

Online Handwriting Recognition for Devanagari Script in Recent Past

Sukhdeep Singh

Assistant Professor in Computer Science, Mata Sahib Kaur Girls College, Talwandi Bhai, Ferozepur, Punjab (India)

ARTICLE DETAILS

Article History

Published Online: 10 October 2018

Keywords

Brahmi, Devanagari, Online handwriting recognition

*Corresponding Author

Email: sransingh13[at]gmail.com

ABSTRACT

Handwriting recognition (HWR) represents to the procedure of understanding handwritten text written on the paper or digital surface. When handwritten text is recognized while writing on the digital surface, it is called online handwriting recognition (OHWR). An OHWR model comprises of different phases as data collection, preprocessing, segmentation, feature extraction and classification etc. The present study has been done for the Brahmi script Devanagari, which is the script of Indian national language Hindi and it is used in all regions of the country. This study shows the work done for online Devanagari handwriting recognition in recent past. This study has taken into consideration the work done for larger units (words or sentences) and smaller units (as characters and strokes). The present study has presented the results from reputed journals/conferences of pattern/handwriting recognition as PR (Pattern Recognition), PAMI (Pattern Analysis and Machine Intelligence), PRL (Pattern Recognition Letters), IWFHR (International Workshop on Frontiers in Handwriting Recognition), IJDAR (International Journal on Document Analysis and Recognition), ICDAR (International Conference on Document Analysis and Recognition) and ICFHR (International Conference on Frontiers in Handwriting Recognition) etc.

1. Introduction

With the passage of time, the technological innovations are being made day by day. It would not be wrong to say that the present era is the technological era. In this technological period, the artificial intelligence is playing a vital role. With the technological advancement, the methods of communicating with computing devices are also changing. The speech recognition is one of the successfully used communication methods between human beings and computing devices. But speech recognition has certain limitations while communicating with computing devices. To avoid the limitations of speech recognition, the handwriting recognition is one of the active of research for last few decades. Handwriting recognition refers to understanding the handwritten text written on paper or digital surface. When the handwritten text is first written on the paper and then it is scanned and understood by the computer, it is called offline handwriting recognition. But on the other hand, when the handwritten text is understood by the computer while writing on the screen or digital surface of the computer, it is called online handwriting recognition. The work in the present study has been done for online Devanagari handwriting recognition.

A great work for offline and online handwriting recognition has been done in past decades. The major work has been carried out for Chinese-Japanese-Korean (CJK) and Latin scripts. The work for offline and online handwriting in Indic scripts is also getting the attention of pattern/handwriting recognition researchers across the world. But as compared to CJK and Latin scripts, the handwriting recognition work for Indic scripts did not receive that much attention in past and such available work for Indic scripts is extremely limited yet.

This available work is especially less for Indic scripts online handwriting recognition in larger units. The most of available work for online handwriting in Indic scripts is for characters or strokes. The Indic scripts are also called the Brahmi scripts. These scripts are used by the people residing around the Indus River. More than ten scripts fall into the family of Indic scripts. The most of these scripts have more than two hundreds and fifty alphabets comprising basic and compound characters [1]. One of the Indic scripts is the Devanagari script. Devanagari is the script of Indian national language Hindi. It is the most popular script of India. The people who use Devanagari script reside in all parts of India. Devanagari is not an isolated script; it shares many properties with other Indic and non-Indic scripts. The stroke order variation challenge present in Latin script is also present in Devanagari script and the symbol order variation challenge present in Devanagari script is also present in Gurmukhi script.

In order to recognize online handwritten words, character or strokes in any script, it is essential to identify the fundamental writing units of scripts. These fundamental units of the script are called symbols. The symbol set for the Devanagari script is shown in figure 1. Here, the symbol 0 is the headline. The independent vowels are shown with the symbols 1 to 11. The consonants with implicit vowel sounds are shown with the symbols 12 to 44. The matras are shown with symbols 48 to 63 and 109 to 110. The sentence ending symbol is the 64. The symbols 65 to 95 represent the half characters in Devanagari. The conjuncts are represented with symbols 45 to 47 and 96 to 108.

