

Devanagari Handwriting Grading System Based on Curvature Features

Munish Kumar¹, Simpel Rani Jindal²

Abstract: Grading of writers in perspective of their handwriting is a challenging task owing to various writing styles of different individuals. This paper presents a framework for grading of Devanagari writers in perspective of their handwriting. This framework of grading can be useful in conducting the handwriting competitions and then deciding the winners on the basis of an automated process. Selecting the set of features is a challenging task for implementing a handwriting grading system of particular language. In this paper, curvature features, namely, parabola curve fitting and power curve fitting have been considered for extracting the vital information of writers, which can be helpful in grading the writers. For obtaining the classification score, k-NN classifier has been considered in the present work. Four printed Devanagari font styles, namely, *Devlys*, *Krishna*, *Krutidev*, and *Utsaah* have been considered for train the proposed model of handwriting evaluation. For evaluating the effectiveness of the proposed framework, we have conducted a mock test of 75 Devanagari writers (Left handed and Right handed) and noticed that the proposed framework performing well suitable for conducting the handwriting competition of Devanagari text writers. This work is also newly in the ground of Devanagari text recognition.

Keywords: Handwritten character recognition, curvature features, parabola curve fitting, power curve fitting, k-NN.

1 Introduction

Optical character recognition is a process of converting the handwritten and printed text into machine processable format. Handwriting recognition is divided into two categories, namely, online and offline. Online handwriting recognition manages the transformation of text information that is composed on an electronic surface with the help of a special pen. In online handwriting, pen strokes information is also accessible which will be helpful in enhancing the precision of online handwriting recognition system. In offline handwriting, the input data will be scanned image of text data. A good number of scholars have exhibited work for printed and handwritten text recognition. Mostly researcher has worked for printed text recognition, and a couple of researchers have done

¹ Department of Computer Applications, GZS Campus College of Engineering & Technology, (Maharaja Ranjit Singh Punjab Technical University), Bathinda, Punjab, India.E-mail: munishcse@gmail.com

² Computer Science and Engineering Section, Yadavindera College of Engineering, Talwandi Sabo, Bathinda, Punjab, India.E-mail: simpel_jindal@rediffmail.com

work for handwriting recognition. For example, Bansal and Sinha (2000) have built up a procedure for Devanagari text recognition. A recognition system based on Hindi handwritten words was presented by Belhe, Paulzagade, Deshmukh, Jetley, and Mehrotra (2012). They have used HMM and tree classifier for recognition and acquired a recognition accuracy of about 89%. Garg, Kaur and Jindal (2010) have segmented the handwritten Hindi text and they have considered structural feature extraction techniques and SVM classifier for recognition of handwritten Hindi characters. Using these features and classifier, they have achieved a recognition accuracy of 89.6%. Pal, Sharma, Wakabayashi and Kimura (2007) have set into motion, a system for offline handwritten Devanagari character recognition. They have achieved a recognition accuracy of 94.2% with five-fold cross validation test. [Kumar, Sharma and Jindal (2014)] have presented a study of various features and classifiers for offline handwritten character recognition. In this study, they have accomplished maximum recognition accuracy of 97.7% with PCA. For exploratory outcomes, they have considered 7,000 examples of 35-class problem. They have also provided a framework for handwriting grading of Gurmukhi writers. They have assessed the execution of their framework with various printed Gurmukhi text styles [Kumar, Sharma and Jindal (2016)]. In this paper, we have proposed a framework for grading of Devanagari writers using curvature features, namely, parabola curve fitting and power curve fitting based features. This framework will be useful for conducting the handwriting competitions and deciding the winners in view of their handwriting. Proposed framework comprises of different stages, like, digitization, pre-preparing, feature extraction, classification and grading based on classification score. We have composed this paper into 5 segments. Preamble and related work is presented in Section 1. In Section 2, we have introduced about Devanagari script. Section 3 comprises of different phases of the proposed framework. Experimental results are described in Section 4. In last Section 5, portrays the conclusion and future extension.

2 Devanagari script

The Devanagari is the most frequently used script which is mostly used in India and furthermore utilizing to compose text in Hindi, Marathi and Nepali dialects. The word Devanagari can be getting from two Sanskrit terms "Deva" means god and "Nagri" means city. Devanagari script has 49 characters which have 36 vyanjan (pure type of consonants) and 13 svar (vowel) as appeared in Fig. 1(a) and 1(b). It also incorporates 'signs and symbol' is the set of modifiers called Matra. In Devanagari script a few characters are similar to each other as appeared in Fig. 2, which makes their recognition a bit difficult.

