A Grading of Writers On Offline Handwritten Recognition System Using Devanagari Script

Er.Rozy[1]
Assistant Professor
Department of Computer Science Engineering
Modern Group of College
Mukerian

Abstract— Devanagari is the most popular script which is mostly used in India.This paper presents a technique for grading of writers based on offline handwritten Hindi characters written by them. This process is complex task because of different person have different handwriting style. This paper have some techniques such as diagonal feature extraction , zoning feature extraction that is used to find the feature of individual character and k-Nearest neighbor classifiers using Euclidean distance method is used for classification score for each character. For testing data set, we collected seventy five different writers sample. For training data set, we considered five printed Devanagari fonts, namely,Mangal, devlys , krishna,Kruti_dev and Utsaah.

Keywords— *Gradation; Feature extraction; Peak extent based features;*

I. INTRODUCTION

Character recognition is a procedure that associates a predefined code to the items draw on paper. The unpredictable way of character recognition field, it is a dynamic zone of research even at this point. Optical character recognition is the recognition of printed or handwritten content by a PC. This includes filtering of the handwritten content, which changes over the paper record into a picture, and after that interpretation of the content picture into character codes. Optical Character Recognition (OCR) system is a most challenging computational process. Optical character recognition is a most mainstream technique used for handwriting recognition.OCR is basically formed of image recognition that is meant to recognize handwriting instead of faces or shapes. OCR was typically used to recognize printed and user handwritten text. Recognition of the offline handwritten *Devanagari* script is very difficult research area. The Recognition of user handwritten character is difficult task because each user has different writing styles. In OCR usage consists of various steps taken after by the real recognition. Recognition of handwritten characters has been great research range for a long time since it has numerous applications in all fields. *Devanagari* script is the script for composing *Hindi* language. Optical Character Recognition (OCR) framework is a most difficult computational process. Optical character recognition is a most mainstream technique utilized for handwriting recognition.OCR is fundamentally formed of picture recognition that is intended to recognize handwriting rather than faces or shapes.

OCR was regularly used to recognize printed and client handwritten content. Recognition of the offline handwritten *Devanagari* script is extremely difficult research zone. The Recognition of client handwritten character is difficult task that every client has different writing styles. Mostly researcher was worked on printed data and a few can work on client handwritten script. This presents a framework grading of writers for evaluating the handwriting of *Devanagari* writers in perspective of their handwriting. This framework of handwriting evaluation might be useful in conducting the handwriting competitions and then deciding the winners on the basis of an automated process. Grading of writers in perspective of their handwriting is a difficult task because every person has different handwritten style. Mostly researcher was worked on printed data and a few can work on client handwritten script. Optical character recognition is the most important part of a document analysis system. The character recognition system is the process of converted client handwritten and printed data into machine format.

II. LITERATURE SURVEY

Garg *et al*. (2013) describe segmentation based approach can be utilized to recognize handwritten *Hindi* character. The framework of segmentation of character recognition is the most important process. Initially, offline handwritten text is segmented into lines, words and character for recognition. Segmentation model finds individual character lines, line segmented into the words, and then word segmented into the characters. Structural features are extracted from the characters and fed into SVM classifier for recognition. The experimental results obtained with the proposed feature set using SVM classifier is very challenging. Ten samples of each character are used for the training phase of the classifier. For 41 characters 410 samples are used for the training phase. All the other characters are used for testing phase. Accuracy of

Characters recognized from correctly segmented characters is 89.6%.

Kumar *et al.* (2011) paper presents a framework for evaluating the handwriting of writers in perspective of their handwriting. This framework of handwriting evaluation might be useful in conducting the handwriting competitions and then winners can be deciding on the basis of an automated procedure. They proposed various features such as zoning feature, directional feature, diagonal feature, Zernike moments and intersection and open end points feature extraction techniques utilized to compute feature from each character. For obtaining the classification score, Bayesian classifiers, k-nearest neighbour and Hidden Markov Model classifier has been considered in the work. In this paper, they collect data from hundred different Punjabi writers and these (W_1, W_2,….., W_{100}) hundred sample are used for grading of handwriting system. They collect five printed *Gurmukhi* fonts (F1, F2… F5, respectively) namely, *Amrit*, *Anandpur Sahib*, *Granthi*, *LMP_TARAN* and *Maharaja*.

