IARJSET

ISSN (Online) 2393-8021
ISSN (Print) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

ISO 3297:2007 Certified

Vol. 3, Issue 10, October 2016

Review of Feature Extraction and Classification Techniques for OHCR in Indian Scripts

Sabeerath .K[1], Baiju K B[2]

Asst Professor, Dept of Computer Science, JDT Islam College of Arts and Science, Vellimadukunnu, Calicut, India [1]

Assistant Professor, Department of Computer Science, NMSM Govt. College Kalpetta, India [2]

Abstract: The development of OHCR (Online Handwritten Character Recognition) is an interesting area in the field of Pattern Recognition. Sustainability of traditional script input methods needs attention in the preservation of neutrality of languages. In the context of Indian scripts, automatic recognition of handwritings contributes a lot in preserving its vital aspects. The South Indian family of languages (Malayalam, Tamil, Kannada, and Telugu) share characteristics of ancient Brahmi script from which they are derived. Research work for accurate OHCR engines for various Indian scripts is still in progress. The paper is intended for familiarizing the works in Online Handwriting Recognition of various Indian scripts. Various features and classification techniques used with recognition accuracies were compared from existing literature. This will be beneficial for the quality selection of features and classification techniques in future works.

Keywords: OHCR, Feature Extraction, Indian scripts, Handwriting Recognition

I. INTRODUCTION

Methods for annihilating the gap in humanemachine coalition has decades of sophisticated studies. With the invention of computers the promptness for a method to input diverse scripts span around the globe has been a necessity for the entire community apart from the scientific realm. In order to cop up with the levels of the society, the preservation of traditional script input methods obtained relevance in machine interaction.

Inputting script in one's own language is a difficulty associated with data processing machines. This will be a weighty task in India, a multi script-language country. An easy way to input regional languages in a traditional way as inputs for data processing is desirable.This lead to the studies for natural interface for data input to the machine.

The force thus made Online Handwriting Recognition a frontier area of research in computer science. In OHCR, traditional handwriting input is linked with modern technologies to preserve the natural script input style. Works are still in progress for high recognition OHCR engines for Indian languages.

The rest of the paper is arranged as Online Handwriting Recognition, Indian languages, Features and Classification methods used in OHCR of various Indian scripts and finally the conclusion and future works are mentioned.

II. ONLINE HCR

Online Handwriting Recognition involves writing on digital touch pads using stylus, writing in paper using digital pen or writing in touch screen displays using fingersand automatically recognized by an OHCR engine. The strokes corresponding to eachcharacter includes coordinate points along the path, pen up, pen down information, time sequence and structural information [1]. The specialty of online handwritingrecognition is the preservation of traditional styles followed in the past era. Eventhough the writer is ignorant of technologies, he can enter the scripts in the dataprocessing machine.

Online handwriting replace keyboards, which is the commonly used input device for data entry. The complexity of using various scripts [2] in keyboards canbe eliminated through online handwriting recognition. Nowadays online recognitionis a popular method in mobile devices, which shows the wide admissibility of thetechnique.

A. Steps in OHCR

In [3] Charles C.Tappert , Suen, C. Y. and Wakahara, Tdescribes the state of the art of Online Handwriting as an active research area for decades to come.OnlineHCR isa sequential process comprising Data acquisition,Pre-processing, Feature extraction, Classification, Post processing. Data acquisition involves obtaining pen trajectories of the writings. The data is acquired as a sequence of points with time and device specific values. The acquired points were pre-processed for removing noises and structural variations pertained to user writing styles.

The data acquisition is a complex task in Online Handwriting and certain issues pertaining to them is to be addressed Data collected from the users may be affected with lot of noises and must be cleansed before analysis. Pre-processing varies in various scripts and usually

IARJSET
ISSN (Online) 2393-8021
ISSN (Print) 2394-1588
International Advanced Research Journal in Science, Engineering and Technology
ISO 3297:2007 Certified
Vol. 3, Issue 10, October 2016

includes duplicate point elimination, de-hooking, smoothening, resampling and normalization [4]. The features are the essential structural and statistical properties which describe the character and the features extracted from the pre-processed co-ordinate values are used in the classifiers. Quality selection of feature affects classification rate [5], hence the features must be selected based on the classification technique to be employed. Commonly used features are coordinate positions, bumps, crossings, cusps, loops, writing directions, angles and aspect ratios [6].