क	ख	ग	घ	ङ
च	छ	ज	झ	ञ
ट	ठ	ड	ढ	ण
त	थ	द	ध	न
प	फ	ब	भ	म
य	र	ल	व	श
ष	स	ह	क्ष	त्र
ज्ञ				

Figure 1(a): Consonants in Hindi

अ आ इ ई उ
ऊ ऋ ए ऐ ओ
औ अं अः

Figure 1(b): Vowels in Hindi

अ and अं	इ and ई	उ and ऊ
ओ and औ	क and फ	ड and ड
छ and ध	ज and त्र	ढ and द
ब and व	प and य	प and ष

Figure 2: Samples of similar characters in Devanagari script

3 Phases of proposed grading system

The meaning of “grading of writers” is to judge the superiority of writing styles related to printed fonts. The grading systems can be used to grade the handwriting of writers and can also be used for verification of signature means determining whether or not the signature is that of a given person. Phases of proposed handwriting grading system are digitization, pre-processing, feature extraction, classification and grading based on the classification score. The architecture of the grading system is given in Fig. 3.

3.1 Digitization

Digitization is the initial stage of the proposed framework for grading of writers. In this stage, paper based handwritten data is converted into electronic shape by scanning the paper on the scanner. The aim of digitization is to create the computerized image which is fed to the pre-preparing stage.

3.2 Pre-processing

Pre-processing is the following phase of the digitization phase which is used to produce the bitmap image of the digital image. Pre-processing also consists of a process for; make the thinned image of the bit mapped image by utilizing parallel thinning algorithm [Zhang and Suen (1984)].

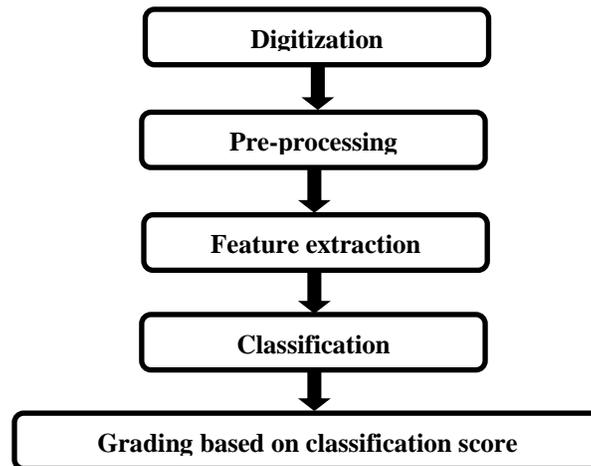


Figure 3: Architecture of proposed grading system

3.3 Feature extraction

Feature extraction is the primary stage of the proposed handwriting grading system. The performance of handwriting grading system depends on features that are being extracted in this phase. So, this phase played a major role for evaluation of the handwriting of different individuals. In this phase, we will extract the features of noise free bitmapped character images written by various writers. In past, structural and statistical features have been used for recognition of different scripts. In this paper, we have explored two new curvature features, namely, parabola curve fitting and power curve fitting. These features are proposed by Kumar, Sharma and Jindal (2014). They have used these feature extraction techniques for offline handwritten Gurmukhi character recognition. They have compared these techniques with other most recent feature extraction techniques and noticed that curvature features perform better than other techniques for their work. So, here, we considered these techniques for evaluating the handwriting of Devanagari writers. These techniques are briefly discussed in following sub-sections.

3.3.1 Parabola curve fitting based features

In parabola curve fitting based technique, firstly character thinned image is divided into equal sized n zones. By using the least square method, this process constructs a parabolic curve that suggests a best fitted to series of foreground pixels to given data. The method of least square is perhaps the most systematic procedure to fit a unique curve through the given data points. A parabola for each zone is fitted using the least square method. Let, a parabola curve equation $y = a_0 + a_1x + a_2x^2$ where a_0, a_1, a_2 three parameters. The values of these three parameters a_0, a_1, a_2 are computed by solving the following equations obtained from the least square method.

$$y = a_0 + a_1x + a_2x^2$$

$$\sum y = na_0 + a_1 \sum x + a_2 \sum x^2$$

$$\begin{aligned}\sum xy &= a_0 \sum x + a_1 \sum x^2 + a_2 \sum x^3 \\ \sum x^2 y &= a_0 \sum x^2 + a_1 \sum x^3 + a_2 \sum x^4\end{aligned}$$

The following steps are used to extract these features:

- The thinned image is divided into n equal sized zones.
- A parabola for each zone is fitted using the least square method and parabola equation $y = a_0 + a_1x + a_2x^2$ where a_0, a_1, a_2 parameters are calculated.
- If any zone does not have a foreground pixel, then values of a_0, a_1, a_2 parameters are taken as zero.