Garg *et al.* (2010) have provided the latest techniques of segmentation stage depend upon structural technique for user written *Devanagari* script. One of the most important methodologies of optical character recognition is Segmentation stage. The offline user written text is segmented into lines, words and character for recognition. The handwritten *Hindi* language text finds character lines are separated into words and word divided into text. The problem occurred in segmentation spread to recognitions.

Kumar *et al*. (2013) proposed modified division point extraction process**.** The features are extracting from input characters by using modified division point extraction technique. They considered linear-support vector machines, *k*-nearest neighbour, and multilayer perceptron as classifiers. For experimental performance, they used 10,500 samples of the isolated, offline, handwritten, basic 35 characters of Punjabi text. They have achieved maximum accuracy of recognition is 84.5%, 85.8% and 89.2% which are dealing with linear-support vector machines, Multilayer Perceptron and k-Nearest Neighbour classifiers, respectively, with a five-fold cross validation technique.

Jayadevan *et al.* (2011) have presented significant progress in research related to the optical character recognition of handwritten and print *Devanagari* script in the previous a few years. From 1970s state can propose machine printed and handwritten *Devanagari* recognition of text is discussed here. Different feature extraction technique and classification, training point and matching technique are used for the recognitions are discussed in different sections of this paper. An effort can have made to solve the report of the result and also try to explain the advantageous direction of the researchers till date.

Kumar *et al.* (2011) have used *k*-Nearest neighbours classifier, hence to find the *k*-NN algorithm using Euclidean distance method writer sample vector to reference printed vector are calculated. They considered one hundred different user samples of offline handwritten *Gurmukhi* character. They have achieved accuracy of recognition is 94.1% using diagonal features and *k*-Nearest neighbours algorithm.

Bansal *et al.* (2014) Offline handwritten text recognition is the dynamic research regions in the field of recognition. Character written by hand is recognition is an extremely difficult research area because every person has their different writing style and size of character, so it is very difficult to recognize the correct handwritten characters and digits. This paper proposed strategy of character recognition has different stages to be specific, Data obtaining, Digitization, pre-processing, Feature extraction, and Classification. They proposed Feature extraction procedure named as Neighbourhood Foreground Pixels Density strategy. It is utilized for discovering different features of character which gives the relevant data about that character. This paper proposed three classifiers in particular, Support Vector Machine (SVM), Naïve Bayes and also Multilayer Perceptron to recognize handwritten *Gurmukhi* characters.

Malik and Singh (2016) Record segmentation is one of the critical stages in machine acknowledgment of any dialect. The character segmentation strategy is explained in this paper. Segmentation model finds individual character lines, line segmented into the words, and then word segmented into the characters. *Devanagari* is one of the acknowledged script in India. It is utilized for lettering *Hindi*, *Marathi*, *Sanskrit* and *Nepali* dialects. In addition, *Hindi* is the third most acknowledged language on the planet. *Devanagari* consists of vowels, consonants and different modifiers. Variety of Gradient, Structural features and artificial neural network (ANN) is proposed in this paper.

Gaurav *et al.* (2013) Optical Character Recognition (OCR) is a requesting field of dynamic research in recognition field and vision system. It can use in actuality applications. It is a kind of record investigation where an examined paper image by scanner that have either printed or written by hand code is offered input to an OCR programming engine is changed over machine-readable text format. The field of OCR is arranged into two sections, initially is discovery of printed characters by machine and the second is recognition of handwritten characters. Recognition of handwritten characters is a critical range of research because of its different application possibilities Feature extraction is the most imperative stage of procedure of OCR. This gives information about various element extraction strategies that are utilized as a part of OCR.

III. METHODLOGY

This framework of handwriting evaluation might be useful in conducting the handwriting competitions and winners can be decided on the basis of an automated framework. “Grading means judge the superiority. The grading systems can be used to grade the handwriting of writers and it’s used for verification of signature means determining whether or not the signature is that of a given person. Grading of writers in

perspective of their handwriting is a difficult task because every person has their different writing style. Selecting the various features is an intricate task. Various feature extractions techniques have been considered for extracting the vital information of writers, which can be used for grading the writers. For obtaining the classification score, classifiers have been considered. A handwriting grading framework has various phases like, digitization, pre-processing, feature extraction, classification and grading based on the classification score. These steps are addressed in following sub-sections.

