

Online Handwriting Recognition for Bangla 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 January 2019

Keywords

Indic, Bangla, Online handwriting recognition

Corresponding Author

Email: sransingh13[at]gmail.com

ABSTRACT

Online handwriting recognition (OHWR) refers to the process of understanding handwritten text while writing on the digital surface. A generic OHWR system consists of various phases like data collection, preprocessing, segmentation, feature extraction and classification etc. The present work has been done for the Indic script Bangla, which is widely used in eastern India and Bangladesh. The present study presents the work done for online Bangla handwriting recognition in recent past. The study considered the work done for smaller units as well as the larger units. The results in the present study have been presented from premier journals/conferences of pattern/handwriting recognition as PAMI (Pattern Analysis and Machine Intelligence), PR(Pattern Recognition), PRL(Pattern Recognition Letters), IJDAR (International Journal on Document Analysis and Recognition), IWFHR (International Workshop on Frontiers in Handwriting Recognition), ICFHR (International Conference on Frontiers in Handwriting Recognition) and ICDAR (International Conference on Document Analysis and Recognition) etc.

1. Introduction

The modern era is the era of technology. Technological inventions are being made with the passage of time. The technological advancement has also resulted in the change of methods for interaction between computers and human beings. The speech recognition and handwriting recognition are vital methods of communicating with computers. Handwriting recognition refers to understanding the handwritten text by computer. There are two types of handwriting recognition: offline handwriting recognition and online handwriting recognition. In case of offline handwriting recognition, the computer understands the handwritten text written on the paper. Online handwriting recognition means that the computer understands the handwritten text while writing on the digital surface or screen. The present work has been done for online handwriting recognition.

Although a lot of work has been done for online handwriting in recent past. But in comparison to Chinese, Japanese, Korean and Latin scripts, the work done for online handwriting recognition in Brahmi scripts is less. The Brahmi scripts are also called the Indic scripts. The Indic or Brahmi scripts are those scripts that are used by people around the Indus River. The major available study of online handwriting in Indic scripts is for smaller units as strokes or characters. The Indic scripts' work for larger units as words or sentences are very limited. There are more than ten Indic scripts and most of these scripts have more than 250 alphabets including basic and compound characters [1]. One of the Indic scripts is the Bangla script. Bangla script is used by more than 200 million people of the world. Bangla is the fifth most popular language in the world and second most popular language of India. Most Bangla speaking people reside in eastern part of India.

অ	আ	ই	ঈ	উ	ঊ	ঋ	এ	ঐ	ও
ঔ	ক	খ	গ	ঘ	ঙ	চ	ছ	জ	ঝ
ঞ	ট	ঠ	ড	ঢ	ণ	ত	থ	দ	ধ
ন	প	ফ	ব	ভ	ম	য	র	ল	শ
ষ	স	হ	ড়	ঢ়	য়	ৎ	ং	ঃ	৷

Figure 1. Bangla basic characters

The alphabets of Bangla script are not only used in Bangla language, these are also used in other Indian languages as: Assamese and Manipuri. The symbol set of modern Bangla script comprises of 40 consonants and 11 vowels. The shapes of two Bangla characters are same. These all symbols of the script are called the *basic characters* of Bangla. The character set of Bangla script is shown in figure 1. In case of Bangla

script, unlike Latin script, there is absence of upper or lower case. The Bangla words are written in left to right direction and most of words have a horizontal line at upper part. In Bangla script, the compound characters are formed by combinations of two consonants or combinations of consonant and vowel. To form compound characters, combining more than two symbols

also exist in Bangla script. The number of compound characters in Bangla is around 280.

There is high degree of variability in styles for writing Bangla text on digital surface. A few samples for online handwritten Bangla characters are demonstrated in figure 2. One of the major difficulties for online handwritten Bangla

text recognition is isomorphic nature of characters. The other problems for online Bangla handwriting are stroke size variation, order variation of strokes, stroke connections and shape variations of strokes, composition of multiple characters in a single stroke and presence or absence of horizontal line in handwritten Bangla words.



Figure 2. Illustration of online handwritten Bangla characters

The present work is a step in direction of online Bangla handwriting, and it will be proved as a useful work for future researchers and readers in the direction of online handwritten Bangla text recognition.

2. Literature Survey

In context of handwriting recognition, it may be worth studying the existing results for online handwritten Bangla text recognition in recent past. The earlier work that we found for

online Bangla handwriting recognition is for smaller units as well as larger units, but the major available studies are for stroke or character recognition. It has been noticed that the major work for online Bangla handwriting recognition has been done in last decade, so we have surveyed the major Bangla script work in last decade. The table 1 demonstrates the recognition results for online Bangla handwriting recognition in recent past.

