

# Online Handwritten Devnagari Word Recognition using HMM based Technique

Prachi Patil  
Master of Engineering  
Dept. of Electronics & Telecommunication  
Dr. D. Y. Patil SOE, Pune, India

Saniya Ansari  
Professor  
Dept. of Electronics & Telecommunication  
Dr. D. Y. Patil SOE, Pune, India

## ABSTRACT

In this paper, online handwritten Devnagari word recognition system is proposed and discussed. The increase in usage of handheld devices which accept handwritten data as input created a demand for application which analyze and recognize data efficiently. Due to the popularity of digital device, we use Smartphone as input device. Input image is drawn on Smartphone. Feature extraction of input image is done by android technology. Using that features HMM recognizes the word. Experimental results show advantages of this method in the field of handwriting recognition.

## Keywords

Android, Devnagari, Feature Extraction, HMM, Recognition.

## 1. INTRODUCTION

Devnagari is one of the most popular scripts in India. Devnagari script is use to write Hindi, Kokani, Marathi, Nepali, Sanskrit, Dogri and Sindhi. It is also used in Urdu. Devnagari script plays an important role in the development of manuscript and literature.

Handwriting recognition has been one of the fascinating and challenging fields for researcher in image processing and pattern recognition in the recent years. Handwriting recognition is mainly divided into two type's offline and online. In offline recognition the writing is usually captured by a scanner so whole image is present thus only static images are available. While in online recognition system input is given by some digital devices such as Digital pad, Tablet PC etc.

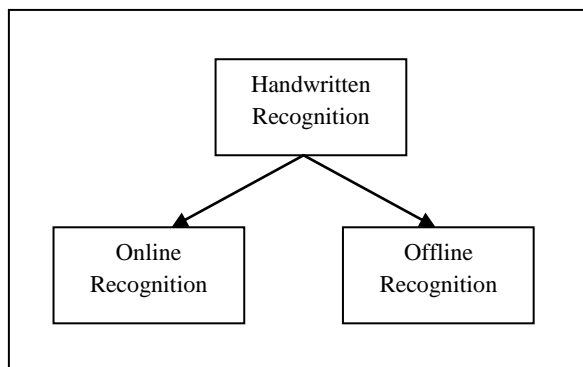


Fig 1: Types of Handwritten Recognition

Devnagari has one distinct feature than other language that is shirorekha. Shirorekha is the upper core of the word and it does not contain any usefull information. Thus in

Devnagari recognition first detect the shirorekha and discard it. Devnagari script is written from left to right. Devnagari consists of vowels and consonants. The complexity of an online handwritten word recognition increases mainly because of various writing styles of writing styles of different writers.

अ	आ	इ	ई	उ	ऊ	ए	ऐ	ओ	औ
क	ख	ग	घ	(ङ)					
च	छ	ज	झ	(ञ)					
ट	ठ	ड	ढ	ण	इ	ड़			
त	थ	द	ध	न					
प	फ	ब	भ	म					
य	र	ल	व						
श	ष	स	ह						
क्ष	त्र	ज्ञ							

Fig 2: Vowels and Consonants in Devnagari Script

For online handwritten recognition system electronic devices are widely used because of their ease. In this paper Smartphone is used as input device for recognition system.

## 2. LITERATURE SURVEY

The literature survey carried out related to technology impact in the study of different script recognition techniques used on different handwritten languages. HMM is widely used in handwritten recognition. HMM based lexicon technique is use for recognition of handwritten Devnagari and Tamil script [1]. In the lexicon-driven technique models each word in the lexicon as a sequence of symbol HMMs according to a standard symbol writing order derived from the phonetic representation. The lexicon-free technique uses a novel Bag-of-Symbols representation of the handwritten word that is independent of symbol order and allows rapid pruning of the lexicon.

Neural network perform computation at a higher rate compared to classical technique. In [2] and [7] Neural network is used for Devnagari script recognition. Anoop Namboodiri [5] presented a method for online recognition of handwritten text by a K nearest neighbor and support vector machine classifier and sequential floating search method for feature extraction. It classified words and lines in an online handwritten document into one of the six major scripts: Arabic, Cyrillic, Devanagari, Han, Hebrew, or Roman.

B. V. Dhandra [6] presented an automatic technique for script recognition at word level based on morphological

reconstruction is proposed for two printed documents of Kannada and Devnagari containing English numerals. The technique developed includes a feature extractor and classifier.

