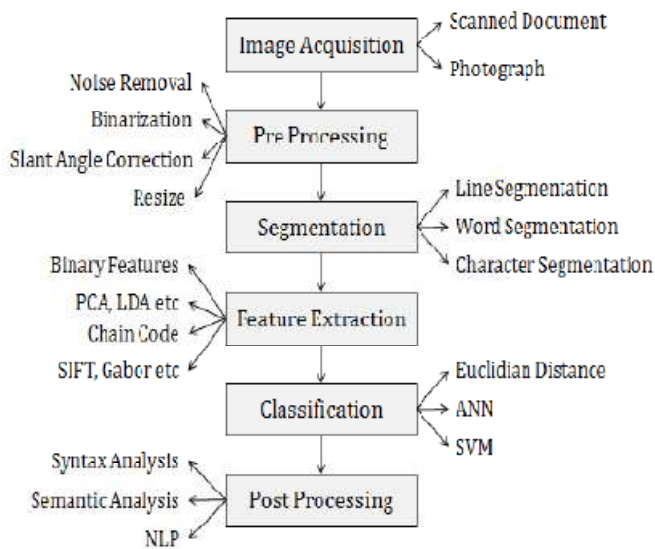


Figure 3: Pattern Recognition Model

III. HANDWRITTEN CHARACTER RECOGNITION METHODOLOGIES

A handwritten character recognition methodology includes feature extraction techniques and classification techniques.

3.1 Feature extraction

In this stage of character recognition, the essential characteristics of the symbols are captured [6]. It is most essential stage. The feature extraction is essential stage along

with the selection of features from the extracted features is more important. The proper features selection will reduce the computations hence the recognition time will reduced. This improves the system performance [7]. According to C.Y. Suen [8], features are obtained from the arrangement of points constituting the character matrix. These features are easily detected as compared to topological features. It provides low complexity and high speed [9]. The number of techniques are used in feature extraction, some of them are moment, zoning, projection histogram, n-topples, crossing and distance.

3.1.1 Statistical features

Statistical features are derived from statistical distribution of pixels and its characteristics measurement of pattern. Features are the measurements which represent the characters such as size, shape and intensity. The lists of the notable statistical features are: Area, Centroid, BoundingBox, MajorAxisLength, MinorAxisLength, Eccentricity Orientation, ConvexHull, ConvexImage, ConvexArea, FilledImage, FilledArea, UlerNumber, Extrema, EquivDimater, Solidity and Extent.

3.1.2 Methods used to extract statistical features

Statistical features are some measures of inherent similarity of distance the following methods are prominently used for the extraction of such features.

i. Moment: - In this case the moments of different points present in a character are utilized as a feature, S. S. Reddi [10]. Redial and angular moment where as generic moments where proposed by Teh and Chin [11].

ii. Zoning: - The frame of character is divided into several overlapping and no overlapping zones and the densities of object pixels in each zone are calculated. And the density is calculated as a no of object pixel in each zone divided by total no of pixels [12].

iii. Projection histogram: - projection histogram gives a no of black pixels in the vertical and horizontal directions of the specified character. It may be vertical, horizontal and left diagonal or right diagonal [13].

iv. N-topple: - In this type the position of black and white pixels is considered as feature. This method provides the no of important properties of the pixels.

v. Crossing and distances: - In this type, features are obtained from counts the character image is crossed by vector in certain directions or at certain angles.

3.2 Structural Features

Structural features are the description of topological and geometrical characteristics of the image of character. Some of the notable and commonly used structural features for character recognition are: Horizontal Line, Vertical Line, End Points, Number of crossing points, Number of curves at top and bottom, Curves opening directions, Number of Contours.

3.2.1 Methods used to extract structural features

The commonly used structural feature extraction methods are given below

3.2.1.1 Geometrical and topological features: - This type of represented takes place an encode some knowledge about the stricter of the object or sorting of components make up that object. Various topological and geometrical representations can be grouped into four categories. a. Extraction and counting topological structure:- In this category, lines, curve, spleens, extreme points, maxima and minima, cups above and below the threshold opening, and cross points, end lines, loops, direction of a stroke from a special point, inflection between two points, horizontal curve at top or bottom, ascending, descending, and middle stroke and relation among the stroke that make up a character as consider as a feature. [13]

b. Measuring and approximating the geometrical Properties:- In this type the characters are represented by the measurement of geometrical quantities such as ratio between height and width of the bounding box of character, the relative distance between last point and last y-min, distance between two points, upper and lower masses of words and word length curvature or change in curvature. [13]

c. Coding: - one of the most popular coding is Freeman's chain code. This code is determined by mapping strokes of character into a two dimensional.

d. Graphs and trees: - words or characters are first portioned into a set of topological primitives such as strokes, holes, cross points etc. These primitives are represented using attributed or relational graphs. Trees can also be represented using the words or characters with a set of features, which has a hierarchical relation.