A. Digitization

Digitization is the primary phase of handwriting grading system. The client hand written text is converted into electronic form by scanning the hand writing sample on a paper that is offline Handwritten character recognition. The aim of digitization is to produce the digital image. Digitization creates the computerized picture, which is fed to the pre-processing stage**.**

B. Pre-Processing

Pre-processing is the preliminary phase of framework of handwriting. This phase receives a binary image characters. Pre-processing is used for reduce the noise in handwritten texts. Pre-processing performed binarization is changing over a gray scale image into dark/black- white image. Where some portion of image is appeared or it is dark/black is shown as 1 and the part where image is not appeared or it is white is shown as 0.

C. Feature extraction

Feature extraction is the best and effective phase of handwriting grading system. It is a vital part of handwriting grading system. This phase is extracting feature from noise free character image. Selecting the different features is an intricate task for implementing a handwriting grading system of particular language. Various feature extractions techniques such as Diagonal feature, zoning based feature, directional features, peak extent based feature extraction and parabola curve fitting have been considered for extracting the vital information of writers, which can be used for grading the writers. In this work, we considered zoning based feature and peak extent based feature extraction: -

1. Zoning based feature

Step I: Split the input image into 25 zones and each of equal size 10×10 pixels as shown in Figure 4.2.
Step II: After that, calculate the number of ON pixels in each zone.
Step III: Normalize the values in feature set to [0, 1].

Z_1	Z_2	Z_3	Z_4	Z_5
Z_6	Z_7	Z_8	Z_9	Z_{10}
Z_{11}	Z_{12}	Z_{13}	Z_{14}	Z_{15}
Z_{16}	Z_{17}	Z_{18}	Z_{19}	Z_{20}
Z_{21}	Z_{22}	Z_{23}	Z_{24}	Z_{25}

Fig. 1.**Zoning based feature**

2. Peak extent based features

In this part, we explain a procedure for feature extraction, to be specific, peak extent based feature. Peak extent based features can be extracted feature horizontally and vertically. In each row of a zone can be calculated the peak extent by added of successive ON pixels, in horizontally peak extent. In each row of a zone can be calculated the peak extent by added of successive ON pixels, in horizontally peak extent.

Few steps are used to acquire these features for a given character: -
Step I: Split the binary image of character into *n* (=100) number of zones, each of equal size 10×10 pixels.
Step II: In each row of a zone can be calculated the peak extent by added of successive ON pixels as shown in Fig 1.
Step III: Thus, in each row of a zone replacing the values of ON pixels by peak extent value.
Step IV: In each Row, find out the greatest value of peak extent. Then 10 peak extent feature of each zone.
Step V: Obtaining the whole sum of these 10 peak extent sub feature. Consider this as a feature for corresponding zone.
Step VI: If any zone has no ON pixel, then peak extent feature value is zero.
Step VII: Normalize of these features by dividing every component of the feature by greatest value in the feature.

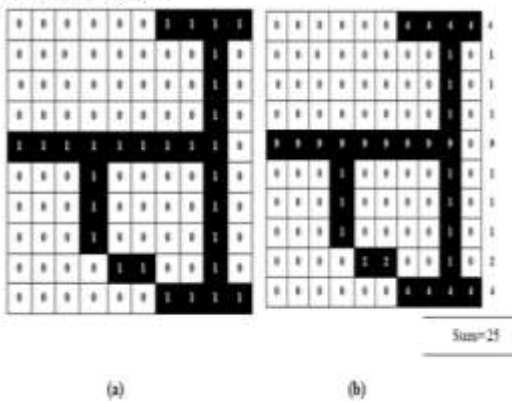

Figure 4.3 Peak extent based features: (a) Zone of binary image, (b) Horizontally peakextent based features

Similarly, in vertically peak extent features, calculated the peak extent by added of successive ON pixels in each column of zone. Utilizing this calculation, we will acquire *n* features to each zone.

