International Journal on Future Revolution in Computer Science & Communication Engineering
Volume: 3 Issue: 7

ISSN: 2454-4248
18 – 22

Opportunities and Challenges of Handwritten Sanskrit Character Recognition System

Shailendra Kumar Singh
Research Scholar, CSE Department
SLIET Longowal, Sangrur, Punjab, India
Sks.it2012@gmail.com

Manoj Kumar Sachan
Assosiate Professor, CSE Department
SLIET Longowal, Sangrur, Punjab, India
manojsachan@gmail.com

Abstract: The rapid growth in the field of internet facilities and digitalization, changes the living way of human being. Due to internet facilities and services, anyone can access data from anywhere. A lot of online data are generating day by day, so that data needs to be processed before extracting the information. Therefore the demand of Natural language Processing (NLP) Techniques has been increased. The Pattern recognition is sub-field of NLP. The field of Pattern Recognition is a branch of machine learning that contributed up to great extent in the Computer Vision and Machine Vision applications. Pattern Recognition is concerned with the recognition of patterns and regularities in data. Handwriting recognition is one of the challenging subtask and current research field under Pattern Recognition, due to different ways of writing and handwriting styles. Handwritten Sanskrit Characters recognition is more complicated than other languages works in online and offline mode, because Sanskrit characters have more consonants and modifiers. In this paper discussed the opportunities and challenges of Handwritten Sanskrit Character Recognition System.

Keywords: *Pattern Recognition, Devanagari, Sanskrit, Sanskrit Character, Character Recognition System, Handwriting.*

*****

I. INTRODUCTION

With the development of new technologies, demand of digital content increases day by day. There are a lot of handwritten contents are stored in libraries, but they are not accessible from outside the libraries. So, there is need of content digitalization for converting handwritten contents into editable and digital form to avail the content online. To generate the digital content, need for the development of an online or offline Handwriting Recognition System with high performance has become essential. Handwriting is a personal skill of every individual. Each person has their own unique style of handwriting; it may be everyday handwriting or their personal signature. Now time is to use the digital signature. The characteristics of handwriting are [17]: slop, specific shape, thickness, average size of letters, pressure to the paper and regular or irregular spacing between letters. These characteristic are used to recognize the handwritten characters. The Handwriting Recognition System (HRS) is used to take handwritten document (texts) as input and produce editable document (digital text) as an output as shown in fig. 1. There are two types of HRS i.e offline and online HRS as discussed in section II.

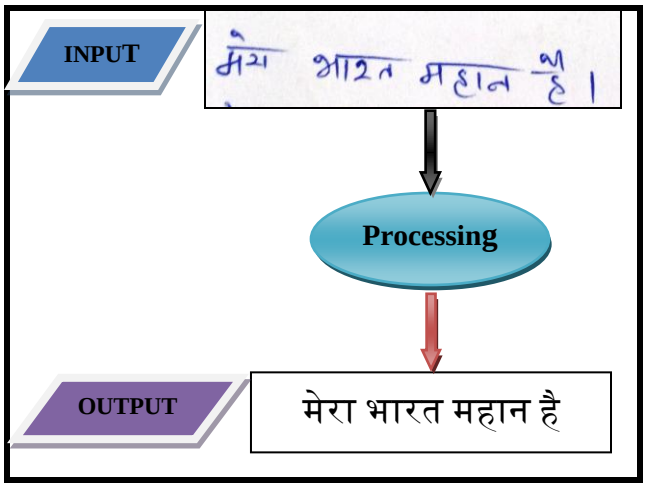


Figure:1 Handwriting Recognition System

Devanagari is an ancient Indian script that is used to write languages such as Sanskrit, Hindi, Marathi and several others. Sanskrit language is written in Devanagari Script. Sanskrit is an ancient Indo Aryan language. In India the state of Uttarakhand has ruled Sanskrit as its second official language [16]. It is also one of official language in India. Most of the Indian Languages have influenced of vocabulary and grammatical base of Sanskrit language. There is a horizontal line at the top of each character or word in Sanskrit language as shown in fig.2. This horizontal line is called as Shirorekha [1]. Sanskrit consists of 48 characters (15 vowels and 33 consonants) as shown in fig.3.