Classification includes the problem of recognizing to which of the available classes a new class member fit. Commonly used classifiers areHidden Markov Models (HMM), Support Vector Machines (SVM), K Nearest Neighbour Classifier (k-NN), Neural Networks(NN), Dynamic Time Warping (DTW), Fourier Descriptors and Daubachies Wavelet Transform (DWT) [7].Post processing is done after classification for the alphanumeric code matching of characters and any dis-ambiguity in confusing sets and applying language specific constructs to improve the recognition phase.

III.OHCR IN INDIAN SCRIPTS

In the following section, we present a review on various techniques proposed in the literature to recognize online Indian scripts. we outline the contributed works for major Indian scripts as follows.

A. Devanagari.

One of the earliest work in Devanagari OHCR has been presented by Scott-D- Connell et al in 2000[8].They have worked with both online and offline HCR using five different classifiers. By using five classifiers they have achieved the average accuracy of 86.5%.A.M Namboodiri et al used 11 spatial and temporal features extracted from the strokes of the words for recognizing online Devanagariscript[9]. They have attained an accuracy of 87.1% on a database of 13,379 words.Joshi et al developed a test vector is based the structural features, such as mean (x, y) values, positional cues and directional codes at the stroke level[10]. A subspace method is used by the system for classification in which each character class is represented by a basis vector which is a set of N eigenvalues. The system has achieved the accuracy of 93.05% on a set of 100 frequently occurring characters.

Abhimanyu and SamitBattacharya proposed a stroke based recognition using Hidden Markov Model[11] .They focussed on recognition of isolated Devanagari characters in the iphone. In another work Swethalakhsmi proposed a two stage proximityanalysis technique for the identification of Devanagari characters from a sequence ofstrokes. Strokes are represented based on spatiotemporal, spectral and spatiostructural features are explored. They have achieved the average accuracy of 95.29%.Lajish V.L and Sunil Kumar presented an online handwritten Devanagari scriptrecognition [12]. They introduced Fuzzy Directional Features (FDF) which can produce directional variance in the handwritten primitives. In another work they introduced a new online feature set called the extended directional feature which candiscriminate similar looking strokes[13].

B. Bangla

U Garain et al proposed a work in bangla character recognition which is focussed on utilizing the cues from the pen trajectory to derive features when tackling the problem of stroke order variation[14]. Directional code feature has been introduced in a work proposed by Bhattacharya et al for recognition of Bangla OHCR [15].MLP has been used as the classifier. A string of features are derived at the sub stroke level by Nakai et al [16]. Based on the similar shape of the graphemes that constitute theideal character shape. The classification algorithm used is HMM.

T Mondal et al presented a comparative study of the performance of the classifiers HMMand a nearest neighbour classifier (based on DTW) [17]. An analytic recognition approach depends on the position of the headline, is adopted by Bhattacharya et al tosegment the input word to a set of substrokes[18].

C. Malayalam

As one of the initial work in Malayalam ,Sreeraj M., and Idicula S proposed an OHCR system with normalised (x, y) coordinates and context bit map features[19].The classification technique used is kohonen networks. The system exhibited a recognition accuracy 88.75% with a recognition time of 15-32 milliseconds. A combination of time domain features and dynamic representation of writing direction along with its curvature is proposed by SreerajM., and Idicula S in another work [20]. The featuresused are writing directions, normalised (x, y) coordinates and aspect ratio of curvature. By using KNN classifier, the system achieved an accuracy of 98.125% even witha small sample size.

A recognition engine using wavelet transform is presented by Primekumar K. P. and Idiculla S. M.[21] .Time domain features of the strokes are used with wavelettransform. The classification algorithm used is simplied fuzzy ARTMAP network, which requires comparatively very less time for training. An accuracy of 97.81% isreported in the work.

