Journal of Advances and Scholarly Researches in Allied Education
Vol. 15, Issue No. 9, October-2018, ISSN 2230-7540

Online Handwritten Words Recognition for Devanagari and Tamil Script in Recent Past

Sukhdeep Singh*

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

Abstract – In the past years, a great work has been done for Chinese, Japanese, Korean, Latin, Arabic and Indic scripts handwriting recognition (HWR). Handwriting recognition is the process to understand handwritten text. The handwritten text can be written on the paper or digital surfaces. When the handwritten text written on the paper is first scanned and then understood by the computer, it is called offline handwriting recognition. But on the other hand, when the handwritten text is understood while writing on the digital surface, it is called online handwriting recognition. Thus, an online handwriting system can recognize/understand digitally handwritten strokes, characters, words and sentences. Online handwritten word recognition (OHWR) system has various phases like data collection and preprocessing, segmentation of strokes, feature extraction, classification and post processing etcetera. The present work has been done for online handwriting recognition in Tamil and Devanagari Indic scripts. This work has been done for online handwriting recognition of larger units (words). The key results presented in this work have been obtained from reputed journals and conferences of pattern recognition and handwriting recognition as Pattern Analysis and Machine Intelligence (PAMI), Pattern Recognition(PR), Pattern Recognition Letters (PRL), International Journal on Document Analysis and Recognition(IJDAR) , International Conference on Document Analysis and Recognition (ICDAR) and International Conference on Frontiers in Handwriting Recognition (ICFHR) etcetera.

Keywords: Indic, Tamil, Devanagari, Online Handwriting Recognition

----------X----------

INTRODUCTION

With the passage of time, the technology is changing at a fast pace. New technological innovations have been proved a boon in the modern era. The modern era has also accepted the change in methods of exchanging information between computing devices and human beings. The artificial intelligence, machine learning and pattern recognition have played a vital role to communicate with computing devices. Speech recognition has been successfully used in these days. But, in order to avoid certain limitations of speech recognition, the handwriting recognition has buzz word in recent past where a tremendous work has been done for both offline and online handwriting recognition. Handwriting recognition is a method of exchanging information between computers and human beings where handwritten text is understood by computers. The handwritten text can be written on the paper or digital surfaces. When the handwritten text written on the paper is first scanned and then understood by the computer, it is called offline handwriting recognition. But on the other hand, when the handwritten text is understood while writing on the digital surface, it is called online handwriting recognition. Thus, an online handwriting system can recognize/understand digitally handwritten strokes, characters, words and sentences. The work done for offline and online handwriting recognition in Chinese, Japanese, Korean, Latin and Arabic scripts is great. When compared with Indic scripts handwriting recognition work to these scripts, the available work for Indic scripts online handwriting recognition is limited. So, Indic scripts require the attention of handwriting recognition researchers across the world. The most of the available studies for online handwriting recognition in Indic scripts is for smaller units (strokes and characters). Thus the Indic scripts online handwriting recognition research area for larger units needs more attention. There are more than 10 Indic scripts. The most of Indic scripts have more than 250 alphabets (basic and compound characters) (Chaudhuri and Pal, 1997). This study has been carried out for online Tamil and Devanagari handwriting recognition. These scripts are among mostly widely used scripts in India. These are not isolated script; these have many similar characteristics with other scripts. As an illustration, the stroke and symbol order variation problem/challenge is common for most of Indic scripts and it is also present in these two scripts.

For online handwritten words recognition in any script, the first step is the identification of

fundamental writing units of script. These fundamental/primary units of the script are known as symbols. The symbol sets for Devanagari and Tamil scripts are shown in figure 1 and 2, respectively. In figure 1, the symbol 0 is the headline in Devanagari which is also called the *shirorekha*. Devanagari independent vowel symbols are presented with the symbol numbers 1 to 11. The consonants having implicit vowel sounds are shown with symbol numbers 12 to 44. Devanagari matras are presented with symbols 48 to 63 and 109 to 110. The symbol 64 is the sentence ending symbol in Devanagari. The symbols 65 to 95 present the half characters in Devanagari. The Devanagari conjuncts are shown with symbol numbers 45 to 47 and 96 to 108.