A simple and an efficient offline handwritten character recognition system using a new type of feature extraction namely, radon feature extraction is proposed by M.K. Mohahmad with recognition accuracy of 90% for 270 features [9]. S. Shelke and S. Apte proposed scheme for handwritten Devnagari character recognition which combines neural network and template matching recognition approach [10]. Ved Agnihotri [12] proposed a new technique of Chromosomes function generation and fitness function for classification by extracting diagonal features from zones of an image. Handwritten Devanagari script recognition system using neural network is presented in this paper. Diagonal based feature extraction is used for extracting features of the handwritten Devanagari script. After that these feature of each character image is converted into chromosome bit string of length 378. In the recognition phase and classification Genetic Algorithm is used.

### 3. PROPOSED METHODOLOGY

#### 3.1 Data Collection

In the proposed method Android based Smartphone is used as an input device. For the experiment Android 4.1 jelly bean is used. Gesture class is used to capture gesture which user will draw screen.

#### 3.2 Feature Extraction

Gesture stroke started on a touchdown and ended on a touch up. A stroke consists of a sequence of timed points. One or multiple strokes form a gesture. Using Gesture stroke class x-y coordinates of gesture on screen is extracted. Extracted features of input data is shown in Figure 3.

**Algorithm 1:** For Data Collection and Feature Extraction

Input: Image drawn on Smartphone  
Output: Extract features for Recognition

Load the library required for gesture. Create WiFi connectivity. For input gesture layout should be set. Once the layout is decided Check whether gesture is drawn or not. Once the gesture is drawn find bounding box stroke. In bounding box four features are extracted such as top, bottom, right and left. Also all the points covered by gesture on the screen.

**Algorithm 2:** Data Flow from Android to Matlab

Input: Extracted features From Android  
Output: Store in file which is readable to Matlab

Http connection is created. For WiFi connection android application is connected to router. Set the file path where features will be store. Once path is created Http buffer read extracted features from algorithm 1 until it becomes null. PHP will store extracted features to file.

#### 3.3 Recognition using HMM

Using PHP features extracted by android is passing to HMM with the help of wifi. The idea of HMM based word recognition systems is to build word models for all hypotheses. Recognition is done by concatenating character models which are built during a training phase. For each word a rating value is calculated, which can be

interpreted as similarity or probability. Viterbi algorithm calculates the likelihood for the best matching character sequence. Also EM algorithm used for matching character. In this each iterations is guaranteed to increase the likelihood of the data. Using this algorithm word recognition is possible. Using HMM model word can be recognized. Figure 4 shows the final output of the recognition system.

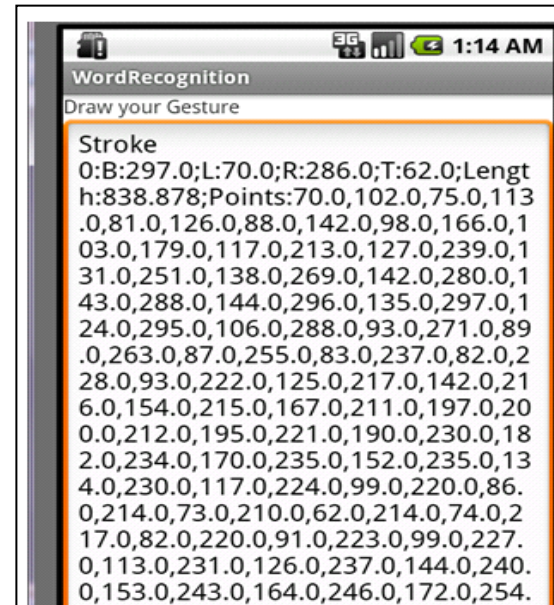


Fig 3: Simulation result for feature Extraction

**Algorithm 3:** Data Flow from Android to Matlab

Input: Features store in file  
Output: Word Recognition by HMM

Create connection to WAMP to read file generated by PHP. Check whether data is read from PHP is compatible to Matlab. Set the maximum iteration to find likelihood data. HMM is used for recognition by using EM algorithm. EM algorithm finds the expected value and maximization probability of likelihood data. Read the likelihood data image. Normalize the image and display it.