Indhu, T. R., and Bhadran, V. K presented an online handwriting recognitionsystem using simplified fuzzy ARTMAP technique[22]. Enormous features are usedin the work which includes,(x, y) coordinates, start quadrant, end quadrant, horizontal and vertical point density, loop, cusp, stroke length, The system achieved98.26% recognition accuracy with simplified fuzzy ARTMAP classification technique.Prime Kumar, K. P., and Idiculla, S. M. extended their work using HMM and SVM[69].They achieved an improved accuracy of 97.97%

IARJSET
ISSN (Online) 2393-8021
ISSN (Print) 2394-1588
International Advanced Research Journal in Science, Engineering and Technology
ISO 3297:2007 Certified
Vol. 3, Issue 10, October 2016

for SVM using Gaussian kernel. Butthe recognition rate dropped to 95.24% for HMM and recognition time was also increased compared to SVM.

Chacko, B. P.,andBabuAnto P achieved 96.83% recognition accuracy using OSELM and SLFN by division point features [24]. Sampath A., Tripti C., and GovindaruV presented a Neural Network based model for handwritten recognition [25]. The direction information of the written character is recorded based on the 8 connectedFreeman chain code. The direction of the pen movement is recorded as feature vector. Back propagation Neural Network is used for classifying characters. Additionaldisambiguation technique is used in post processing stage to identify confusing pairs.A recent work of Steffy Maria and Joseph achieved a recognition accuracy of 90%using SVM The directional and curvature features are extracted and trained in LIBSVM tool available in Matlab [26].

D. Tamil

Majority of the work in Dravidian scripts have been reported in Tamil. In the work,Niranjan Joshi., Sita G., Ramakrishnan A. G., and Madhvanath S demonstrated atemplate based elastic matching technique [27]. The features used are (x, y) coordinates, quantized slopes and dominant point coordinates, the system combineddifferent features and formed seven different schemes. The work describes recognitionaccuracy, recognition speed and number training templates with dynamic time warping distance measure. The system achieved a maximum of 95.90% accuracy with aspeed of 32.6 characters per second.

A novel approach has been described by Aparna K. H., Subramanian V., KasirajanM., Prakash G. V., and Chakravarthy V. S with a string of shape features[28]. Using this string representation, an unknown stroke is identified by comparing it with adatabase of strokes using a flexible string matching procedure and finite state automation. A recognition rate of 77.84% has been achieved by the system. Neural networkshas been adopted in Ishwarya and Kannan R.J with multilayer perceptron [29]. Thefeature used is Fourier descriptor. Test results indicate that Fourier descriptors withback propagation network provides a recognition accuracy of 97%. ShashikiranK.,Prasad K., Kunwar R., and Ramakrishnan, A. analysed the performance of TamilScripts with HMM and Statistical Dynamic Time Warping (SDTW)[39].The resultsshow that the best result of 85% for HMM is possible using a simpler SDTW model.

The work of Murthy and Venkatesh N involves two stage recognition with PCAand NN classifier adopted at the first stage followed by combined feature combination and DTW [30]. The system uses (x, y) coordinates, quantised slope, coordinatesof accurate dominant points and Quartile Features. Based on the primary classifier output and prior knowledge, a classifier is chosen for the second phase. Thesystem achieved 90.2% recognition accuracy for Tamil Scripts. RiturajKanwar andRamakrishnan.A.G has projected a fractal coding method with features as fractaldimensions[31]. This technique exploits the redundancy in data, thereby achievingbetter compression and usage of lesser memory. The fractal code features were appliedin separate classifiers (HMM, SVM, SDTW). The recognition achieved an accuracyof 90% with SDTW.

E. Kannada

Wavelet features of the Kannada script is used in the recognition method proposed by Samuel R. D. S.,SrinivasaRaoKunte [32]. The system achieved a recognition accuracy of 95% for basic Kannada characters. The structural and spatio-temporal features are employed in a work (Prasad M. M., SukumarM.,andRamakrishnana.G.[33]) in the recognition of Kannada Scripts. The strategy used here is Divide and Conquer. A character is segmented into three strokes units and recognized separately.The method is proposed to overcome the complexity of huge number compound characters. The segmented units are preprocessed and the extracted features are mapped to sub-space using PCA. The system achieved an accuracy of 81% with KNN classifier.

KunwarR.,Mohan P., Shashikiran K., &Ramakrishnan a. G [34] presented an unrestricted recognizer using a smoothed first derivatives of (x, y) coordinates andSDTW as the classifier. The system achieved 88% of recognition accuracy with fasterrecognition time compared to the recognition time of DTW.

