IJCSE International Journal of Computer Sciences and Engineering Open Access
Research Paper Vol.-7, Issue-3, March 2019 E-ISSN: 2347-2693

Improvement in the Online Handwritten Kannada Numeral Recognition with the Difference Feature

M. Mahadeva Prasad

Dept. of Studies in Electronics, Post Graduate Centre, University of Mysore, Hassan, India

Corresponding author: prasada9@gmail.com

DOI: https://doi.org/10.26438/ijcse/v7i3.868870 | Available online at: www.ijcseonline.org

Accepted: 12/Mar/2019, Published: 31/Mar/2019

***Abstract*—**The paper discusses the performance of the online handwritten recognition module designed for Kannada numerals. In the proposed system, there is an improvement in the recognition accuracy of the recognition module when compared to the previous ones. The difference feature has given the improved performance of the recognition module. The difference feature vector is formed by computing the difference of consecutive x- and y- coordinate values of online handwriting. The writer independent experiments are carried out by dividing the collected online handwritten Kannada numeral data into a disjoint set of training and testing data sets. Out of 1400 numeral data samples, 700 numeral data samples are used for training the recognition modules and the remaining 700 numeral data samples are used for testing. The collected online handwritten data are subjected to preprocessing and feature extraction. The difference feature is extracted from the preprocessed data. The extracted features are mapped to lower-dimensional feature space by using OLDA. The subspace features are used to train and test the recognition module. Classification of the test data is carried out by using the nearest neighbor classifier. The average recognition accuracy of 99.3% is achieved.

***Keywords*—** Online Handwriting, Kannada Numerals; Handwriting Recognition Module; Difference Feature; Nearest Neighbor Classifier;

I. INTRODUCTION

In the field of pattern classification, various methods have been adopted to improve the recognition performance of the pattern classifiers. Some of the parameters used to measure the performance of the pattern classifier are; recognition accuracy, requirement of system resources, and recognition speed. The main data processing modules of a typical pattern classifier are; preprocessing, feature extraction, and pattern classifier [1]. Many efforts have been made to devise different methods to fine tune each of these blocks to improve the overall performance of the recognition module.

The collected data is not suitable for the direct processing due to the noise or redundant components in it. In some cases, there is a need for the segmentation of individual patterns in the data to get the region of interest. To remove redundant components or to segment the data, various preprocessing and segmentation techniques are employed. Feature of a pattern gives salient characteristics of a pattern and helps in representing the pattern characteristics in an abstract and precise manner. Therefore, the performance of the pattern classifier depends on the extraction of relevant feature. A relevant feature is the one by removing this feature from the feature set decreases the performance of the pattern classifier [2]. Therefore, numerous efforts have been made to find the relevant optimal features to boost the performance of the recognition system. The role of pattern classifier is to assign the class label to a test pattern by matching it with the training data samples. It does by comparing the test pattern to all the training data patterns. Some of the commonly used classifiers that are used in pattern classification are; nearest neighbor, HMM, SVM, decision tree, neural networks, etc. Various techniques like classifier fusion, multi-stage classification, etc. have been introduced to improve the classification performance.

The focus of this paper is to increase the recognition accuracy and hence improve the overall performance of the online handwritten Kannada numeral recognition module with the use of better discriminating feature. A discriminating feature is the one which has the better inter-class discrimination capability and it reduces the misclassification of data patters. In this direction, numerous efforts have been made to extract features that represent the data pattern in an abstract fashion and also have the better discriminating capability. Some of the works related to the extraction of the useful features from the online handwritten data are given in the next section.

The rest of the paper is organized as follows: In section II, some of the research work carried out to extract features from the online handwritten data is presented. Information about the database used in the present work is given in section III.

Section IV highlights the various methods used for preprocessing the online handwritten numerals. The experimental results are given in section V. The concluding remarks are given in section VI.

II. RELATED WORK

In paper [3], the tangent slope angle features are extracted from the online handwriting and addresses the problem of integrating the directional and linear data into a multivariate probability density function. The curvature and the direction of writing feature at every point of online handwriting are extracted in [4]. In [5], the rotation invariant features are extracted from the online handwritten alpha-numeric characters to handle the characters which are written with some orientation. Writing with orientation either to left or right side is one of the writing styles of people. The features which represent both direction and structural information of online handwritings are discussed in [6]. The Delaunay triangulation based shape feature extraction from the online handwritten data is explained in [7].

In the present work, the online handwritten Kannada numeral recognition module is designed by extracting the difference features. The OLDA subspace algorithm [8] is used to map the difference feature to the lower-dimensional feature space.

III. DATABASE

Tablet PC is used to collect the online handwritten numeral data. The data is collected from 140 writers, who are the native people of Karnataka. Each one of them has contributed from one trial to a maximum of five trials. After annotation of

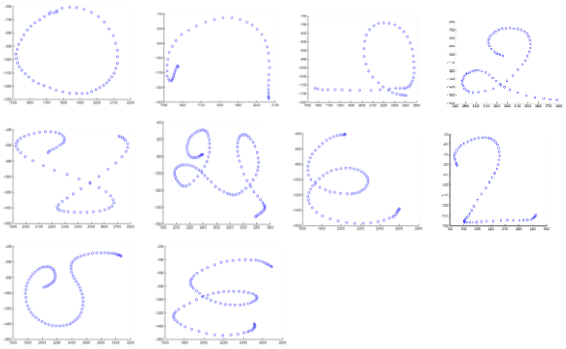

Figure 1. Sample of online handwritten Kannada numeral data samples coorespond to Indo-Arabic numerals 0 to 9.

the data, the database consists of 140 trials in each class of Kannada numeral patterns 0 to 9. Hence, there are 1400 numeral samples in the database. Fig. 1 shows the sample of online handwritten Kannada numerals 0 to 9.