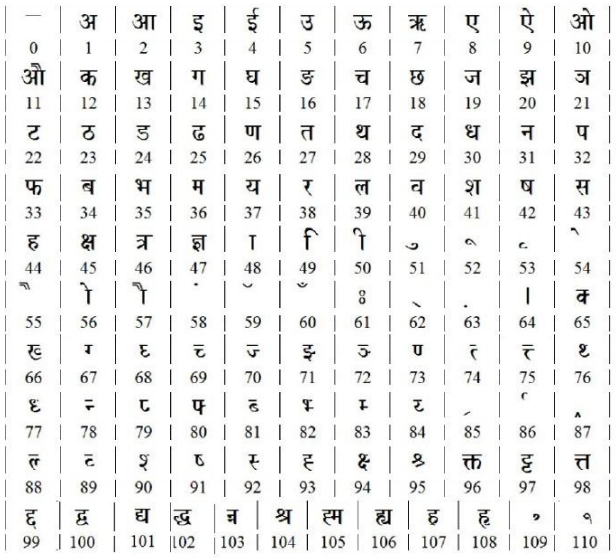

Figure 1. Symbols for Devanagari script

In figure 2, the symbols 0 to 10 represent the Tamil vowels, the symbol 11 is the special symbol, the symbols 12-33 are Tamil consonants and these consonants/symbols have implicit vowel sound, and the symbols 72-80 show the Tamil script matras. When symbol 72 is placed over the Tamil script consonants, the Tamil consonants are altered to their half character forms. The symbol 81 is always used with symbol 75 in Tamil script. The symbols 81 and 82 are conjuncts. The Tamil symbol 83 is the dot/period. The rest Tamil symbols are made with consonants and vowels combinations, and these are known as syllabic units of Tamil script.

Figure 2. Symbols for Tamil script

All Indic script writers do not handwrite their text in the same way and the same is present for Devanagari and Tamil script digital handwriting, and there is a great degree of unpredictability and variability of writing styles of Devanagari and Tamil writers for writing these scripts text. The similar stroke or symbol shapes are the major challenges for online Devanagari and Tamil handwriting recognition. These similar shaped strokes or symbols are called Isomorphic natured strokes or symbols. There are two causes for isomorphic nature of strokes and symbols. Some strokes and symbols have isomorphic nature originally and others because of the writing style of writers. The stroke size and order variation, symbol size and order variation, stroke shapes and stroke connections variation, multiple characters composition in a one stroke and the presence or absence of headline in handwritten Devanagari and Tamil texts are among the key challenges for online handwriting recognition in these two scripts.

The present study is an important step for online handwritten word recognition in Devanagari and Tamil script and it will be a very useful for future researchers and readers to carry out work for these scripts online handwriting recognition for larger units as words.

LITERATURE SURVEY

For online handwriting recognition in Devanagari and Tamil scripts, it is essential to analyze and survey the existing studies for online handwritten Devanagari and Tamil text recognition in past. The available work for Devanagari and Tamil scripts online handwritten text recognition has been done for smaller units as strokes and characters, and larger units as words. The most of existing works for online handwritten Devanagari and Tamil text recognition has been done for strokes and characters. But, in view to see the scope of online handwriting recognition work for larger units, the

present study has reviewed the related work for Devanagari and Tamil scripts for larger units as words. As other Indic scripts, the most work for Devanagari and Tamil scripts online handwriting recognition has been carried out in past two decades. The table 1 presents the Devanagari and Tamil scripts online handwriting recognition work in the most recent past.

Table 1.Online handwritten word recognition results for Devanagari and Tamil script

Sr. No.	Authors and references	Script	Year	Classification technique applied	Accuracy rate (%)
1	Ghosh and Roy [2]	Devanagari	2016	Zone wise slopes of dominant points (ZSDP) approach, Hidden Markov Models (HMM)	93.82
8	Urala et al. [3]	Tamil	2014	SVM + bigram	89.2 (Symbol level accuracy, GNote data), 74.5 (Word level accuracy, GNote data)
9	Urala et al. [3]	Tamil	2014	SVM + bigram	83.22 (Symbol level accuracy, tablet PC data), 54.2 (Word level accuracy, tablet PC data)
10	Urala et al. [3]	Tamil	2014	SVM	78.52 (Symbol level accuracy, tablet PC data), 40.05 (Word level accuracy, tablet PC data)
12	Sundaram and Ramakrishnan [4]	Tamil	2013	Dominant overlap criterion segmentation (DOCS), SVM	50.9
13	Sundaram and Ramakrishnan [4]	Tamil	2013	Attention feedback segmentation (AFS), SVM	64.9
5	Bharath and Madhvanath [5]	Devanagari	2012	HMM	87.13
14	Bharath and Madhvanath [5]	Tamil	2012	HMM	91.8
15	Sundaram and Ramakrishnan [6]	Tamil	2011	Dominant overlap segmentation (DOS), SVM	86.9 (Symbol level accuracy)
20	Bharath and Madhvanath [7]	Tamil	2007	HMM	94.49, 93.17 and 92.15 for 5k, 10k and 20k words, respectively