3.3.2 Power curve fitting based features

In power curve fitting technique, again character thinned image is divided into equal sized n zones. By using the least square method, this process constructs a power curve that suggests a best fitted to series of foreground pixels to given data. The method of least square is perhaps the most systematic procedure to fit a unique curve through the given data points. A power curve for each zone is fitted using the least square method. Let a power curve equation $y = ax^b$, the values of these two parameters a, b are computed by using least square method.

The following steps are used to extract these features:

- The thinned image is divided into n equal sized zones.
- A power curve for each zone is fitted using least square method and power equation $y = ax^b$ where a, b two parameters are calculated.
- If any zone does not have a foreground pixel, then values of a, b parameters are considered as zero.

$$\begin{aligned}y &= ax^b \\ \log y &= \log a + b \log x\end{aligned}$$

Let us take, $\log y = Y$, $\log a = A$ and $\log x = X$

$$\begin{aligned}Y &= A + bX \\ \sum Y &= nA + b \sum X \\ \sum XY &= A \sum X + b \sum X^2\end{aligned}$$

3.4 Classification

Classification phase is also one of the important phases of the proposed grading system. Classification phase uses the features extracted in the previous phase for classification of writers based on their handwriting. In this work, we have considered Nearest Neighbours (NN) classifier. In the NN classifier, Euclidean distance from the applicant vector to stored vector is computed. The Euclidean distance between an applicant vector and stored vector is given by,

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

Here, N is the aggregate number of elements in a list of feature vector, x_k is the library stored vector and y_k is the applicant feature vector. The class of the library stored feature delivering the smallest Euclidean distance, when compared with the candidate feature vector, is allocated to the input character.

4 Data set and experimental results

In training dataset, four printed Devanagari font styles, namely, *Devlys* (F_1), *Krishna* (F_2), *Krutidev* (F_3), and *Utsaah* (F_4) are taken. For evaluating the effectiveness of the proposed framework, a mock test of 75 writers has been conducted. In this mock test, 35 writers were left-handed and 40 writers were right-handed. So, we have additionally observed that framework won't impact, regardless of whether the writer is left-handed or whether right-handed. A few samples from training and testing data set are shown in Fig. 4-5, respectively. Experimental results of the proposed grading system based on curvature features, in view of the qualities acquired by NN classifier are presented in this section. Classification scores obtained with NN classifier are standardized to [0, 100] in order to give the grade in percentage form. Feature-wise results of proposed grading system for 75 writers are provided in the following sub-sections.

Script Character	<i>Devlys</i> (F_1)	<i>Krishna</i> (F_2)	<i>Krutidev</i> (F_3)	<i>Utsaah</i> (F_4)
क	क	क	क	क
ख	ख	ख	ख	ख
ग	ग	ग	ग	ग
घ	घ	घ	घ	घ

Figure 4: A few samples of training dataset

Script Character	Writers			
	w_1	w_2	w_3	w_4
क	क	क	क	क
ख	ख	ख	ख	ख
ग	ग	ग	ग	ग
घ	घ	घ	घ	घ

Figure 5: A few samples of testing dataset

4.1 Grading using parabola curve fitting based features

In this sub-section, experimental results of the proposed grading system with parabola

curve fitting based features and k -NN classifier are presented. Using this feature, it has been noticed that writer W_{16} (with a score of 100) is the best writer. Graphically results of writers (W_1, W_2, \dots, W_{75}) based on this feature extraction technique, are presented in Fig. 6.

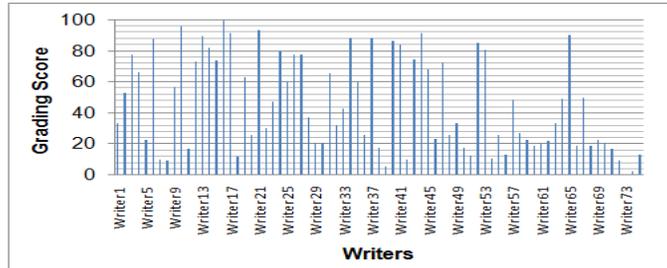


Figure 6: Writer wise grading score using parabola curve fitting based features

4.2 Grading using power curve fitting based features

In this sub-section, experimental results of the proposed grading system with power curve fitting based on features and k -NN classifier are presented. Using this feature, it has been noticed that writer W_{72} (with a score of 100) is the best writer. Graphically results of writers (W_1, W_2, \dots, W_{75}) using power curve fitting based features are presented in Fig. 7.