—	अ	आ	इ	ई	उ	ऊ	ऋ	ए	ऐ	ओ
0	1	2	3	4	5	6	7	8	9	10
औ	क	ख	ग	घ	ङ	च	छ	ज	झ	ञ
11	12	13	14	15	16	17	18	19	20	21
ट	ठ	ड	ढ	ण	त	थ	द	ध	न	प
22	23	24	25	26	27	28	29	30	31	32
फ	ब	भ	म	य	र	ल	व	श	ष	स
33	34	35	36	37	38	39	40	41	42	43
ह	क्ष	त्र	ज्ञ	ा	ि	ी	ु	ू	े	ै
44	45	46	47	48	49	50	51	52	53	54
ँ	ो	ौ	ँ	ँ	ँ	ॠ	ँ	ँ	।	व
55	56	57	58	59	60	61	62	63	64	65
ॠ	ॡ	ॢ	ॣ	।	॥	०	१	२	३	४
66	67	68	69	70	71	72	73	74	75	76
६	७	८	९	०	१	२	३	४	५	६
77	78	79	80	81	82	83	84	85	86	87
७	८	९	०	१	२	३	४	५	६	७
88	89	90	91	92	93	94	95	96	97	98
८	९	०	१	२	३	४	५	६	७	८
99	100	101	102	103	104	105	106	107	108	109
९	०	१	२	३	४	५	६	७	८	९

Figure 1.Symbol set defined for Devanagari

There is great amount of changeability and unpredictability of writing style of Devanagari writers for writing Devanagari text on digital surface. One of the key challenges for online Devanagari handwriting recognition is the isomorphic (similar shape) nature of strokes or characters. Some Devanagari characters/strokes are in isomorphic nature originally and others because of the writing style of writers. The stroke order and size variation, stroke connections and shape variations of strokes, composition of multiple characters in a single stroke and the absence or presence of headline in handwritten Devanagari words are the other challenges for online Devanagari handwriting recognition.

The present study is an important step and effort for online Devanagari handwriting recognition and it will be definitely proved as a valuable study for future researchers/readers in the direction of online Devanagari handwriting recognition.

2. Literature Survey

In case of online handwriting recognition, it is useful to study the prior studies and results for online handwritten Devanagari text recognition in recent decades. The available studies for online Devanagari handwriting recognition are for

both the smaller and larger units. The most of the available studies for online Devanagari handwritten text recognition are for smaller units as characters or strokes. As other Brahmi scripts, the key work for online Devanagari handwriting recognition has been carried out in twenty first century. So we have surveyed the major work for Devanagari script in last decade. The table 1 shows the recognition results for online Devanagari handwriting recognition in the most recent past.

a) Online handwriting recognition for Devanagari strokes and characters

In 2006, Swethalakshmi et al. [7] presented a novel study to recognize isolated Devanagari characters in online handwriting mode. In their work, they employed SVM the statistical recognition technique for classification. For Devanagari character recognition, based on proximity analysis, they formed stroke groups explicitly. They obtained recognition results as 96.69% and 97.27% for 42 classes and 82 classes, respectively. A notable work has been done for the development of benchmarked datasets for online handwritten Devanagari character recognition. Most of these datasets are also freely available.

Table 1. Recognition results for online Devanagari handwriting

Sr.no.	Authors and references	Year	Units	Classification techniques	Recognition rate (%)
1	Ghosh and Roy [2]	2016	Words	Zone wise slopes of dominant points (ZSDP) approach, Hidden Markov Models (HMM)	93.82

2	Ghosh and Roy[3]	2015	Characters	ZSDP, Support Vector Machine (SVM)	90.63
3	Ghosh and Roy [3]	2015	Characters	Characters Zone wise structural and directional features (ZSD), SVM	85.10
4	Chowdhury et al. [4]	2013	Characters	Levenshtein distance metric	83.95
5	Bharath and Madhvanath [5]	2012	Words	HMM	87.13
6	Mondal et al. [6]	2010	Characters	Point-float feature, HMM	82.43
7	Mondal et al.[6]	2010	Characters	Point-float feature, Multilayer perceptron (MLP)	83.30
8	Mondal et al. [6]	2010	Characters	Chain-code feature, HMM	87.13
9	Mondal et al. [6]	2010	Characters	Chain-code feature, MLP	86.15
10	Swethalakshmi et al. [7]	2007	Characters	SVM	96.69 (42 classes) and 97.27 (82 classes)

