e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:05/Issue:10/October-2023 Impact Factor- 7.868 www.irjmets.com

A REVIEW OF ONLINE HANDWRITTEN CHARACTER AND WORD RECOGNITION FOR INDIAN LANGUAGES

Dr. Vishal Naik*1, Heli Mehta*2

*1Asst. Prof., Udhna College, Surat, Gujarat, India.

*2Asst. Prof., RNGPIT, Bardoli, Gujarat, India.

DOI : https://www.doi.org/10.56726/IRJMETS45209

ABSTRACT

In this paper, we present a review for online handwritten character recognition by comparing various pre-processor, feature extraction methods and algorithms for Indian languages, language wise performance of the system is also compared. A review for online handwritten word recognition is also presented with various algorithms and post-processing methods for Indian languages. Performance of various systems and languages are described like Gujarati with 91.5%, Bangla with 97.89%, Gurmukhi with 99.3%, Assamese with 95.07% and Hindi with 90.78% accuracy .

Keywords: OHWR, HCR, OCR, SVM, ANN, CNN, Indian Language.

I. INTRODUCTION

Optical character recognition system can be used to identify the handwritten characters. Handwritten character recognition (HCR) can be classified into two types - offline and online.

In an online handwritten character recognition system, the system should work like a real-time system. The user writes a character in a computer using a pen or stylus with the help of some tool. The sequence of coordinate pixels is considered as a written character. Different pre-processing algorithms are applied to a written character for removing noise, size variation, etc. A preprocessed character is used to extract meaningful features from it. Extracted features are used as an input to the classification algorithm. The classifier returns a matching character according to the training data. A classified character can further be processed using post-processing methods to improve accuracy and association of characters.

Indian languages have large and complex character set compare to English and other Latin scripts. Indian scripts include constants, vowels and composite characters representing a combination of constants and vowels. There is a similarity between characters of different Indian languages, mainly based on a geographic location of languages used. Many characters in different Indian languages requires multiple strokes to write. Such a complex character set makes traditional keyboard not practical for Indian languages. Most of the Indian languages have major differences among each other and due to that, there cannot be a single handwritten character recognition system for all the Indian languages. We need to develop separate systems for every Indian language.

Online handwritten character recognition system is used to take input from the user as a handwritten character in the computer system. The user can use digital pen, stylus, touch screen device, or mouse to write a character.

Using an online handwritten character recognition system, a person can write in a traditional Indian language with the help of a mouse or a pen or a stylus or any such device. A person can write a character as easily as he is writing using a pen and paper. When a user writes a character using a pen, active pixel values are recorded.

Recorded pixel values are used for further processing of feature extraction and recognition. This system requires specific training data which can be generated from different types of users and different types of devices. This system should be high performance and should recognize characters in real time.

II. LITERATURE REVIEW

The online character recognition system has training data of all character classes of a particular language. Online character recognition system should be fast and accurate because the user is going to use the system in real-time. We have compiled and compared different work of researchers of major Indian languages.

The following tables 1 to 7 shows and compare classifier, feature set and accuracy of online handwritten character recognition of Assamese, Tamil, Devanagari, Malayalam, Gurmukhi, Bangla and Gujarati Indian languages.

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:05/Issue:10/October-2023 Impact Factor- 7.868 www.irjmets.com

Table 1. Comparison of character recognition system for Assamese

Sr. No.	Language	Classifier	Features	Accuracy	Author(s)
1	Assamese	SVM	Posterior feature	99.52%	S. Mandal [1]
2	Assamese	Combined HMM & SVM	Coordinate sequence 1st & 2nd order derivative	96.17%	H. Choudhury [2]
3	Assamese	HMM SVM	1st & 2nd order derivative Baseline features	95.10%	S. Mandal [3]
4	Assamese	HMM	Pixel coordinates	93.35%	H. Choudhury [4]
5	Assamese	HMM SVM	Statistical Directional	76.24% 76.56%	S. Mandal [5]