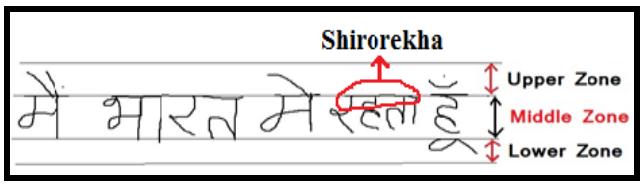
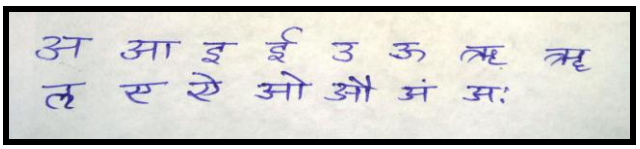
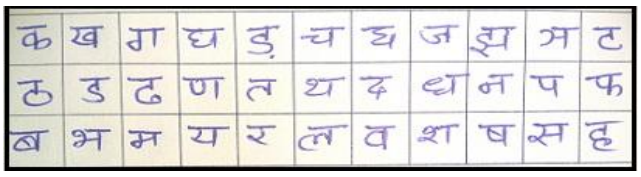


Figure:2 Shows Shirorekha and Zones



(a)Vowels



(b) Consonants

Figure:3 Sanskrit alphabet in modern Devanagari Script;

(a) Vowels and (b) Consonants

The NLP techniques are used in HRS. There are a large number of works done using the NLP techniques such as Sentiment analysis [2, 3], Script recognition [4, 5], information extraction etc. The main aim of Handwriting recognition is to

 ISSN: 2454-4248
18 – 22

get handwritten document and convert it into editable text. There are two form of handwriting recognition system. One is Online, in which handwritten document is captured by writing on device as shown in Fig.4. Second is offline handwriting recognition system, in which input is in form of image captured by scanner or camera after writing on a piece of paper as shown in Fig. 5. We discussed the challenges of Sanskrit character recognition.

II. TYPES OF HANDWRITING RECOGNITION SYSTEM

Online handwriting recognition system means that the electronic device (machine) recognizes the writing while the user writes on an electronic device. Depending on the speed of the device (computer) and the recognition technique, the recognition lags behind the writing to a greater or lesser extent. On-line recognition systems need only to be fast enough to keep up with the writing. Average writing rates are 1.5-2.5 characters/s for Sanskrit.

(a) (b) (c)

Figure:4 Devices to Capture Online handwriting; (a) Digimemo; (b) Tablet PC; (c) PDA

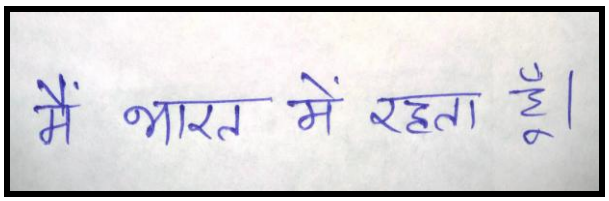


Figure:5 Image of handwritten document captured by Camera.

On-line handwriting recognition system requires a transducer that captures the writing as it is written. The most common of these devices is the electronic tablet or digitizer, which typically has a resolution of 200 points/in, a sampling rate of 100 points/s, and an indication of "inking" or *pen down.* The framework of online is discussed in [3,15].

Off-line handwriting recognition, is performed after the writing is completed. It can be performed after years, months or days. In offline character recognition system a piece of paper is used to write the character and scan directly into the system by an optical scanner or camera. The optical scanner converts the image of writing into a bit pattern. Scanners have x and y resolutions of typically 300-400 points/in. The process of offline HRS for Gurmukhi and Devanagari scripts is discussed in [5, 6].