3.3 Neural Network Approach

Neural Network approach used better recognition of handwritten character recognition as compared to statistical and structural approach. Because of its distinctive inherent

property of reconciling learning. Neural Network has an ability to automatically take into account the peculiarities of different handwriting styles. Some of the neural network pattern recognition is are feed Forward Network, Back propagation network, Hopfield Model, Self –organizing Network .

3.4 Classification

Classification is the decision making stage of handwritten character recognition with homogeneous characteristics. It can carry out on the basis of stored features of the features space, i.e. structural and global features etc. [14]

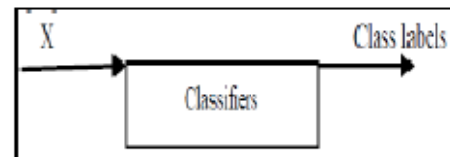


Figure 4: Classifier Technique

Where X is the feature vector.

Some classification methods are temple matching, Statistical technique, Structural techniques, and artificial neural network [15].

- Template matching: This is the simplest form pattern recognition. The given pattern that is to be recognized is compared with the stored pattern. Style and size are ignored while matching the pattern.
- Statistical method: In this task, can determined the category of given pattern belongs. By making observations and measurement process, a set of numbers is prepared, which is to prepare a measurement vector. [16]
- Structural method: It is good for determine handwritten text. In this type, classifies the input pattern on the basis of components of the relationships among these components. The structural approach is that it provides a good symbolic description of the image. [16]
- Artificial neural network: - It component of inter connected element called neurons. This algorithm is non algorithmic and trainable. The neural network is for pattern classification class is the feed forward network and other used for classification purpose or conventional neural network.
- Kernel method: Most important kernel methods are support vector machine, kernel principle component analysis etc. support vector machine are a group of

supervised learning methods that can be applied into training and testing test. Different types of kernel functions of SVM are linear kernel, polynomial kernel, Gaussian radial basis function and sigmoid.

Classifiers used for character recognition problem include k-nearest neighbor classifier, hidden Markov model (HMM), support vector machine (SVM), and artificial neural network (ANN) etc. Jain et al. [17] give a review of statistical pattern recognition techniques. In [18], Pal and Singh train neural network to recognize uppercase handwritten characters based on Fourier descriptors of character boundaries as features. In [19], recognition of handwritten alphabets using neural network and zoning based diagonal features is addressed.

IV. COMPARATIVE STUDY OF DEVANAGARI CHARACTER RECOGNITION

As compare to other script the Handwritten Devanagari character recognition is much more complex task and it is due to the sum of the silent feature of the Devanagari script

- The script is written from left to write and after the completion of each word a space is used as a separator.
- One may write devanagari with or without "Shirorekha" ,that is, horizontal line placed on the top of the character.
- Vowels can occur independently or in conjunction with diacritics.
- It is not a case sensitive script.
- A compound character is a combination of two or more characters and/or vowels. It user dependent.[Complexity of devanagari script]

Analysis of different handwritten character methods: This analytical study results in comparison of feature extraction methods and different classifiers used for handwritten Devanagari character recognition system.

Devanagari script is employed for writing many languages some of them are Hindi, Marathi, Nepali and Sanskrit. A Aggarwal *et. al* are recommended several techniques for recognition of Devanagari handwritten Characters because of its logical simplicity, easy use and high recognition rate, Gradient features and obtained recognition accuracy of 94%. P E Ajmire *et. al* have proposed histogram oriented gradient which is employed to find the article in 3 dimensional atmospheres (3-D) that can be used for 2-D image of character for handwritten Devanagari vowels recognition. The classifier