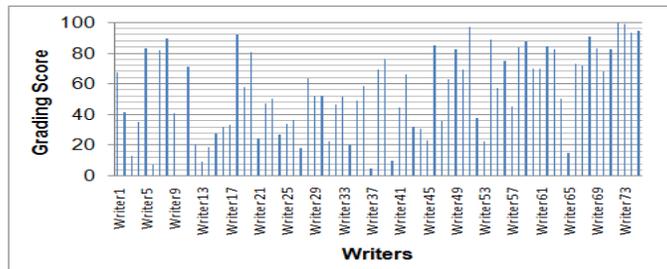


Figure 7: Writer wise grading score using power curve fitting based features

4.3 Final grading using the average of parabola curve fitting and power curve fitting based features

Here, average grading, based on above mentioned two features has been presented. It has been observed that writer W_{16} (with an average score of 66.06) is the best writer. Final average grading scores of writers (W_1, W_2, \dots, W_{75}) considered in this study are shown in Fig. 8.

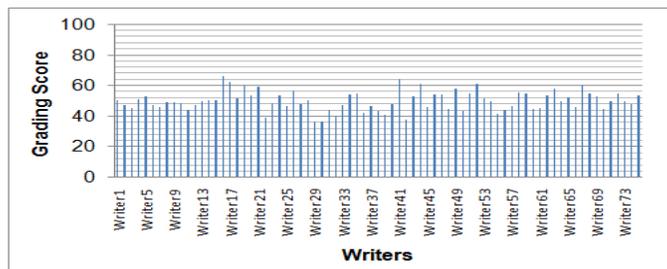


Figure 8: Final grading of writers using curvature features

5 Conclusions and Future Scope

In this paper, a framework has been presented for calculating the grading score of the Devanagari writers. Curvature features have been considered for extract the information related to the handwriting of individuals. The framework, proposed in this paper, is tested with four well known printed Devanagari text fonts, namely, *Devlys*, *Krishna*, *Krutidev* and *Utsaah*. Proposed system can be used as a decision support system for conducting the handwriting competitions of Devanagari writers. We have also concluded that if a writer's handwriting is very artistic, left-handed or right-handed etc. then the proposed system will not fail. This framework can likewise be stretched out for grading of writers using offline handwritten characters of different scripts after building the training dataset of these scripts.

References

- Bansal, V., Sinha, R. M. K.** (2000): Integrating knowledge sources in Devanagari text recognition system. *IEEE Transactions on Systems, Man and Cybernetics-Part A*, vol. 30, no. 4, pp. 500-505.
- Belhe, S., Paulzagade, C., Deshmukh, A., Jetley, S. and Mehrotra, K.** (2012): Hindi handwritten word recognition using HMM and symbol tree. *Proc. of the Workshop on Document Analysis and Recognition*, pp. 9-14.
- Garg, N. K., Kaur, L. and Jindal, M. K.** (2010): A new method for line segmentation of handwritten Hindi text. *Proc. of the 7th International Conference on Information Technology: New Generations*, pp. 392-397.
- Kumar, M., Sharma, R. K., and Jindal, M. K.** (2014): Efficient Feature Extraction Techniques for Offline Handwritten Gurmukhi Character Recognition. *National Academy Science Letters*, vol. 37, no. 4, pp. 381-391.
- Kumar, M., Sharma, R. K. and Jindal, M. K.** (2016): A Framework for Grading Writers using Offline Gurmukhi Characters. *Proceedings of the National Academy of Sciences- Physical Science-A*, vol. 86, no. 3, pp. 405-415.
- Pal, U., Sharma, N., Wakabayashi, T. and Kimura, F.** (2007): Off-line handwritten character recognition of Devanagari script. *Proc. of the 9th International Conference on Document Analysis and Recognition*, pp. 496-500.
- Zhang, T. Y. and Suen, C. Y.** (1984): A fast parallel algorithm for thinning digital patterns. *Communications of the ACM*, vol. 27, no. 3, pp. 236-239.