In 2010, the cost free availability of these datasets has motivated Mondal et al. [6] to demonstrate the benchmarked results for recognition of online handwritten Devanagari character recognition of Tamil, Telugu, Bangla and Devanagari scripts. In their work, they have employed the point-float features values and chain code histogram feature values to denote feature vectors. Further, they employed the MLP, nearest neighbour and HMMs for classification job. In their work, they obtained recognition accuracy results for six different combinations by employing every classifier with every feature extraction technique. Based on their experimentation, they made conclusion that the chain code histogram feature values give the best accuracy rates irrespective the classification technique employed. They also made conclusion that the best recognition results are obtained with NN classification technique irrespective the point-float and chain code histogram based feature extraction technique used. In 2013, Chowdhury et al. [4] considered the similarity of train and test samples for online handwritten Devanagari character recognition and other Indic scripts as Bangla, Tamil and Telugu also. They proposed a novel technique to recognize isolated characters in Brahmi scripts. In their work, to calculate similarity between samples, they employed Levenshtein distance metric and feature vector based on position and shape information. Recently, in 2015, Ghosh and Roy [3] employed SVM classification technique to recognize online handwritten Devanagari character recognition. In their work, they used ZSDP and ZSD approaches for feature extraction and they achieved better recognition results using ZSDP.

b) Online handwritten Devanagari words recognition

From the recent decade's literature, it is found that there is not much work done for online Devanagari handwriting to recognize online handwritten words. Only Bharath and Madhvanath [5], and Ghosh and Roy [2] made their vital

contribution in this direction. In 2012, Bharath and Madhvanath [5] considered the important work done for Latin, CJK and Arabic scripts and they referred the same work for online handwriting recognition in Indic scripts. In Indic scripts, they have also used this work for Devanagari script. They analyzed the differences and relationships/similarities of Devanagari script and other Indic and non-Indic scripts for online handwritten character recognition. As an illustration, the stroke order variation challenge in Devanagari is also faced by Latin and other Indic scripts. On the other hand, the symbol order variation problem faced by Devanagari script is not faced by Latin or CJK scripts. The comparison for zone based features for online handwritten Devanagari and Bangla word recognition has been done by Ghosh and Roy [2] in 2016. In their work, they used HMM statistical model for classification.

3. Conclusion

In this study, the authors have demonstrated the work carried out for online Devanagari handwriting recognition in most recent decade. The work has been done for smaller and larger units both. In the past studies, it has been observed that the smaller units' work has been done for digits, characters and strokes. In recent studies, the larger units work for online Devanagari handwriting recognition has been done for words. The present study is an important guide for future readers and researchers to carry out further/advance research in the direction of online handwritten Devanagari text recognition. Our study has demonstrated that there is the great necessity to carry out more research work for online Devanagari handwriting recognition for larger units as words/sentences. For online handwritten Devanagari sentence recognition, a lot to be done in the near future.

References

1. B. B. Chaudhuri and U. Pal, "An OCR system to read two Indian language scripts: Bangla and Devanagari (Hindi)", In: *Proceedings of 4th International Conference on Document Analysis and Recognition*, 1997.
2. R. Ghosh and P. P. Roy, "Comparison of Zone-Features for Online Bengali and Devanagari Word Recognition Using HMM", In: *Proceedings of 15th International Conference on Frontiers in Handwriting Recognition (ICFHR)*, 435-440, 2016.
3. R. Ghosh and P. P. Roy, "Study of two zone-based features for online Bengali and Devanagari character recognition", In:

- Proceedings of 13th International Conference on Document Analysis and Recognition (ICDAR)* 401–405, 2015.
4. S. D. Chowdhury, U. Bhattacharya, and S. K. Parui, "Online Handwriting Recognition Using Levenshtein Distance Metric", *In: Proceeding of 12th International Conference on Document Analysis and Recognition*, 79–83, 2013.
 5. A. Bharath and S. Madhvanath, "HMM-Based Lexicon-Driven and Lexicon-Free Word Recognition for Online Handwritten Indic Scripts", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34 (4), 670–682, 2012.
 6. T. Mondal, U. Bhattacharya, S. K. Parui, and K. Das, "On-line handwriting recognition of Indian scripts - the first benchmark", *In: Proceeding of the 12th International Conference on Frontiers in Handwriting Recognition*, 200–205, 2010.
 7. H. Swethalakshmi, A. Jayaraman, V. S. Chakravarthy and C. Sekhar, "Online handwritten character recognition of Devanagari and Telugu characters using support vector machine", *In the proceedings of 10th IWFHR*, 367-372, 2006.