D. Classification

Classification score is one of the successful phases of grading system. Classification score is one of the effective stages of grading system. The main objective of Classification phase is that features are extracted from character is used for obtain classification score. Classification is the decision making step of the grading system. The classifier evaluation might be useful in conducting the handwriting competitions and then deciding the winners on the basis of an automated process. Features of texts are comparing with already stored training dataset for conducting the handwriting competition of *Devanagari* text writers. Some classifier techniques are- *k*- Nearest Neighbours (*k*-NN), Hidden Markov Model and Bayesian classifiers.

1. *k*- Nearest Neighbour (*k*-NN) classifier

Initially all the training set of grading system is given then the mock testing samples starts finds the nearest neighbour which match closely to the testing sample from the training set given with the help of nearest neighbour algorithm for conducting the handwriting competition of *Devanagari* text writers. The *k*-Nearest neighbour algorithm based upon neighbourhood in the feature space is used for classifying of character. The k-Nearest neighbour classifiers using Euclidean distance method is used for classification score for each character. In this classifier, hence to find the *k*-NN algorithm using Euclidean distance method writer sample vector to reference printed vector are calculated. The Euclidean distance between sample vector and printed vector is show below:

Euclidean distance (d) = $\sqrt{\sum_{K=1}^{N}(x_k - y_k)^2}$

Where, N means total number of features, x_k be the reference stored vector and y_k be the writer sample vector. Compute the smallest Euclidean distance between sample vector and printed vector.

IV. Experimental Results

In this section, we show experimental results for printed and handwritten font and character recognition using different classification techniques.

Script Character	*Devlys* (F_1)	*Krishna* (F_2)	*Kruti Dev* (F_3)	Utsaah (F_4)
अ	अ	अ	अ	अ
आ	आ	आ	आ	आ
इ	इ	इ	इ	इ
ई	ई	ई	ई	ई
उ	उ	उ	उ	उ

Fig. 3. Samples of printed characters from four *Devanagari* fonts (Training data set)

As discussed in section, the gradation results, in view of the qualities acquired by *k*-NN classifier are presented in this section. Classification scores acquired with *k*-NN classifier are standardized to [0, 100] in order to give the grade in percentage form.

1. Grading using *k*-NN classifier

We proposed grading using *k*-NN classifier classifier the features, namely, zoning and peak extent based feature elements as an input to *k*-NN classifier. We perform a handwriting grading system based on *Devanagari* characters. This framework of handwriting evaluation might be useful in conducting the handwriting competitions and winner can be decided on the basis of process.In below the experimental results for zoning and peak extent based feature are presented.

1.1 Grading using zoning based features

In this part, gradation results of writers in view of zoning based features and *k*-NN classifier are used. Using zoning based feature, it has noticed that font *TF* (with a score of 100) is the best font. On similar lines, it has additionally been observed that writer W_{33} (with a score of 95.38) is the best writer.

1.2 Grading using peak extent based features

When we use peak extent based feature elements as an input to *k*-NN classifier in proposed grading system. *Devanagari* font style *TF* (with a score of 100) turns out

to be the best text style. It has likewise been seen that writer W_{44} (with a score of 91.31) is the best writer among the seventy-five writers taken in this study.

1.3 Final grading using the average of zoning and peak extent based features

Here, average grading, based on above mentioned two features has been presented. It was observed that if we use the average score of these two features and *k*-NN classifier, then font *TF* (with an average score of 100) is the best font and writer W_{44} (with an average score of 98.62) is the best writer. Final average grading scores of the font *TF* and writers (W_1, W_2, …, W_{75})

Table 1: Average grading of writers using *k*-NN classifier

Feature type	Best font and first writer
Zoning features	*TF* (100); W_{33} (95.38)
Peak extent feature	*TF* (100); W_{44} (91.31)
Average with zoning and peak extent based features	*TF* (100); W_{44} (98.62)