A recognizer for mobile devices has been suggested by Keerthi Prasad, ImranKhan and Naveen R Chanukotimath [35] using two approaches PCA and DTW. Thereported recognition accuracy is 88% for the PCA and 64% for DTW. The recognitiontime of PCA approach is 0.8 seconds which is fairly better compared to DTW witha recognition time of 55 seconds. Venkatesh N and A G Ramakrishnan proposed anovel method with two stage classifiers[30]. The features used are quartile features,preprocessed (x, y) coordinates, quantized slopes and accurate dominant points. Theprimary classifier used is DTW. Based on the result of primary classifier and priorknowledge, secondary classifier is applied to obtain a recognition rate of 92.6% forKannada characters.

F. Thelugu

H. Swethalakshmi developed a system for online HCR of Telugu and Devanagari writing using Support Vector Machines (SVMs) [36]. Each stroke is represented as an n-dimensional feature vector depending on the choice of the number of points for stroke representation. The features chosen to represent the curve are the coordinates of points in the pre-processed stroke. The stroke is then classified using SVM with 73.30% of accuracy. Babu V. J.,PrasanthL.,Sharma R.R., and Bharath,a presented an online handwritten symbol recognition system for Telugu

IARJSET
ISSN (Online) 2393-8021
ISSN (Print) 2394-1588
International Advanced Research Journal in Science, Engineering and Technology
ISO 3297:2007 Certified
Vol. 3, Issue 10, October 2016

is based on Hidden Markov Models (HMM)[37]. Normalized x-y coordinates, normalized first derivatives, normalized second derivatives, curvature and aspect ratio are used as thefeatures. The system achieved the accuracy of 98.7% on a dataset containing 29,158train samples and 9,235 test samples.

Mandalapu D., Prasanth L., Babu V. J., Sharma R. R.,andRao G. V. P. described character based elastic matching using local features for recognizing online Telugu handwritten data [38]. Dynamic Time Warping (DTW) has been used withfour different feature sets: x-y features, Shape Context (SC) and Tangent Angle (TA)features, Generalized Shape Context feature (GSC) and the fourth set containing x-y,normalized first and second derivatives and curvature features. Nearest neighbourhood classifier with DTW distance was used as the classifier with the recognition accuracyof 89.77%.

IV. CONCLUSION

The work presented a comparison of different recognition techniques used in OHCR of Indian scripts. Maximum recognition rate of 98% is reported with DFT, DCT and SVM classifier in Tamil script. In Malayalam, structural and directional features on SFAM technique produced a recognition rate of 98.26%.The existing literature in Telugu reported a recognition rate of 98.7% with HMM classifier and frequency, time domain features. An accuracy of 95% is reported for Kannada characters with wavelet features and Neural Networks. The maximum recognition rate attained in various Indian scripts uses different methods. Most of the classifiers used structural features or its refined form extensively. No similarity in the recognition scheme is visible in maximum recognition rate. But majority of the work employs DTW and SVM. The comparison of various recognition schemes reveals that feature selection has a major role in recognition accuracy in Indian scripts. A method used in one of the script gives lesser recognition rate in another script is also observed.

The study recommends that even better recognition methods are needed to achieve more promising recognition rate with minimum time for both writer independent and dependent OHCR methods in Indian scripts. The future works in OHCR ofIndian scripts may be focused on feature dimensions, class size and recognition time with optimum standards and Bench mark databases.