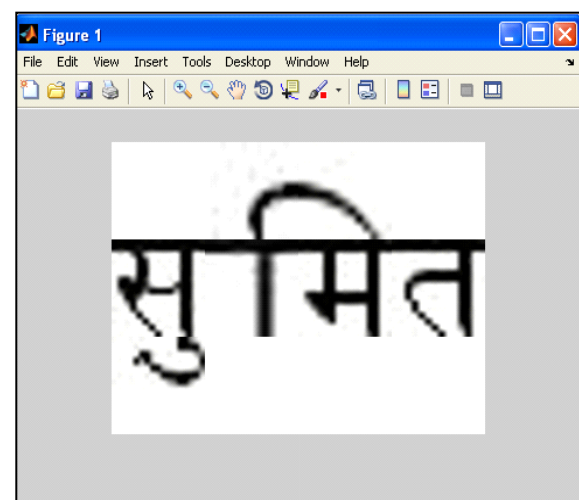
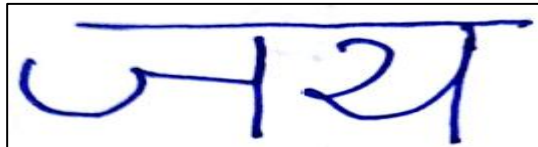


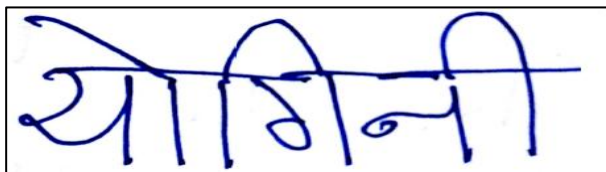
Fig 4: Simulation of final output

#### 4. EXPERIMENTAL RESULT AND DISCUSSION

In Devnagari script vowels, consonant, lower and upper modifier present. With the help of this modifier script is written. But sometimes this modifier becomes difficult to recognize. For experimental analysis we divided word into two types. Type1 word does not contain any modifier while Type2 word consists of lower and upper modifier.



Type 1 word



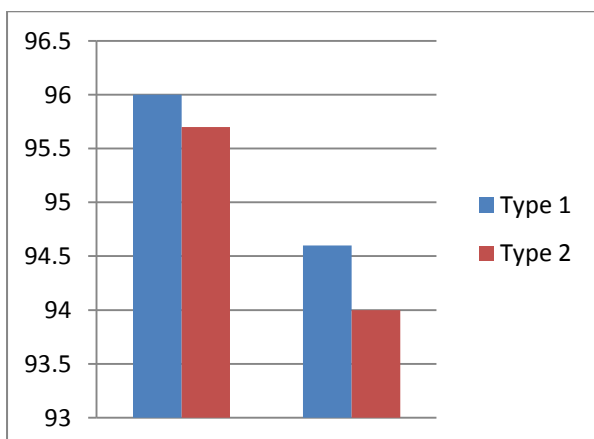
Type 2 Word

Fig 5: Different type of word

Following table summarize the results obtained by utilizing the presented system to recognize the collected handwritten word using the collected character template database. Following table shows the first column of entries to measures the accuracy of the system, by specifying the percentage of words written that were recognized correctly. Subsequent columns denote the accuracy within the top 5 and 10 candidates from the lexicon, respectively.

Table 1 : Average Accuracy of different type of word

Word	Accuracy/top5	Accuracy/top10
Type 1	96.00	94.6
Type 2	95.7	94.00



We have conducted test to know the recognition rate of our application. For this we test our application to 50 and 100 words by different writers. Result of this test is summarized in following table.

Table 2: Average Accuracy of different set of word

Word	Correctly recognized	Misrecognized	%Recognition
50	48	2	96.00%
100	94	6	94.00%

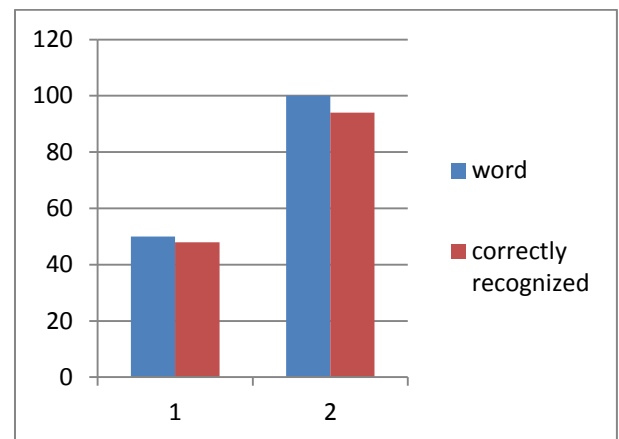


Fig. 7 Graph of accuracy of different set of word

Word is consists of characters and modifiers. So recognition accuracy depends on characters. Following graph shows the accuracy of the word up to 3 characters and up to 5 characters. Accuracy of the word up to 3 characters is higher than 5 characters.