Table 2. Comparison of character recognition system for Tamil

Sr. No.	Language	Classifier	Features	Accuracy	Author(s)
1	Tamil	Naïve Bayes	Pixel coordinates	91.81%	R. Kunwar [6]
2	Tamil	HMM BOS	Writing direction Curvature Slope	91.80%	A. Bharath [7]
3	Tamil	Connected component	Blobs Stems features	77.84%	K. H. Aparna [8]

Table 3. Comparison of character recognition system for Devanagari

Sr. No.	Language	Classifier	Features	Accuracy	Author(s)
1	Devanagari	SVM HMM	Pixel coordinates	97.27%	H. Swetha lakshmi [9]
2	Devanagari	HMM SVM	Zone wise slope of dominant points	93.3% 97.11%	R. Ghosh [10]
3	Devanagari	Template matching	DTW	97%	K. C. Santosh [11]
4	Devanagari	SVM	Structural Zone wise directional Zone wise slope	90.63%	R. Ghosh [12]
5	Devanagari	HMM BOS	Writing direction Curvature Slope	87.13%	A. Bharath [13]

Table 4. Comparison of character recognition system for Malayalam

Sr. No.	Language	Classifier	Features	Accuracy	Author(s)
1	Malayalam	k-NN	Pixel coordinates Direction Curvature Aspect ratio	98.12%	M. Sreeraj [14]
2	Malayalam	HMM SVM	Pixel coordinates Direction Curvature	97.97%	K.P. Prime kumar [15]

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:05/Issue:10/October-2023 Impact Factor- 7.868 www.irjmets.com

			Angular features		
3	Malayalam	SVM DDAG	Pixel coordinates Direction Curvature Moments	95.78%	A. Arora [16]
4	Malayalam	k-NN MLP SVM	Accurate dominant points Intersections	90.39% 93.17% 95.12%	Baiju KB [17]

Table 5. Comparison of character recognition system for Gurmukhi

Sr. No.	Language	Classifier	Features	Accuracy	Author(s)
1	Gurmukhi	SVM	X & Y projection	99.75%	H. Singh [18]
2	Gurmukhi	SVM HMM	RDP Chain code	98.21% 98.27%	S. Singh [19]
3	Gurmukhi	K-means clustering	Direction Loops	94.69%	A. Sharma [20]
4	Gurmukhi	k-NN MLP SVM	Spatial temporal Spectral features	89.35% 89.89% 89.64%	R. Kaur [21]
5	Gurmukhi	HMM	Zoning features	88.40%	K. Verma [22]

Table 6. Comparison of character recognition system for Bangla

Sr. No.	Language	Classifier	Features	Accuracy	Author(s)
1	Bangla	CNN	Pooling	99.40%	S. Sen [23]
2	Bangla	SMO	Mass distribution Chord length krill-herd	98.57%	S. Sen [24]
3	Bangla	SVM	COG based global & local	98.26%	S. Sen [25]
4	Bangla	MLP	Hausdorff Distance Directed HD	95.57%	S. Sen [26]
5	Bangla	SVM	Transition counts, centre of gravity, & topological	95.49%	S. Sen [27]

Table 7. Comparison of character recognition system for Gujarati

Sr. No.	Language	Classifier	Features	Accuracy	Author(s)
1	Gujarati	SVM k-NN	derivative of pixel values, zoning, normalized chain code	94.65%	Vishal [28]
2	Gujarati	SVM	zoning features dominant point-based normalized chain code	94.13%	Vishal [29]
3	Gujarati	SVM	Structural	91.63%	Vishal [30]

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:05/Issue:10/October-2023 Impact Factor- 7.868 www.irjmets.com

		MLP k-NN	Statistical	86.72% 90.09%	
4	Gujarati	SVM	zoning and chain code directional features	95%	Vishal [31]

The handwritten word recognition systems have large classes set which includes characters, numerals, diacritics, and modifiers. These systems require one addition step after recognition of a stroke. This step is known as a post-processing which works on recognized characters. The researchers are working in Gujarati, Bangla, Gurmukhi, Gurmukhi and Hindi Indian languages but limited work is done.