Differences in capturing of handwriting data between online and offline recognition system. Online data capture means that the machine data are being captured as a person writes. While Off-line data capture means that the machine data are captured some time after the writing is created. Once the handwriting data is captured then processed by the recognizer. Although on-line data are recognized immediately in most applications, while offline data can be recognized later.

An advantage of on-line devices is that they capture the temporal or dynamic information of the writing. This information consists of the number of strokes, the *order* of the strokes, the *speed* of the writing within each stroke, and the *direction* of the writing for each stroke. A *stroke* is the writing from pen down to pen up. Most on-line transducers capture the trace of the handwriting or line drawing as a sequence of coordinate points [7]. While offline conversion of scanned data to line drawings usually costly and imperfect preprocessing to extract contours and to thin or skeletonize them [8].

The main disadvantage of on-line handwriting recognition system is that the writer is required to use special device to write. Unfortunately, current on-line device is not as comfortable and natural to use as pen and paper [9].

III. RELATED WORK

N. Dwivedi et. al., 2013 [10] and S.P. Patil et. al., 2014[11], presented the handwritten Sanskrit word recognition system using a Prewitt's operation for the edge detection. They used Freeman chain code (FCC) to represent the image character. FCC is generated by randomized algorithm and Support vector machine (SVM) was used for the classification.

R.Dineshkumar et al., 2014[12], presents an experimental assessment of the efficiency of BPNN and RBF neural network in terms of accuracy in recognition. They reported the comparative study of Sanskrit handwritten character recognition by using BPNN and RBF networks with 91% and 84% accuracy respectively. There are 43 training and testing samples are used.

S.S.Magare et. al., 2014 [1], presented offline Handwritten Sanskrit compound character recognition using Hough Transform and Euclidean Distance with an accuracy of 94%. Hough Transform is applied on the binarized edge map to generate the Hough image of it. The Prewitt operator is used in image processing, particularly within edge detection algorithms. Euclidean Distance classifier is used to classify features.

R. Dineshkumar et al., 2015 [13], proposed an off line handwritten character recognition framework using feed forward neural network. The handwritten Sanskrit character is resized into 20x30 pixels. System recognized the handwritten names and reading documents with 98% accuracy.

Table: 1 Related Work

S.No	Authors	Objective	Methods	Accuracy
1	N. Dwivedi et. al. ,2013 [10]	Sanskrit Character Recognition	Prewitt's Operator, SVM	
2	R. Dineshkumar et al. , 2014[12]	Sanskrit Character Recognition	Radial Based Function,	84%
			Back-Propagation Neural Network	91%
3	S.S.Magare et. al.,2014 [1]	Sanskrit Character Recognition	Hough Transform, Prewitt Edge Detection, Euclidean Distance	94%
4	S.P.Patil et. al., 2014[11]	Online Handwritten Sanskrit Character Recognition	SVM	-

 ISSN: 2454-4248
18 – 22

5	R. Dineshkumar et al. , 2015 [13]	Recognizing the handwritten names, reading documents	Artificial Neural Network.	98%

IV. CHALLENGES OF SANSKRIT CHARACTER RECOGNITION

Every written language has an alphabet of characters (letters), punctuation symbols etc. Every languages has basic property of writing characters , which makes communication easier is that differences between different characters are more significant than differences between different drawings of same character [7]. Either in offline or online handwriting recognition system, the character recognition is basically affected by variations in writing style of different writes and also different styles of writing of same person on different times.

A. Handwriting Style Variations

Different person have various handwriting styles as shown in fig. 6. Even same writer have various handwriting style over the time. Many times a writer finds herself /himself unable to recognize her/his own handwriting. Hence, practically it is much difficult to recognize handwriting by machine efficiently [7].

Deformed distortion, noise, overlapping, skews, slants, geometry are inserted by different writes in many ways. Geometric properties such as position, and aspect ratio and size vary. In handwriting recognition system many time errors arise due to the confusion among the similar shaped characters. There are many similar shaped characters in Devanagari script [14]. Some groups of similar shaped characters are shown in Fig. 7. There is very less difference in characters within corresponding groups. Characters can look similar although their number of strokes, and the drawing order and direction of the strokes may vary considerably.