Table 1. Recognition results for online Bangla handwriting

Sr. no.	Authors and references	Year	Units	Classification techniques	Recognition rate (%)
1	Bhattacharya et al. [2]	2017	Words	Hidden Markov Model (HMM)	89.29 (All test set words)
2	Bhattacharya et al. [3]	2016	Words	Support Vector Machine (SVM)	94.3 (Character level accuracy)
3	Ghosh and Roy [4]	2016	Words	Zone wise slopes of dominant points (ZSDP) approach, HMM	90.23
4	Ghosh and Roy [5]	2015	Characters	ZSDP, SVM	92.48
5	Ghosh and Roy [5]	2015	Characters	Zone wise structural and directional features (ZSD), SVM	87.48
6	Samanta et al. [6]	2015	Words	DCEseg, HMM	84.67
7	Samanta et al. [6]	2015	Words	Minseg, HMM	85.69
8	Samanta et al. [6]	2015	Words	ANGseg, HMM	84.85
9	Srimany et al. [7]	2014	Words	SVM	87.73
10	Samanta et al. [8]	2014	Words	HMM	89.65
11	Frinken et al. [9]	2014	Words	Bi-directional long shortterm (BLSTM) neural network	55.05 (Complex words)
12	Frinken et al. [9]	2014	Words	BLSTM neural network	89, 85.9 (Lexicon driven word)

					recognition)
13	Surinta et al. [10]	2013	Digits	Contour angular technique (CAT), SVM	95.9, 40% train data
14	Surinta et al. [10]	2013	Digits	Hotspot technique (HOT), SVM	92.7, 40% train data
15	Surinta et al. [10]	2013	Digits	Gray pixel-based method (GPB), SVM	95.9, 40% train data
16	Surinta et al. [10]	2013	Digits	Black and white down scaled method (BWS), SVM	94.1, 40% train data
17	Chowdhury et al. [11]	2013	Characters	Levenshtein distance metric	86.24
18	Bhattacharya et al. [12]	2013	Words	Rule based segmentation, SVM	97.45
19	Bhattacharya et al. [13]	2012	Words	Rule based segmentation, SVM	97.68 (Stroke level accuracy)
20	Mondal et al. [14]	2010	Characters	Point-float feature, HMM	81.55
21	Mondal et al. [14]	2010	Characters	Point-float feature, MLP	84.40
22	Mondal et al. [14]	2010	Characters	Chain-code feature, HMM	90.72
23	Mondal et al. [14]	2010	Characters	Chain-code feature, MLP	89.82
24	Fink et al. [15]	2010	Words	HMM	87.70
25	Bhattacharya et al. [16]	2008	Words	MQDF	82.34
26	Parui et al. [17]	2008	Characters	HMM, look-up tables	87.7 (character level accuracy), 84.6 (Stroke level accuracy)
27	Bhattacharya et al. [18]	2007	Characters	Direction code histogram features, MLP	93.90 and 83.61
28	Roy et al. [19]	2007	Characters	Quadratic classifier	91.13

a) Online handwritten Bangla characters and digits recognition

In 2007, Roy et al. [19] presented a novel scheme for online Bangla handwriting recognition, where they tested their system on 2500 Bangla numeral data and 12500 Bangla character data and attained an accuracy of 98.42% and 91.13%, respectively. In the same year, Bhattacharya et al. [18] described a novel technique to recognize online handwritten basic Bangla characters. In their work, they used a dataset of 7043 online handwritten Bangla characters. In 2008, Parui et al. [17] proposed a novel approach to recognize online handwritten Bangla characters using HMM. They recognized Bangla characters in two stages. In 2013, Chowdhury et al. [11] recognized online handwritten Bangla characters using Levenshtein distance metric, where they considered both direction and shape based characteristics to extract features. The work to compare feature and pixel-based methods for recognizing handwritten Bangla digits has been done by Surinta et al. [10] in the same year. In 2015, Ghosh and Roy did a novel work for feature extraction work for online handwritten Bangla characters.

b) Online handwritten Bangla words recognition

In 2008, Bhattacharya et al. [16] proposed an analytic approach to recognize online handwritten Bangla cursive words, where ten thousands online handwritten word samples provided by Bengali writers of different groups with respect to age, education, sex and income have been used. Later, in 2010,

Fink et al. [15] proposed an approach to recognize cursive Bangla words using HMM, where they used a sub-stroke level feature representation of the script. In two successive years (2012-13), Bhattacharya et al. [12][13] employed rule based

segmentation for online handwritten Bangla words, where they used SVM as a classification technique in their work. In 2014, Frinken et al. [9] used improved BLSTM neural networks for recognition of online handwritten complex Bangla words, where they investigated various encoding schemes of Bangla compound characters and compared the recognition results. In 2014, Samanta et al. [8] implemented different segmentation schemes and explored that the performance of HMM-based recognition is not dependent on segmentation strategy or scripts. In the same year, Srimany et al. [7] presented a study of a data driven approach for Bangla words and combined multiple SVM classifiers with RBF kernels for recognition. In 2015, Samanta et al. [6] proposed a script independent technique for online handwritten words recognition, where they used Bangla script also in their work. In 2016, Ghosh and Roy [4] employed ZSDP approach to recognize online handwritten Bangla words, where they used HMM as a classification technique in their work. Recently, in two successive years (2016-17), Bhattacharya et al. [2][3] developed a system for online Bangla handwriting recognition and proposed a solution to remove noise from online handwritten Bangla words, where they employed HMM and SVM classification techniques.