Gujarati: Vishal et al. [32] have used a SVM with RBF kernel for classification. The hybrid feature set is used with directional and zoning features. Location based and mapping rule based post-processing method is used for further processing. They have achieved an accuracy of 95.3% for individual characters , 91.5% for individual words, and 83.3% for sentences.

Bangla: K. Chowdhury et al.[33] have used different geometric values as fuzzy features. The system had three-stages which are grouping of segments, compare features with a database, and aggregation of recognized characters. The result showed an accuracy of 77%.

N. Bhattacharya et al.[34] have used segmentation using the offline horizontal histogram. The feature set included chain code directions features. Support Vector Machine (SVM) was used for classification. The result showed an accuracy of 97.89%.

S. Sen et al. [35] have used multi-stroke character segmentation method was proposed which uses stroke-level busy zone for segmentation. The feature set had 136 feature values which include global and local features. The system was tested using MLP, SVM, and Bayes network classifiers. The result showed an accuracy of 95.33% using SVM-polynomial, 94.95% using SVM-linear, 92.95% using SVM-Gaussian, 92.65% using MLP, and 86.63% using Bayes network classifier.

N. Bhattacharya et al. [36] have used skew correction, zone segmentation, word size normalization, word slicing, and ordering procedures are used in pre-processing. The feature set included N-Pen++ features, vertical position, writing direction, pen up/down, ascenders, descenders, and Context map features. The HMM classifier was used for classification. The result showed an accuracy of 90.53%.

Gurmukhi: N. Kumar and S. Gupta[37] have used Deep Artificial Neural Network was used for classification. The feature set included the Local Binary Pattern (LBP), directional, and regional features. Classification was performed using simple mapping method. The result showed an accuracy of 99.3%.

Assamese: S. Mandal et al. [38] have presented work on smoothing, resampling, and size normalization methods were used in pre-processing. The Hidden Markov Model was used for classification. Top 2 results of the HMM classifier was checked and if it belonged to confusing pair then its score was calculated based on discriminative states. The feature set had 16 features which included ascender, descender, local and directional features. The result showed an accuracy of 93.69% using HMM for character and 95.07% using the proposed method for word recognition.

Hindi: S. Belhe et al. [39] have presented work on smoothing using a Gaussian filter and normalization methods were used in pre-processing. Histogram of Oriented Gradients (HOG) features was used in the feature set. Gaussian kernel Hidden Markov Model (HMM) was used for classification. A stroke group-based tree structure was used in post-processing. They achieved an accuracy of 89%.

S. Malakar et al. [40] have used feature set with 89 feature values which include centroid, projection, area, aspect ratio, pixel ratio, pixel density, longest run, projection length, and other local features. The system was tested on different classifiers. The result showed an accuracy of 90.78% using MLP, 90.76% using SMO, 87.75% using LRM, 72.94% using Naïve Bayes, and 82.77% using the multi-class classifier.

III. RESULTS AND DISCUSSION

The comparison of various classifier, feature set and accuracy of online handwritten character recognition for various Indian languages. For online handwritten character recognition, following are the best result achieved by the researcher for various Indian languages. For Assamese, the result showed an accuracy of 97.67% for upper letters & processing time of 162.34 milliseconds, and 96.05% for lower letters & processing time of

335.56 milliseconds.[1] For Tamil, the result showed an accuracy of 91.81%. [6] For Devanagari, the result showed an accuracy of 97.27% using SVM and 83.08% using HMM. [9] For Malayalam, the result showed an accuracy of 98.12%. [14] For Gurmukhi, the result showed an accuracy of 94.8% for character and 99.75% for zone identification. [18] For Bangla, the result showed an accuracy of 99.40% using max pooling and softmax function. [23] For Gujarati, the result showed an accuracy of 94.65% using two-layer classification using SVM and k-NN. [28].