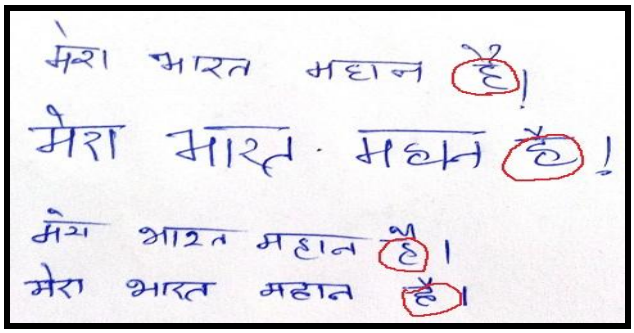


Figure:6 Various Writing Style of Same Character

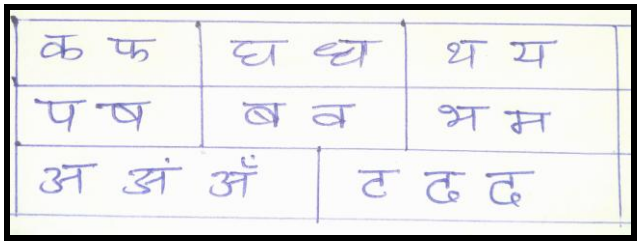


Figure:7 Similar shaped Devanagari Character

B. Constrained and Unconstrained Handwriting

The characters to recognize may be constrained or unconstrained. Because unconstrained documents include all possible style variation, so such document are much difficult to recognize.

In constrained document format, the handwritten samples are written in standard format that make the characters easy to recognize. The constrained document has box discrete and space discrete nature of characters or words. In box discrete nature, each character is written in separate standard sized box. In space discrete nature, characters are written have much space from one-another to make the segmentation and recognition easier as shown in Fig.8.

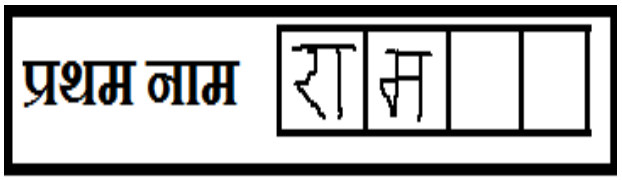


Figure:8 Constrained Handwriting

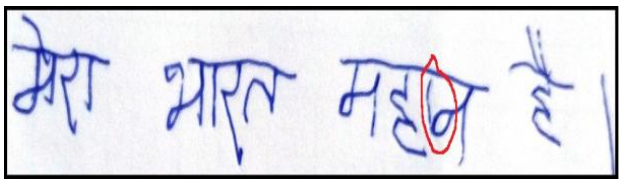


Figure:9 Unconstrained Handwriting: Overlapping of Characters

Unconstrained documents consist of touching, overlapping and cursive characters. In cursive writing there is variability in number of stokes which makes recognition difficult. To segment the characters from each other in touching characters are difficult and overlapping characters makes this situation worse as shown in Fig. 9.

C. Writer Dependent or Independent Recognition

Writer dependent recognition system is used to recognize the samples of only those writers whose samples are takes to train the recognition system. In writer dependent system, all possible style variations can be trained to system, hence a higher recognition rate can be obtained.

At the other hand, writer independent system needs the generalization of the recognition system also to recognize the handwritten samples of unknown writers. Hence, it needs to train the system with all possible and commonly used style variations. Hence it needs to train the system with large number of samples taken from large number of different types of writers, to make the recognition system generalized. Due to recognition of unknown samples, the recognition rate of writer independent system is comparatively lower. In practice, writer independent systems are in more demand because of generalized application.