REFERENCES

[1] RejeanPlamondon and Sargur N Srihari. Online and on-line handwriting recognition: a comprehensive survey. IEEE Transactions on pattern analysis and Machine intelligence, 22(1):63{84, 2000.
[2] PJ Antony and KP Soman. Computational morphology and natural language parsing for indian languages: a literature survey. International Journal of Computer Science and Engineering Technology (IJCSET), 3:136,146, 2012.
[3] Tappert, C. C., Suen, C. Y., &Wakahara, T. (1990). The state of the art in online handwriting recognition.
[4] Huang, B. Q., Zhang, Y. B, &Kechadi, M. T. (2007). Preprocessing Techniques for Online Handwriting Recognition. Seventh ISDA (793–800).
[5] Kc, S., &Nattee, C. (2010). A Comprehensive Survey on On-line Handwriting Recognition Technology and its Real Application to the Nepalese Natural Handwriting. In Kathmandu University Journal of Science, Engineering and Technology (Vol. 5, pp. 31–55).
[6] Bharath, A., &Madhvanath, S. (2008). Online Handwriting Recognition for Indic Scripts Online Handwriting Recognition for Indic Scripts.
[7] Anitha Mary M.O Chacko, D. P. (2014). Handwritten Character Recognition in Malayalam Scripts - A Review. In International Journal of Artificial Intelligence & Applications (IJAIA) (Vol. 5, pp. 79–89).
[8] Scott D Connell, RMK Sinha, and Anil K Jain. Recognition of unconstrained online devanagari characters. In Pattern Recognition, 2000. Proceedings. 15th International Conference on, volume 2, pages 368,371. IEEE, 2000.
[9] Anoop M Namboodiri and Anil K Jain. Online handwritten script recognition.IEEE Transactions on Pattern Analysis and Machine Intelligence, 26(1):124-130, 2004.
[10] Niranjan Joshi, G Sita, AG Ramakrishnan, V Deepu, and SriganeshMadhvanath. Machine recognition of online handwritten devanagari characters. In Eighth International Conference on Document Analysis and Recognition (ICDAR'05), pages 1156{1160. IEEE, 2005.
[11] Abhimanyu Kumar and Samit Bhattacharya. Online devanagari isolated character recognition for the iphone using hidden markov models. In Students' Technology Symposium (TechSym), 2010 IEEE, pages 300{304. IEEE, 2010.
[12] VL Lajish and Sunil Kumar Kopparapu. Fuzzy directional features for unconstrained on-line devanagari handwriting recognition. In Communications (NCC),2010 National Conference on, pages 1-5. IEEE, 2010.
[13] VL Lajish and Sunil Kumar Kopparapu. Online handwritten devanagari stroke recognition using extended directional features. In Signal Processing and Communication Systems (ICSPCS), 2014 8th International Conference on, pages 1-5.IEEE, 2014.
[14] UtpalGarain, BB Chaudhuri, and TT Pal. Online handwritten indian script recognition: a human motor function based framework. In Pattern Recognition,2002. Proceedings. 16th International Conference on, volume 3, pages 164-167.IEEE, 2002.
[15] Ujjwal Bhattacharya, Bikash K Gupta, and S Parui. Direction code based features for recognition of online handwritten characters of bangla. In Ninth International Conference on Document Analysis and Recognition (ICDAR 2007),volume 1, pages 58{62. IEEE, 2007.
[16] Mitsuru Nakai, Naoto Akira, Hiroshi Shimodaira, and Shigeki Sagayama. Substroke approach to hmm-based on-line kanji handwriting recognition. In Document Analysis and Recognition, 2001. Proceedings. Sixth International Conference on, pages 491-495. IEEE, 2001.
[17] T Mondal, U Bhattacharya, SK Parui, K Das, and V Roy. Database generation and recognition of online handwritten bangla characters. In Proceedings of the International Workshop on Multilingual OCR, page 9. ACM, 2009.
[18] U Bhattacharya, A Nigam, YS Rawat, and SK Parui. An analytic scheme for online handwritten bangla cursive word recognition. Proc. of the 11th ICFHR,pages 320-325, 2008.
[19] M Sreeraj and Sumam Mary Idicula. On-line handwritten character recognition using kohonen networks. In Nature &Biologically Inspired Computing, 2009.NaBIC 2009. World Congress on, pages 1425-1430. IEEE, 2009.
[20] M Sreeraj and Sumam Mary Idicula. k-nn based on-line handwritten character recognition system. In Integrated Intelligent Computing (ICIIC), 2010 First International Conference on, pages 171{176. IEEE, 2010.
[21] KP Primekumar and Sumam Mary Idiculla. On-line malayalam handwritten character recognition using wavelet transform and sfam. In Electronics Computer Technology (ICECT), 2011 3rd International Conference on, volume 1, pages 49-53. IEEE, 2011.