For online handwritten word recognition, following are the best result achieved by the researcher for various Indian languages. For Gujarati, the result showed an accuracy of 95.3% for individual characters , 91.5% for individual words, and 83.3% for sentences. [32] For Bangla, The result showed an accuracy of 97.89% using SVM. [34] For Gurmukhi, The result showed an accuracy of 99.3% using DANN. [37] For Assamese, The result showed an accuracy of 95.07% using HMM. [38] For Hindi, The result showed an accuracy of 90.78% using MLP, 90.76% using SMO, 87.75% using LRM, 72.94% using Naïve Bayes, and 82.77% using the multi-class classifier. [40].

IV. CONCLUSION

An online handwritten character recognition system can be used by a normal person to write in a traditional Indian languages. A normal mouse or a pen or a stylus or any such device can be used for better and fast communication with the system. Comparison of classifier, feature set and accuracy of online handwritten character recognition of Gujarati, Assamese, Tamil, Devanagari, Malayalam, Gurmukhi, and Bangla Indian languages. For online handwritten character recognition, Gujarati, Assamese, Devanagari, Gurmukhi and Bangla language performed best among all Indian languages with SVM, ANN and CNN learning algorithms with variety of feature set according to the language.

For online word recognition system, good work has been done for Gujarati, Bangla, Gurmukhi, Assamese and Hindi languages with the help of single and multi-layer classification algorithms, variety of feature set and various post-processing methods. Such an online word recognition system can provide a normal person very efficient writing and communication facility with computer system.

V. REFERENCES

[1] S. Mandal, S. R. M. Prasanna, and S. Sundaram, “GMM posterior features for improving online handwriting recognition”, Expert Syst. Appl., vol. 97, pp. 421–433, 2018.

[2] H. Choudhury, S. Mandal, S. Devnath, S. R. M. Prasanna, and S. Sundaram, “Combining HMM and SVM based stroke classifiers for online Assamese handwritten character recognition”, in 12th IEEE International Conference Electronics, Energy, Environment, Communication, Computer, Control: (E3-C3), INDICON 2015, 2016, pp. 1–6.

[3] S. Mandal, H. Choudhury, S. R. M. Prasanna, and S. Sundaram, “Frequency count based two stage classification for online handwritten character recognition”, 2016 Int. Conf. Signal Process. Commun. SPCOM 2016, 2016.

[4] H. Choudhury, S. Mandal, S. Devnath, S. R. M. Prasanna, and S. Sundaram, “Comparison of assamese character recognizer using stroke level and character level engines”, 2015 21st Natl. Conf. Commun. NCC 2015, 2015.

[5] S. Mandal, S. R. M. Prasanna, and S. Sundaram, “An improved discriminative region selection methodology for online handwriting recognition”, Int. J. Doc. Anal. Recognit., 2018.

[6] R. Kunwar, U. Pal, and M. Blumenstein, “Semi-supervised Online Learning of Handwritten Characters Using a Bayesian Classifier”, 2013 2nd IAPR Asian Conf. Pattern Recognit., pp. 717–721, 2013.

[7] A. Bharath and S. Madhvanath, “HMM-based lexicon-driven and lexicon-free word recognition for online handwritten indic scripts”, IEEE Trans. Pattern Anal. Mach. Intell., vol. 34, no. 4, pp. 670–682, 2012.

[8] K. H. Aparna, V. Subramanian, M. Kasirajan, G. V. Prakash, and V. S. Chakravarthy, “Online Handwriting Recognition for Tamil”, Ninth Int. Work. Front. Handwrit. Recognit., pp. 438–443, 2004.

[9] H. Swethalakshmi, “Online handwritten character recognition of Devanagari and Telugu Characters using support vector machines”, Tenth Int. Work. Front. Handwrit. Recognit., pp. 1–6, 2006.