D. Personal and Situational Aspects

Personal factors include writer's writing style which might be affected by handedness- either left handed or right handed as shown in Fig.10. A good recognition requires neat and clean handwriting and this writing style also depends on profession of writers to some extent. The situational aspects depend on the facts either writer is interested or not to write, how much attention a writer is paying, text is written giving

 ISSN: 2454-4248
18 – 22

proper time or in hurry, whether there was any interruption while writing, how was the quality of material used for writing etc.

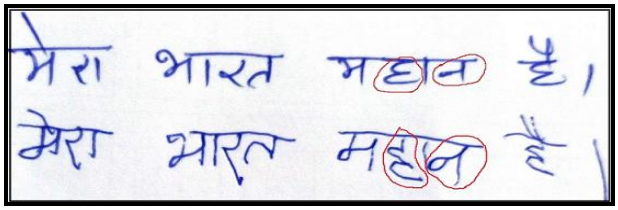


Figure:10 Same Writer writing with different hand; 1st is right handed and 2nd is left handed.

E. Different Sequence of Strokes

The sequence of strokes and number of strokes for a same character may various, due to different way of writing by different writers as shown in fig.11.

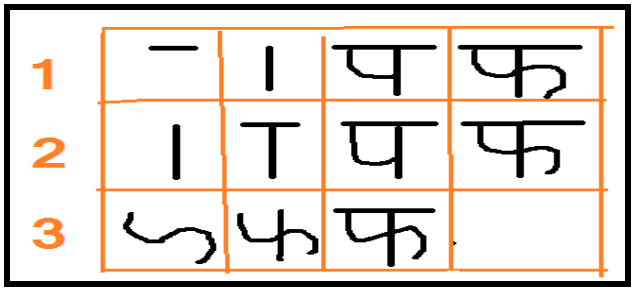


Figure:11 Sequence of Strokes for Same Character

F. Different Number of Strokes

A stroke is created by the set of coordinate points (x, y). Any language have a single character or symbol, which is used by the user for writing, may formed by one or more than one strokes. Fig.12. shows handwritten character "va" of Sanskrit language. "va" character is written with the help of three different strokes as mentioned in Fig. 12 (a) and also with two different strokes as mentioned in Fig. 12 (b). Same character has different number of stroke due to different writing style by different writers as shown in Fig. 12.

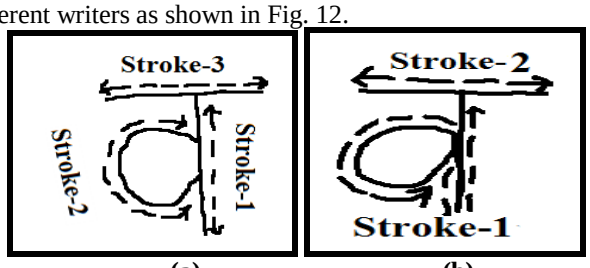


Figure:12 Different Number of Strokes for Same Character

G. Number of stroke classes

There are various composite characters in Sanskrit language, so there is a large number of stroke classes are possible. These stroke classes represent consonants, vowels and modifiers or combinations of consonants and vowels. The recognition system complexity increases due to a large number of stroke classes and the shape complexity of various stokes. The composite character "त्र ″ is made by one half character and one full character as shown in fig.13.

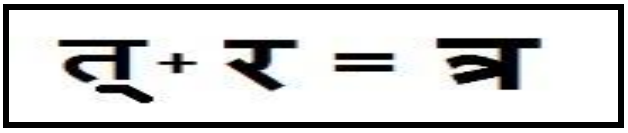


Figure:13 Composite Character

H. Directionality of writing

There exist big variations in the directionality of writing strokes and stroke segments which could affect the uniformity in stroke representation using certain features. It is necessary to identify writing direction invariant features for representing the stroke. Sanskrit language is written from left to right while Urdu is written from right to left.