3. Conclusion

In the present study, we have presented the work done for online handwriting in Bangla script in recent past. The study has been carried out for smaller units as well as larger units. It has been noticed that the work for smaller units has been done for strokes, characters and digits. In the literature, the work for larger units has been done for online handwritten Bangla words. The present work is a guideline to future researchers to carry out further research in the area of online Bangla handwriting recognition. The presented study investigated that there is the need to work more for online Bangla handwriting

recognition for larger units as sentences or words. For online

handwritten Bangla sentence recognition, a lotto be done yet.

References

1. B. B. Chaudhuri and U. Pal, "An OCR system to read two Indian language scripts: Bangla and Devnagari (Hindi)", In: Proceedings of 4th International Conference on Document Analysis and Recognition, 1997.
2. Nilanjana Bhattacharya, Umapada Pal, and ParthaPratim Roy. "Cleaning of Online Bangla Free-form Handwritten Text", ACM Trans. Asian Low-Resour. Lang. Inf. Process, 17, 1, Article 8, (Nov 2017), 25 pages.
3. S. Bhattacharya, D. S. Maitra, U. Bhattacharya, and S. K. Parui, "An End-to-End System for Bangla Online Handwriting Recognition", (October 2016), 373–378.
4. R. Ghosh and P. P. Roy, "Comparison of Zone-Features for Online Bengali and Devanagari Word Recognition Using HMM", (October 2016), 435–440.
5. R. Ghosh and P. P. Roy, "Study of two zone-based features for online Bengali and Devanagari character Recognition", (August 2015), 401–405.
6. O. Samanta, A. Roy, U. Bhattacharya, and S. K. Parui, "Script independent online handwriting recognition", (August 2015), 1251–1255.
7. A. Srimany, S. D. Chowdhuri, U. Bhattacharya, and S. K. Parui, "Holistic recognition of online handwritten words based on an ensemble of svm classifiers", In Document Analysis Systems (DAS), 11th IAPR International Workshop, IEEE, 2014, 86–90.
8. O. Samanta, U. Bhattacharya, and S. K. Parui, "Smoothing of HMM parameters for efficient recognition of online handwriting", Pattern Recognition, 47, 11, 2014, 3614 – 3629.
9. VolkmarFrinken, Nilanjana Bhattacharya, Seiichi Uchida, and Umapada Pal, "Improved BLSTM Neural Networks for Recognition of On-Line Bangla Complex Words", In Structural, Syntactic, and Statistical Pattern Recognition, PasiFränti, Gavin Brown, Marco Loog, Francisco Escolano, and Marcello Pelillo(Eds.), Springer Berlin Heidelberg, Berlin, Heidelberg, 404–413.
10. O. Surinta, L. Schomaker, and M. Wiering, "A Comparison of Feature and Pixel-based Methods for Recognizing Handwritten Bangla Digits", In: Proceedings of International Conference on Document Analysis and Recognition, 2013, 165–169.
11. 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, 2013, 79–83.
12. N. Bhattacharya, U. Pal, and F. Kimura, "A System for Bangla Online Handwritten Text", (August 2013), 1335–1339.
13. N. Bhattacharya and U. Pal, "Stroke Segmentation and Recognition from Bangla Online Handwritten Text", (September 2012), 740–745.
14. 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, 2010, 200–205.
15. G. A. Fink, S. Vajda, U. Bhattacharya, S. K. Parui, and B. B. Chaudhuri, "Online Bangla word recognition using sub-stroke level features and hidden Markov models", In: Proceeding of the 12th International Conference on Frontiers in Handwriting Recognition, 2010, 393–398.
16. U. Bhattacharya, A. Nigam, Y. S. Rawat, and S. K. Parui, "An Analytic Scheme for Online Handwritten Bangla Cursive Word Recognition", 2008.
17. S. K. Parui, K. Guin, U. Bhattacharya, and B. B. Chaudhuri, "Online handwritten Bangla character recognition using HMM", Proceedings of 19th International Conference on Pattern Recognition, 2008, 1–4.
18. U. Bhattacharya, B. K. Gupta, and S. K. Parui, "Direction Code Based Features for Recognition of Online Handwritten Characters of Bangla", In: Proceedings of the 9th International Conference on Document Analysis and Recognition, 2007, 58–62.
19. K. Roy, N. Sharma, T. Pal, and U. Pal, "Online Bangla Handwriting Recognition System", Proceedings of International Conference on Advances in Pattern Recognition, 2007, 117–122.