SVM is employed for the classification with the accuracy of recognition of handwritten Devanagari vowels is 92.11%. P M. Kamblea *et. al* presents a system for handwritten Marathi character recognition. The system has been evaluated on big amount of hand written Marathi characters. It shows the conclusion of R-HOG primarily based feature extraction methodology and FFANN based classification is simpler with accrued process speed and accuracy 97.15%. K. V. Kale *et. al* that extracts Zernike moment features and classified using SVM and k-NN approach. In future the system is extended to handle additional compound characters with different options like orthogonal moments and might classified through advanced patterns classification and neural network approaches. The recognition rate is with SVM is 98.33% & KNN is 95.82% . P E Ajmire *et. al*, Handwritten Devanagari (Marathi) Compound Character Recognition using Seventh Central Moment proposed system with different kinds of compound characters, the average accuracy of recognition rate is 93.87%. Arjun Singh *et. al* obtained the results as SVM gives reliable classification accuracy. It uses four experiments with different feature extraction techniques, Chain Code, ZoneBased Centroid, Distance Profile, and Background Directional Distribution. These features are applied to SVM classifier. It results the recognition rate with accuracy 97.61%. Sandhya Arora *et. al*, proposes an off-line handwritten Devanagari character recognition system consisting of three different features using simple feed forward Multilayer Perceptrons. The character is said to be classified by any one of the classifier. It was giving 92.16% accuracy. N. Sharma *et. al* presented a quadratic classifier based system for the recognition of unconstrained off-line Devanagari handwritten characters are obtained based on the directional chain codes of the contour of the character. It has a better than results reported 98.86%. Hanmandlu *et. al.* have used membership functions of fuzzy sets for handwritten Devanagari script recognition. M Jangid proposed the feature vector is form from the three feature extraction techniques for a handwritten Devanagari character. The two different people who belongs to different age group have been prepared the handwritten Devanagari Character, and obtained the 94.89 % recognition accuracy.

SNo.	Author	Feature extraction method	Classifier	Size	Recognition Rate (%)
1	A Aggarwal <i>et al</i> [20]	Gradient Features	SVM	7000	94%
2	P E Ajmire <i>et. al</i> [21]	Histogram Gradient	SVM	5400	92.11%
3	P. Kamble <i>et. al</i> [22]	Rectangle Histogram Oriented Gradient	FFANN	4000	97.15%
4	K. V. Kale <i>et. al</i> [23]	Generic Moment	SVM KNN	27,000	98.33% & 95.82%
5	P E Ajmire <i>et. al</i> [24]	7th Order Central Moment	SVM	6000	93.87%
6	Arya Singh <i>et. al</i> [25]	Chain Code, Zone Based Centroid	SVM ANN	20,000	97.61%
7	Sandhya Arora <i>et. al</i> [26]	Shadow Feature	MLP	4900	92.16%
8	N. Sharma <i>et. al</i> [27]	Chain Code Features	Quadratic Classifier	---	98.86%
9	Hanmandlu <i>et al</i> (Numerals) <i>et. al</i> [28]	Normalized distance	Fuzzy	4750	95%
10	M Jangid [29]	Statistical features	SVM	12240	94.89%

IV. CONCLUSION

In this paper, we have surveyed the character recognition work that has been done on Devanagari scripts also analyzed the recognition rate using different features using various classifier techniques. We have assessed the work done for handwritten characters. We have presented recognition accuracies achieved for character recognition of different scripts. We have seen that the efficient techniques used handwritten character so that accuracy of recognition may be increased. The novel features can be proposed to improve the recognition accuracy of different scripts. There is a need to develop the standard database for Devanagari, Gurmukhi scripts etc.

REFERENCES

- Xiaolin, Li and Dit- Yan Yeung, "On-line handwritten alphanumeric character recognition using dominant stroke", Pattern recognition, Journal of pattern recognition, Vol.30, ISSUE 1(Published by Elsevier Jan 1997)
- P.E. Ajmire, R.V. Dharaskar and V.M. Thakare, "Handwritten Devanagari (Marathi) Compound Character recognition using Seventh Central Movement", International Journal of Innovative Research in Computer and Communication Engineering, Vol.3, Issue 6, June 2015.
- Mansi Shah And Gordhan B Jethava, "A Literature Review On Hand Written Character Recognition" Indian Streams Research Journal Vol -3 , ISSUE -2, March.2013
- Pal, U., and B.B. Chaudhuri, 2004. Indian Script Character Recognition: a Survey, Pattern Recognition, 37: 1887-1899.
- Mantas, J., 1986. An overview of character recognition methodologies, Pattern recognition, 19(6): 425-430.
- Er. Harpreet Kaur, "A Survey of Feature Extraction and Classification Techniques Used In Character Recognition for Indian Scripts," International Journal of Engineering Sciences, Vol. 3, Issue December 014.