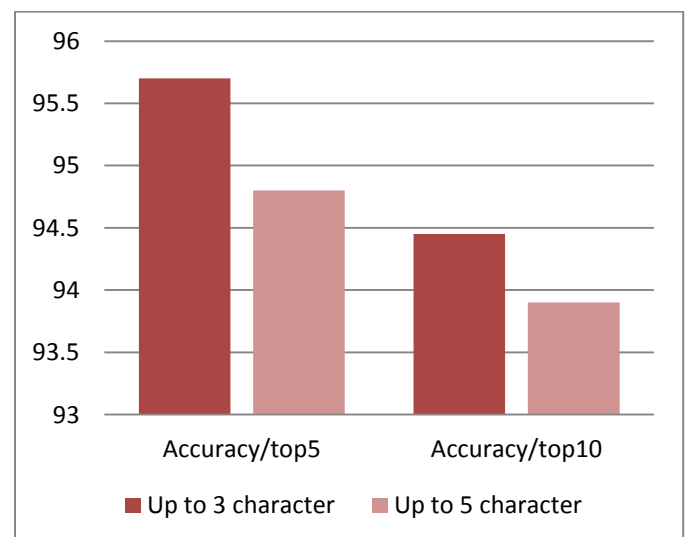
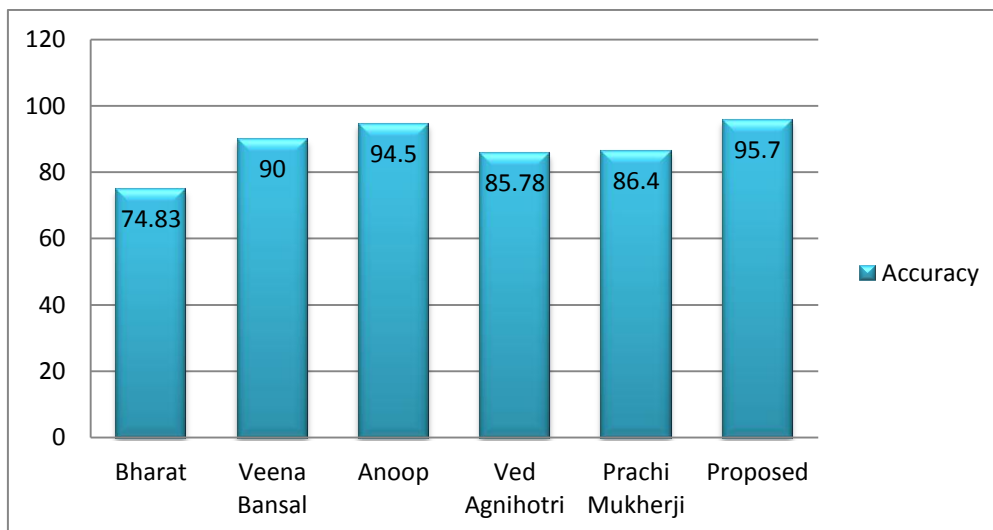


Fig. 7.3 Graph of accuracy of different word

In table 1 shows the comparison of how the proposed method stacks up against other methods, which we reviewed as part of the literature survey

**Table 3: Comparison of proposed method with existing method**

Author	Year	Feature Extraction Method	Recognition Method	Accuracy
Bharat et.al [1]	2012	Npen recognition	Hidden Markov Models (HMM): lexicon Driven and lexicon free.	74.83%
Veena Bansal et.al[3]	2008	Vertical feature bar, horizontal zeroes, crossing moments.	Tree classifier	90%
Anoop Namboodiri [5]	2004	sequential floating search method	K nearest neighbor and support vector machine classifier.	94.5 %
Ved Agnihotri[12]	2012	Diagonal features	Genetic Algorithm	85.78%
Prachi Mukherji et.al[20]	2009	Structural features like endpoint, cross point, junction points	Tree classifier	86.4%.
Proposed	2013	Android based	Hidden Markov Models (HMM):	95.70%



**Fig. 6 Graph of accuracy comparison of proposed with existing method**

## 5. CONCLUSION & FUTURE SCOPE

In this paper a simple online handwritten Devnagari word recognition system using HMM is proposed. The main objective of this paper is to use android technology in Devnagari feature extraction. Using these features HMM can recognize the word. The success of any recognition system is depends on feature extraction and classifier which is used to classify the unknown input to well define class. Using proposed method accuracy of recognition is enhancing.