[22] TR Indhu and VK Bhadran. Malayalam online handwriting recognition system:A simplified fuzzy artmap approach. In 2012 Annual IEEE India Conference (INDICON), pages 613-618. IEEE, 2012.
[23] KP Primekumar and Sumam Mary Idiculla. On-line malayalam handwritten character recognition using hmm and svm. In Signal Processing Image Processing & Pattern Recognition (ICSIPR), 2013 International Conference on, pages 322-326. IEEE, 2013.
[24] Binu P Chacko and Anto P Babu. Online sequential extreme learning machine based handwritten character recognition. In Students' Technology Symposium (TechSym), 2011 IEEE, pages 142-147. IEEE, 2011.
[25] AmrithaSampath, C Tripti, and V Govindaru. Freeman code based online handwritten character recognition for malayalam using back propagation neural networks. International journal on advanced computing, 3(4):51-58, 2012.
[26] Steffy Maria Joseph and Abdul Hameed. Online handwritten malayalam character recognition using libsvm in matlab. In Communication, Signal Processing and Networking (NCCSN), 2014 National Conference on, pages 1-5. IEEE, 2014.
[27] Niranjan Joshi, G Sita, AG Ramakrishnan, and SriganeshMadhvanath. Comparison of elastic matching algorithms for online tamil handwritten character 80 recognition. In Frontiers in Handwriting Recognition, 2004. IWFHR-9 2004.Ninth International Workshop on, pages 444-449. IEEE, 2004.
[28] KH Aparna, Vidhya Subramanian, M Kasirajan, G Vijay Prakash,VSChakravarthy, and SriganeshMadhvanath. Online handwriting recognition for tamil. In Frontiers in Handwriting Recognition, 2004. IWFHR-9 2004. Ninth International Workshop on, pages 438-443. IEEE, 2004.
[29] MV Ishwarya. An improved online tamil character recognition using neural networks. International Journal of Advanced Science and Technology, 42:1-10, 2012.
[30] VenkateshNarasimha Murthy and AngaraiGanesanRamakrishnan. Choice of classifiers in hierarchical recognition of online handwritten kannada and tamilaksharas. J. UCS, 17(1):94-106, 2011.
[31] RiturajKunwar and AG Ramakrishnan. Online handwriting recognition of tamil script using fractal geometry. In 2011 International Conference on Document Analysis and Recognition, pages 1389-1393. IEEE, 2011.
[32] Samuel, R. D. S. (n.d.). Wavelet Features based Recognition of Handwritten Kannada Characters SJ College of Engineering, 417–420
[33] Prasad, M. M., Sukumar, M., &Ramakrishnan, a. G. (2009). Divide and conquer technique in online handwritten Kannada character recognition. Proceedings of the International Workshop on Multilingual OCR - MOCR '09, 1.
[34] RiturajKunwar, P Mohan, K Shashikiran, and AG Ramakrishnan. Unrestricted kannada online handwritten akshara recognition using sdtw. In 2010 International Conference on Signal Processing and Communications (SPCOM), pages 1-5. IEEE, 2010.
[35] G Keerthi Prasad, Imran Khan, and Naveen Chanukotimath. On-line hindi handwritten character recognition for mobile devices. In Proceedings of the International Conference on Advances in Computing, Communications and Informatics,pages 1074{1078. ACM, 2012.
[36] HariharanSwethalakshmi, AnithaJayaraman, V SrinivasaChakravarthy, and C Chandra Sekhar. Online handwritten character recognition of devanagari and telugu characters using support vector machines. In Tenth International workshop on Frontiers in handwriting recognition. Suvisoft, 2006.
[37] V Babu, L Prasanth, R Sharma, GV Rao, and A Bharath. Hmm-based online handwriting recognition system for telugu symbols. In Ninth International Conference on Document Analysis and Recognition (ICDAR 2007), volume 1, pages 63-67. IEEE, 2007.
[38] KP Primekumar and Sumam Mary Idiculla. On-line malayalam handwritten character recognition using wavelet transform and sfam. In Electronics Computer Technology (ICECT), 2011 3rd International Conference on, volume 1, pages 49-53. IEEE, 2011.
[39] K Shashikiran, KolliSai Prasad, RiturajKunwar, and AG Ramakrishnan. Comparison of hmm and sdtw for tamil handwritten character recognition. In 2010 International Conference on Signal Processing and Communications (SPCOM),pages 1-4. IEEE, 2010.