IV. CONCLUSION

In our research, we contributed our effort in proposing a system to recognize Offline Handwritten *Devanagari* Characters. We gather data from seventy five (*W1, W2, … , W75*) different writers and different wide issues are happened in client handwriting data because every person has their different writing style, so it is very difficult to recognize the correct handwritten characters and digits. We have used two features extraction techniques namely, zoning features and peak extent features. For obtaining the classification score, *k*-NN classifiers have been considered. The experimental results of this proposed system are graphically. Using zoning based feature, it have been observed that font TF with a score of 100% is the best font and observed that writer W_{33} with a score of 95.38% is the best writer. Using peak extent based features, it has been observed that font TF with a score of 100% is the best font and observed that writer W_{44} with a score of 91.31% is the best writer. Final grading using the average of zoning and peak extent based features observed that font TF with an average score of 100% is the best font and writer W_{44} with an average score of 98.62% is the best write. This framework can likewise be stretched out for grading writers using offline handwritten characters of different scripts after building the dataset of these scripts.

V. REFERENCES

1. Badhika S 2013 An Empirical Study on Identification of Strokes and their Significance in Script Identification, *International Journal of Modern Engineering Research* (IJMER), 3(4):2232-2238.
2. Bag S and Harit G (2013) A survey on optical character recognition for Bangla and Devanagari scripts, *SADHANA*, 38(1):133–168.
3. Bansal S, Garg M, and Kumar M 2014 A Technique for Offline Handwritten Character Recognition, *International Journal of Computing and Technology*, 1(2): 210-215.
4. Barawkar M B 2015 Devanagari Character Recognition A Review, *IOSR Journal of Electronics and Communication Engineering*, 5-9.
5. Bhowmik, T K, Ghanty, P, Roy A and Parui S K 2009. SVM-based hierarchical architectures for handwritten Bangla character recognition, *International Journal on Document Analysis and Recognition*, 12(2):97-108.
6. Garg N K, Kaur L, Jindal M K 2010 Segmentation of Handwritten Hindi Text, *International Journal of Computer Applications*, 1:22-26.
7. Garg N K, Kaur L, Jindal M K 2011 The Hazards in Segmentation of Handwritten Hindi Text, *International Journal of Computer Applications*, 29:30-34.
8. Garg N K, Kaur L, Jindal M K 2013 Recognition of Offline Handwritten Hindi Text Using SVM, *International Journal of Computer Applications*, 7(4):395-401.
9. Gaurav Y. Tawde, Mrs. Jayashree M. Kundargi 2013 An Overview of Feature Extraction Techniques in OCR for Indian Scripts Focused on Offline Handwriting, *International Journal of Engineering Research and Applications*, 3(1):919-926.
10. Goyal N and Jain S 2015 Optimized Hindi Script Recognition using OCR Feature Extraction Technique, *International Journal of Advanced Research in Computer and Communication Engineering*, 4(8):419- 424.
11. Herekar R R and Dhotre S R 2014 Handwritten Character Recognition Based on Zoning Using Euler Number for English Alphabets and Numerals, *IOSR Journal of Computer Engineering (IOSR-JCE)*, 16(4):75-88.

12. Joshi M R and Sabale V V 2015 Offline character recognition for printed text in Devanagari using Neural Network and Genetic Algorithm, *International Journal of Advanced Research in Computer and Communication Engineering*, 4(5):175-179.

13. Kaur P and Singh S 2015 Perform Evaluation of Various Feature Extraction Techniques for Gurmukhi Script, *International Journal of Computer Applications*, 119(19):16-22.

14. Kumar M, Jindal, M K, Sharma R K 2011 Offline Handwritten *Gurmukhi* Character Recognition Using Curvature Feature, *Proceedings of International Conference on Advances in Modelling, Optimization and Computing*, 981-989.

15. Kumar M, Sharma R K, Jindal M K. 2011 Classification of Characters and Grading Writers in Offline Handwritten *Gurmukhi* Script, *Proceedings of International Conference on Image Information Processing*, 1-4.

16. Kumar M, Sharma R K, Jindal M K 2011 *k* -Nearest Neighbour Based Offline handwritten Gurmukhi Character Recognition, *Proceedings of International Conference on Image Information Processing*, 1-4.

17. Kumar M, Jindal, M K, Sharma R K 2011 MDP Feature Extraction Technique for Offline Handwritten Gurmukhi Character Recognition, *Smart Computing Review*, 3(6):397-404.