a) Online handwritten Devanagari words recognition

From the past available studies, it is found that there is a limited work done for online Devanagari handwriting to recognize online handwritten words. Bharath and Madhvanath [5], and Ghosh and Roy [2] have contributed for Devanagari and Tamil scripts' work in this direction. In 2012, Bharath and Madhvanath [5] have taken into consideration the important work done for Latin, Chinese, Japanese, Korean and Arabic scripts, and they referred and used the same work for online handwriting recognition in Brahmi scripts. In Brahmi scripts, they have employed this work for Northern Indic script Devanagari. They analyzed and studied the differences and similarities/ relationships of Devanagari script and other Indic and non-Indic scripts for online handwriting recognition. For example, the stroke order variation challenge faced by Devanagari is also present in Latin and other Indic scripts. But, on the other side, the symbol order variation problem present in Devanagari script is not present in Latin or CJK scripts. An important comparison for features (zone based) in online Devanagari and Bangla word handwriting recognition has been made by Ghosh and Roy [2] in 2016. In their work, they employed hidden Markov models for classification.

b) Online handwriting recognition for Tamil script words

Bharath and Madhvanath [5] have used the HMM based word modelling for online handwritten Tamil words recognition in 2007. For experimentation in their work, they employed different sized datasets of Tamil words and got best recognition with lexicon size of 1k words. In their work, they recognized Tamil words in writer independent environment of handwriting. In 2011, Sundaram and Ramakrishnan [7] provided an innovative segmentation approach to segment Tamil words in online mode of handwriting. This lexicon free segmentation approach is applicable to rest of Indic and non- Indic scripts also which are non-cursive in nature. In 2012, Bharath and Madhvanath [5] have considered the important work done for Latin, Chinese, Japanese, Korean and Arabic scripts, and they employed the output of non-Indic scripts work in Indic scripts online handwritten word recognition work. In Indic scripts, they have employed these studies for Tamil script also. In their work, they made study and analysis to find the similarities and differences of Tamil and other Indic and non-Indic scripts for online handwriting recognition. For example, the stroke order variation challenge present in Tamil script is also tackled by other Indic and non-Indic scripts. On the other side, the symbol order variation problem faced by Tamil script is not present in Latin or other non-Indic scripts. In 2013, Sundaram and Ramakrishnan [5] proposed a two moduled segmentation technique for online handwritten Tamil words. This technique is lexicon free and script dependent, and over and under segmentation is solved by it. In 2014, Urala et al. [3] described a complete system for online handwriting recognition of isolated Tamil words. Further, they have also made enhancement of their study to paragraph handwriting recognition. In their work, they described segmentation, preprocessing, feature extraction, classification and bigram-based post processing phases of online handwritten Tamil word recognition.

CONCLUSION

The present study has demonstrated the most recent online handwriting recognition work done for Devanagari and Tamil scripts. This work has been done for recognition of online handwritten words. In past studies, it has been observed that the most of

available works for Devanagari and Tamil scripts online handwriting has been done for the smaller units. The present study is an important step for future researchers and readers to do further research for online handwritten Devanagari and Tamil text recognition. This study has showed that there is the great requirement to do further research work in online Devanagari and Tamil handwriting recognition for larger units as words and sentences. For online handwritten Devanagari and Tamil script sentence recognition, a lot to be done yet.

REFERENCES

1. B. B. Chaudhuri and U. Pal (1997). "An OCR system to read two Indian language scripts: Bangla and Devanagari (Hindi)", In: *Proceedings of 4th International Conference on Document Analysis and Recognition*.

2. R. Ghosh and P. P. Roy (2016). "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),* pp. 435–440.

3. K. B. Urala, A. G. Ramakrishnan, and S. Mohamed (2014). "Recognition of open vocabulary, online handwritten pages in Tamil script", *International Conference on Signal Processing and Communications*, 1–6.

4. S. Sundaram and A. G. Ramakrishnan (2013). "Attention-Feedback Based Robust Segmentation of Online Handwritten Isolated Tamil Words*", ACM Transactions on Asian Language Information Processing,* 12(1).

5. A. Bharath and S. Madhvanath (2012). "HMM-Based Lexicon-Driven and Lexicon-Free Word Recognition for Online Handwritten Indic Scripts", *IEEE Transactions on Pattern Analysis and Machine Intelligence,* 34 (4), pp. 670–682.

6. S. Sundaram and A. G. Ramakrishnan (2011). "Lexicon-free, novel segmentation of online handwritten Indic Words", In: *Proceedings of 11th International Conference on Document Analysis and Recognition*, pp. 1175–1179.

7. A. Bharath and S. Madhvanath (2007). "Hidden Markov Models for Online Handwritten Tamil Word Recognition", In: *Proceedings of 9th International Conference on Document Analysis and Recognition*.

Corresponding Author

Sukhdeep Singh*

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

sransingh13@gmail.com