- Muhammad Aarif Mohamed , Dewi Nasien, Haswadi Hassan and Habibollah Haron, "A Review on Feature extraction and Feature Selection for Handwritten Character Recognition", International Journal of Computer Science and Applications, Vol. 6, 2015.
- C.Y. Suen, "Character recognition by computer and applications", in Handbook of pattern recognition and Image processing, New York: Academic, pp. 569-586, 1986.
- Priyanka Varma and Pratibha Singh, "A survey on handwritten character recognition using Artificial Neural Network," International Journal of Engineering applied sciences and Technology Vol. 1, Issue 5, 2016.
- S.S. Reddi, "Radial and angular moment invariants for image identification", IEEE Transactions on PAMI, Vol. 3(2), pp.240-242, 1981.
- C. H. Teh and R.T. Chin, "On image analysis by the method of moments", IEEE Transactions on PAMI, Vol. 10(4), pp.496-513, 1988.
- Pritpal Singh and Sumit Budhiraja, "Feature Extraction and Classification Techniques in O.C.R. System for Handwritten Gurumukhi Script- A Survey," International Journal of Engineering Research and Applications, Vol. 1, Issue 4, pp. 1736-1739, August 2012.
- C.Y. Suen, "Character recognition by computer and applications", in Handbook of Pattern Recognition and Image Processing, New York: Academic, pp. 569-586, 1986.
- Ratnashil N Khobragade and Dr. Nitin A koli, "A survey on recognition of Devnagari Script," International Journal of Computer applications & Information Technology Vol. II, Issue I, January 2013.
- P. E. Ajmire, R. V. Dharaskar, V. M. Thakare, "Statistical Techniques for Feature Extraction for Handwritten Character Recognition"- A Survey, January 2014.
- P. E. Ajmire, R. V. Dharaskar, V. M. Thakare and S E Warkhede, "Structural Features for Character Recognition System-A Review", International Journal of Advanced Research in Computer Science, Volume 3, No. 3, May-June 2012.
- A. K. Jain, P. W. Duin, and J. Mao, "Statistical pattern recognition: a review," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 22, no. 1, Jan. 2000.
- A. Pal and D. Singh, "Handwritten English character recognition using neural network," International Journal of Computer Science and Communication, vol. 1, no. 2, pp. 141-144, July-Dec 2010.
- J. Pradeep, E. Srinivasan 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, Feb. 2011.
- A Aggarwal, R Rani, RenuDhir "Handwritten Devanagari Character Recognition Using Gradient Features", International Journal of Advanced Research in Computer Science and Software Engineering, Vol.2, Issue 5, May 2012
- P E Ajmire, R V Dharaskar, V M Thakare "Feature Extraction And Selection For Handwritten Devanagari Vowels

- Recognition", International Journal of Research in Engineering and Technology, Vol. 05, May-2016
22. P. M. Kamble, R. S. Hegadi "Handwritten Marathi character recognition using R-HOG Feature, International Conference on Advanced Computing Technologies and Applications (ICACTA-2015)
23. K. V. Kale, P. D. Deshmukh, S. V. Chavan "Zernike Moment Feature Extraction for Handwritten Devanagari (Marathi) Compound Character Recognition", International Journal of Advanced Research in Artificial Intelligence, Vol. 3, No.1, 2014
24. 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 (An ISO 3297: 2007 Certified Organization) Vol. 3, Issue 6, June 2015
25. Arjun Singh, Kansham Angphun Maring "Handwritten Devanagari Character Recognition using SVM and ANN", International Journal of Advanced Research in Computer and Communication Engineering Vol. 4, Issue 8, August 2015
26. S. Arora, D. Bhattacharjee, M. Nasipuri, D. K. Basu, and M. Kundu, "Recognition of non-compound handwritten Devanagari characters using a combination of MLP and minimum edit distance," International Journal of Computer Science and Security, vol. 4, no. 1, pp. 1-14, 2010.
27. N. Sharma, U. Pal, F. Kimura, and S. Pal, "Recognition of offline hand-written Devanagari characters using quadratic classifier," in Proc. Indian Conference on Computer, Vision and Graphics, Image Processing 2006, pp. 805-816.
28. M. Hanmandlu ; O.V. Ramana Murthy ; Vamsi Krishna Madasu "Fuzzy Model Based Recognition of Handwritten Hindi Characters", IEEE, 07 January 2008
29. Mahesh Jangid "Devanagari Isolated Character Recognition by using Statistical features", International Journal on Computer Science and Engineering (IJCSSE), ISSN : 0975-3397 , Vol. 3 No. 6 June 2011