18. Kumar M, Jindal, M K, Sharma R K 2016 A Novel Framework for Grading of Writers using Offline *Gurmukhi* Character, *Proceedings of the National Academy of Sciences, India Section A: Physical Sciences*, 86(3):405-415.

19. Lehal G S, Singh C and Lehal R 2001 A Shape Based Post Processor for Gurmukhi OCR, IEEE Proceedings of the Sixth International Conference on Document Analysis and Recognition (ICDAR'01), 1105-1109.

20. Malik N, Singh A 2016 Character recognition of offline handwritten Devanagari script using artificial neural network, *International Journal of Advanced Computing Research*, 2(2):35-41.

21. Moni B S and Raju G (2011) Modified Quadratic Classifier and Directional Features for Handwritten Malayalam Character Recognition, *Computational Science - New Dimensions & Perspective*, 30-34.

22. Padma M C and Vijaya P A 2009 Monothetic Separation of Telugu, Hindi and English Text Lines from a Multi Script Document, IEEE International Conference on Systems, Man, and Cybernetics, 4870-4875.

23. Pal U, Roy K R, Roy K and Kimura F 2009 Indian Multi-Script Full Pin-code String Recognition for Postal Automation, *IEEE 10th International Conference on Document Analysis and Recognition*, 456-460.

24. Pradeep J, Srinivasan E and Himavathi S 2011 Diagonal based feature extraction for handwritten alphabets recognition system using neural network, *International Journal of Computer Science & Information Technology (IJCSIT)*, 3(1):27-38.

25. Sarma B, Mehrotra K, Naik R K, Prasanna S R M, Belhe S and Mahanta C 2013 Handwritten Assamese Numeral Recognizer Using HMM & SVM Classifiers, *IEEE*.

26. Sastry P N, Krishnan R and Ram B V S 2010 Classification and identification of telugu handwritten characters extracted from palm leaves using decision tree approach, ARPN Journal of Engineering and Applied Sciences, 5(3): 22-32.

27. Shanthi N and Duraiswamy K 2010 A novel SVM-based handwritten Tamil character recognition system, *Pattern Analysis and Applications*, 13(2):173-180.

28. Sharma A, Verma V (2016) Handwritten Hindi Character Recognition, *International Journal of Advanced Research in Electronics and Communication Engineering (IJARECE)*, 5:1455-1460.

29. Singla G and kumar P (2013) extract the punjabi word with edge detector from machine printed document images, *International Journal of Computer Science & Engineering Technology (IJCSET)*, 4(5):543-545.

30. Sharma D and Jhajj P 2010 Recognition of Isolated Handwritten Characters in Gurmukhi Script, *International Journal of Computer Applications*, 4(8):9-17.

31. Sharma N, Pal U, and Kimura F 2006 Recognition of Handwritten Kannada Numerals, *9th International Conference on Information Technology*, 133-136.

32. Sreeraj M and Idicula S M 2010 k-NN based On-Line Handwritten Character recognition system, *IEEE International Conference on Integrated Intelligent Computing*, 171-176.

33. Singh G, Lehri G 2012 Recognition of Handwritten Hindi Characters using Backpropagation Neural Network, *International Journal of Computer Science and Information Technologies*, 3(4):4892-4895.

34. Rani V, Luxmi V 2013 Character Segmentation of Handwritten Document in Devanagari Script, *International Journal of Computer science and technology*, 4(1):95-97.

35. Tapia E and Rojas R 2003 Recognition of On-line Handwritten Mathematical Formulas in the E-Chalk System, *IEEE Proceedings of the Seventh International Conference on Document Analysis and Recognition*,1-5.

36. Tanuja K, Kumari U V and Sushma T M 2015 Handwritten Hindi Character Recognition System Using Edge detection & Neural Network, *International Journal of Advanced Technology and Engineering Exploration*, 2(6):71-75.

37. Wanchoo A S, Yadav P, Anuse A 2016 A Survey on Devanagari Character Recognition for Indian Postal System Automation, *International Journal of Applied Engineering Research*, 11:4529-4536.