[10] R. Ghosh and P. P. Roy, “Study of Zone-Based Feature for Online Handwritten Signature Recognition

and Verification in Devanagari Script", in International Conference on Computer Vision and Image Processing, 2017, pp. 523–530.

[11] K. C. Santosh and E. Iwata, "Stroke-Based Cursive Character Recognition", Adv. Character Recognit., 2012.

[12] R. Ghosh and P. P. Roy, "Study of two zone-based features for online Bengali and Devanagari character recognition", Proc. Int. Conf. Doc. Anal. Recognition, ICDAR, vol. 2015–Novem, pp. 401–405, 2015.

[13] A. Bharath and S. Madhvanath, "HMM-based lexicon-driven and lexicon-free word recognition for online handwritten indic scripts", IEEE Trans. Pattern Anal. Mach. Intell., vol. 34, no. 4, pp. 670–682, 2012.

[14] M. Sreeraj and S. Idicula, "k-NN Based On-Line Handwritten Character Recognition System", Integr. Intell. Comput. (ICIIC), 2010 First Int. Conf., pp. 171–176, 2010.

[15] K. P. Primekumar and S. M. Idiculla, "On-line Malayalam Handwritten Character Recognition using HMM and SVM", Int. Conf. Signal Process., Image Process. Pattern Recognit. [ICSIPR], pp. 1–5, 2013.

[16] A. Arora and A. M. Namboodiri, "A hybrid model for recognition of online handwriting in Indian scripts", in International Conference on Frontiers in Handwriting Recognition, ICFHR 2010, 2010, pp. 433–438.

[17] K. Baiju and S. K, "Online Recognition of Malayalam Handwritten Scripts - A comparison using KNN, MLP and SVM", Intl. Conf. Adv. Comput. Commun. Informatics, no. Sept. 21-24, 2016, pp. 1–5, 2016.

[18] H. Singh, R. K. Sharma, and V. P. Singh, "Efficient zone identification approach for the recognition of online handwritten Gurmukhi script", Neural Comput. Appl., vol. 3456789, 2018.

[19] S. Singh, A. Sharma, and I. Chhabra, "A dominant points-based feature extraction approach to recognize online handwritten strokes", Int. J. Doc. Anal. Recognit., vol. 20, no. 1, pp. 37–58, 2017.

[20] A. Sharma and K. Dahiya, "Online Handwriting Recognition of Gurmukhi and Devanagari Characters in Mobile Phone Devices", in International Journal of Computer Applications, 2012, pp. 37–41.

[21] R. Kaur and M. Singh, "Stroke based online handwritten Gurmukhi character recognition", 2016 Int. Conf. Adv. Comput. Commun. Informatics, ICACCI 2016, pp. 598–601, 2016.

[22] K. Verma and R. K. Sharma, "An Efficient Writing-Zone Identification Technique for Online Handwritten Gurmukhi Character Recognition", Proc. Natl. Acad. Sci. India Sect. A Phys. Sci., 2017.

[23] S. Sen, D. Shaoo, S. Paul, R. Sarkar, and K. Roy, "Online Handwritten Bangla Character Recognition Using CNN : A Deep Learning Approach", Intell. Eng. Informatics. Adv. Intell. Syst. Comput., vol. 695, pp. 413–420, 2018.

[24] S. Sen, M. Mitra, A. Bhattacharyya, R. Sarkar, F. Schwenker, and K. Roy, "Feature Selection for Recognition of Online Handwritten Bangla Characters", Neural Process. Lett., 2019.

[25] S. Sen, A. Bhattacharyya, A. Das, R. Sarkar, and K. Roy, "Design of Novel Feature Vector for Recognition of Online Handwritten Bangla Basic Characters", in the First International Conference on Intelligent Computing and Communication. Advances in Intelligent Systems and Computing, 2016, vol. 458, pp. 485–494.