This application can be used for conjugate word. Also it can be used for script recognition.

## 6. REFERENCES

- [1] Bharat, Sriganesh Madhvanath, "HMM-Based Lexicon-Driven and Lexicon-Free Word Recognition for Online Handwritten Indic Scripts", *IEEE*, April 2012.
- [2] K. Y. Rajput and Sangeeta Mishra "Recognition and Editing of Devnagari Handwriting Using Neural Network", *Proceedings of SPIT-IEEE Colloquium and International Conference*, Mumbai, India Vol. 1, 66.
- [3] Sandhya Arora, Debotosh Bhattacharjee, Mita Nasipuri, Latesh Malik, "A Two Stage Classification Approach for

- Handwritten Devanagari Characters”, *Proceedings of the Fifth International Conference on Document Analysis and Recognition*, 1999, pp.653-656.
- [4] Anoop Namboodiri, “Online Handwritten Script Recognition”, *IEEE* Vol. 26 No. 1, January 2004.
- [5] B. V. Dhandra et. al, “Word-wise script identification from Bilingual Documents Based on Morphological Reconstruction”, *IEEE* Vol. 32, No. 12, 2006.
- [6] Naveen Sankaram, C. V. Jawahar, “Recognition of Printed Devnagari Text Using BLSTM Neural Network”, *ICPR*, Nov. 11-15, 2012.
- [7] U. Bhattacharya, A. Nigam, “An Analytic Scheme for Online Handwritten Bangla Cursive Word Recognition”.
- [8] M.K. Mohahmed Althaf, M. Baritha Begum, “Handwritten Characters Pattern Recognition using Neural Networks”, *International Conference on Computing and Control Engineering*, April 2012.
- [9] S. Shelke, S. Apte, “A Novel Multiclassifier Scheme for Unconstrained Handwritten Devnagari Character Recognition”, *Proc. 12<sup>th</sup> Int. Conf. Frontiers in Handwriting Recognition, Kolkata, India*, pp. 215-219, Nov. 16-18, 2010.
- [10] Shanthi N and Duraiswami K, “A Novel SVM -based Handwritten Tamil character recognition system”, *Springer Pattern Analysis & Applications*, Vol-13, No. 2, 173-180, 2010.
- [11] Ved Agnihotri, “Offline Handwritten Devnagari Script Recognition”, *IJCSI International Journal of Computer Science Issues*, 2012.
- [12] Suresh Kumar C and Ravichandran T, “Handwritten Tamil Character Recognition using RCS algorithms”, *Int. J. of Computer Applications*, (0975 – 8887) Volume 8– No.8, October 2010.
- [13] U. Garain, B.B. Chaudhuri “Segmentation of Touching Characters in Printed Devnagari and Bangla Scripts Using Fuzzy Multifactorial Analysis”, *Proceedings of the 6th International Conference on Document Analysis and Recognition*.
- [14] M.C. Padma, P.A. Vijaya, “Identification of Telugu, Devnagari and English Scripts using Discriminating features”, *IJCSIT*, Vol 1, No 2, November 2009.
- [15] C. V. Jawahar, M. N. S. S. K. Pavan Kumar, S. S. Ravi Kiran “A Bilingual OCR for Hindi-Telugu Documents and its Applications”, *Proc. of the 11th ICPR*, vol. II, pp. 200-203, 1992.
- [16] M.K. Mohahmed Althaf, M. Baritha Begum, “Handwritten Characters Pattern Recognition using Neural Networks”, *International Conference on Computing and Control Engineering*, April 2012.
- [17] S. Shelke, S. Apte, “A Novel Multiclassifier Scheme for Unconstrained Handwritten Devnagari Character Recognition”, *Proc. 12<sup>th</sup> Int. Conf. Frontiers in Handwriting Recognition, Kolkata, India*, pp. 215-219, Nov. 16-18, 2010.
- [18] B. V. Dhandra et. al, “Word-wise script identification from Bilingual Documents Based on Morphological Reconstruction”, *IEEE* Vol. 32, No. 12, 2006.
- [19] Prachi Mukherji, Priti P. Rege, —Shape Feature and Fuzzy Logic Based Offline Devnagari Handwritten Optical Character Recognition, *Journal of Pattern Recognition Research* 4 (2009), pp 52-68.