V. CONCLUSION

A large number of Sanskrit handwritten documents, books and holy books are stored in libraries. There is need to convert these handwritten texts into editable or digital form, to make available online. So, that people can easily access these documents through internet. Different types of HRS have been developed but they are limited up to some Indian languages. In coming days the demand and scope of Online HRS will increased, due to growth in Mobile technologies. There are also a number of challenges for recognition of handwritten Sanskrit characters and words. Researcher can do more work in field of Online Handwritten Sanskrit Character recognition using temporal information (strokes and directions of strokes) of writing on electronic devices.

References

[1] Magare, Sujata S., and Ratnadeep R. Deshmukh. "Offline handwritten sanskrit character recognition using hough transform and euclidean distance." International Journal of Innovation and Scientific Research 10 (2014): 295-302.

[2] Singh, Shailendra Kumar, and Manoj Kumar Sachan. "Importance and Challenges of Social Media Text." *International Journal of Advanced Research in Computer Science* 8.3 (2017).

[3] Singh, Shailendra Kumar, et al. "Sentiment Analysis of Twitter Data Set: Survey." *International Journal of Applied Engineering Research* 9.22 (2014): 13925-13936.

[4] Singh, G., and M. Sachan. "A framework of online handwritten gurmukhi script recognition." International Journal of Computer Science and Technology (IJCST) 6 (2015): 52-56.

[5] Singh, Gurpreet, and Manoj Sachan. "Multi-layer perceptron (MLP) neural network technique for offline handwritten Gurmukhi character recognition." Computational Intelligence and Computing Research (ICCIC), 2014 IEEE International Conference on. IEEE, 2014.

[6] Sharma, Preeti, and Manoj Kumar Sachan. "A Review on Character Segmentation of Touching and Half Character in Handwritten Hindi Text." *International Journal of Advanced Research in Computer Science* 8.3 (2017).

[7] Charles C. Tappert, Ching Y. Suen and Toru Wakahara. "The State of the Art in On-Line Handwriting Recognition". IEEE TRANSCIIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE , Vol. 12, No. 8, pp. 787-808, August 1990.

[8] C. Y. Suen. "Character recognition by computer and applications." In T. Y. Young and K. S. Fu, editors, Handbook of Pattern Recognition and Image Processing. Academic Press, Inc., Orlando, FL, 1986, pp. 569–586.

 ISSN: 2454-4248
18 – 22

[9] C. C. Tappert, A. S. Fox. J. Kim, S. E. Levy, and L. L. Zimmerman. "Handwriting recognition on transparent tablet over flat display." in *1986 SID Inc. Symp. Dig. Tech. Papers*, May 1986. Pp.308-312.

[10] Namita Dwivedi,K. Srivastava & N. Arya. "SANSKRIT WORD RECOGNITION USING PREWITT'S OPERATOR AND SUPPORT VECTOR CLASSIFICATION". 2013 IEEE International Conference on Emerging Trends in Computing, Communication and Nanotechnology (ICECCN 2013).

[11] Sonal P.Patil, Priyanka P. Kulkarni. "Online Handwritten Sanskrit Character Recognition Using Support Vector Classification". Journal of Engineering Research and Applications, Vol. 4, Issue 5(Version 1), May 2014, pp.82-91.

[12] R. Dineshkumar and J. Suganthi ."A Comparative Study for Handwritten Sanskrit Character Recognition Using BPNN and RBF Networks" . Recent Advances in Electrical Engineering, 2014.

[13] R. Dineshkumar and J. Suganthi. "Sanskrit Character Recognition System using Neural Network". Indian Journal of Science and Technology, *Vol 8(1), 65–69, January 2015.*

[14] U. Pal, T. Wakabayashi and F. Kimura. "Comparative Study of Devnagari Handwritten Character Recognition using Different Feature and Classifiers". 10th International Conference on Document Analysis and Recognition, pp. 1111-1115, IEEE, 2009.

[15] Sachan, Manoj. "Analysis of Shape Variations in Recognition of Online Gurmukhi Script". International Conference on Electrical and Electronics: techniques and Applications (EETA2015), 2015, pp. 312-316.

[16] https://en.wikipedia.org/wiki/Sanskrit

[17] https://en.wikipedia.org/wiki/Handwriting