[26] S. Sen, R. Sarkar, K. Roy, and N. Hori, "Recognize Online Handwritten Bangla Characters Using Hausdorff Distance-Based Feature", in the 5th International Conference on Frontiers in Intelligent Computing: Theory and Applications. Advances in Intelligent Systems and Computing, 2017, vol. 515, pp. 541–549.

[27] S. Sen, A. Bhattacharyya, P. K. Singh, R. A. M. Sarkar, K. Roy, and D. Doermann, "Application of Structural and Topological Features to Recognize Online Handwritten Bangla Characters", ACM Trans. Asian Low-Resour. Lang. Inf. Process., vol. 17, no. 3, 2018.

[28] Vishal Naik and Apurva Desai, "Online Handwritten Gujarati Character Recognition: Two-Layer Classification Approach", International Journal of Research in Advent Technology, Vol.7, No.6, June 2019 DOI: 10.32622/ijrat.762019128

[29] Vishal Naik and Apurva Desai, "Multi-layer Classification Approach for Online Handwritten Gujarati Character Recognition", Computational Intelligence: Theories, Applications and Future Directions—Volume II, Advances in Intelligent Systems and Computing, pp 595-606,2019 https://doi.org/10.1007/978-981-13-1135-2_45

[30] Vishal Naik and Apurva Desai, "Online handwritten Gujarati character recognition using SVM, MLP, and K-NN", 8th International Conference on Computing, Communication and Networking Technologies (ICCCNT), 2017 DOI: 10.1109/ICCCNT.2017.8203926

[31] Vishal Naik and Apurva Desai, "Online Handwritten Gujarati Numeral Recognition Using Support Vector Machine", International Journal of Computer Sciences and Engineering, Vol.-6, Issue-9, Sept. 2018 DOI: 10.26438/ijcse/v6i9.416421

[32] Vishal Naik and Apurva Desai, "Online Handwritten Gujarati Word Recognition", International Journal of Computer Vision and Image Processing, Vol-9, Issue-1, 2019 DOI: 10.4018/IJCVIP.2019010103

[33] K. Chowdhury, L. Alam, S. Sarmin, S. Arefin, and M. M. Hoque, "A fuzzy features based online handwritten Bangla word recognition framework", 2015 18th Int. Conf. Comput. Inf. Technol. ICCIT 2015, pp. 484–489, 2016.

[34] N. Bhattacharya, U. Pal, and F. Kimura, "A System for Bangla Online Handwritten Text", in 12th International Conference on Document Analysis and Recognition, 2013, pp. 1335–1339.

[35] S. Sen, S. Chowdhury, M. Mitra, F. Schwenker, R. Sarkar, and K. Roy, "A novel segmentation technique for online handwritten Bangla words", Pattern Recognit. Lett., 2018.

[36] N. Bhattacharya, U. Pal, and P. P. Roy, "Stroke-Order Normalization for Online Bangla Handwriting Recognition", in 14th IAPR International Conference on Document Analysis and Recognition, 2017.

[37] N. Kumar and S. Gupta, "Offline Handwritten Gurmukhi Word Recognition Using Deep Neural Networks", Int. J. Pure Appl. Math., vol. 119, no. 12, pp. 14749–14767, 2018.

[38] S. Mandal, H. Choudhury, S. R. M. Prasanna, and S. Sundaram, "Exploring Discriminative HMM States for Improved Recognition of Online Handwriting", in 24th International Conference on Pattern Recognition (ICPR), 2018, pp. 3753–3758.

[39] S. Belhe, C. Paulzagade, A. Deshmukh, S. Jetley, and K. Mehrotra, "Hindi handwritten word recognition using HMM and symbol tree", Proceeding Work. Doc. Anal. Recognit. - DAR '12, p. 9, 2012.

[40] S. Malakar, P. Sharma, P. K. Singh, M. Das, R. Sarkar, and M. Nasipuri, "A Holistic Approach for Handwritten Hindi Word Recognition", Int. J. Comput. Vis. Image Process., vol. 7, no. 1, pp. 59–78, 2017.