IV. PREPROCESSING AND FEATURE EXTRACTION

Due to the limitations of the data collection unit and the writing styles of different writers, the collected data is not suitable for the direct processing. Therefore, the collected data is subjected to various preprocessing steps [9]. At first, the data is filtered by using the Gaussian low-pass filter to remove jitters and repeated data points. Then each of the filtered data samples is sampled using linear interpolation technique so that the all numeral data samples have 30 numbers of points. Then, they are size normalized so that every character fit into a bounding box of size one.

Let the vector D indicate the preprocessed online handwritten numeral data sample.

$$D = [d_1, d_2, d_3, \dots\dots\dots\dots d_{30}] \quad (1)$$

Where the vectors, $d_i = (p_i, q_i)$ are given by,

$$p_i = (x_i - x_{min})/(x_{max} - x_{min}) \quad (2)$$

$$q_i = (y_i - y_{min})/(y_{max} - y_{min}) \quad (3)$$

In Equations (2) and (3), the values (x_{min}, x_{max}) and (y_{min}, y_{max}) correspond to minimum and maximum values of x- and y- coordinate values of writing on the Tablet PC. From the size normalized x- and y- coordinate values, the consecutive x- and y- coordinate difference values are computed. The expressions used to compute the x- and y-difference values are given in Equations (4) and (5) respectively. The computed values are normalized so that their final values lie between 0 and 1. A vector F, comprising of the normalized difference values is used as a feature to train and test the online handwritten numeral data samples.

$$x_d(n) = x(n) - x(n-1) \quad (4)$$

$$y_d(n) = y(n) - y(n-1) \quad (5)$$

$$F = \{(x_{d1}, y_{d1})^T, (x_{d2}, y_{d2})^T, \dots, (x_{d30}, y_{d30})^T\} \quad (6)$$

V. EXPERIMENTS AND REULTS

The collected online handwritten numeral data is segmented into a disjoint set of training and testing data sets. Out of the available 140 numeral data samples per numeral pattern, 70 data samples are used for training the recognition system and the remaining 70 data samples are used to test the recognition system. The numeral data samples earmarked for training the recognition system are subjected to preprocessing and feature extraction. Based on the expressions given in Equations (4) and (5), consecutive difference values from the normalized x- and y-coordinate value are computed. A vector comprising of the normalized x- and y-difference value is framed and is used as the feature vector. These feature vectors of online handwritten numeral values are mapped to lower-dimensional feature space using OLDA. These subspace features are used to train and test the online handwritten numeral recognition module. From the simulation experiments the average

recognition accuracy of 99.3% is achieved. Table 1 gives the comparison of the recognition accuracy of the proposed online handwritten recognition numeral with the earlier works.

Table 1 reveals that the proposed online handwritten Kannada numeral recognition module has yielded the average recognition accuracy of 99.3%. This value of recognition accuracy is about 0.7% higher than that of the recognition module proposed in [12].

TABLE I. COMPARISION OF THE PROPOSED ONLINE HANDWRITTEN KANNADA NUMERAL RECOGNTION MODULE WITH THE EARLIER MODULES.

Sl. No.	Features used	Average Recognition Accuracy in %
1.	Trajectory and Deviation Features [10]	96.9
2.	DFT of Trajectory and Deviation Features [11]	98.2
3.	Writing Direction Feature [12]	98.6
4.	Difference Feature in the proposed work	99.3

VI. CONCLUSION

In this paper, the performance of the designed online handwritten Kannada numeral recognition module is presented. The consecutive difference values of normalized x- and y-coordinate values are used as feature. The subspace algorithm, OLDA is used for mapping the extracted features to the sub-space. With the nearest neighbor as the pattern classifier, the average recognition accuracy of 99.3% is achieved.

ACKNOWLEDGMENT

The work carried out is part of the project work sanctioned under Minor Research Project of University of Mysore. The author would like to thank the concerned authorizes of University of Mysore for sanctioning the minor research project. During this project, the online handwritten numeral data has been collected. The author extends his thanks to all those who actively participated in providing their handwritten numeral data samples for this project.

REFERENCES

[1] R O Duda, P E Hart, and D Stork, Pattern Classification, Second Edition, John Wiley and Sons (Asia) Pvt. Ltd., 2006.
[2] A.L. Blum, P. Langley, "Selection of relevant features and examples in machine learning," Artificial Intelligence, Vol. 97, pp. 245–27, 1997.
[3] Claus Bahlmann, "Directional features in online handwriting recogntion," Pattern Recognition, Vol. 39, pp. 115-125, 2006.
[4] S. Jaeger, S Manke, J Reichert, and Waibel, "Online handwriting recognition: the NPen++ recognizer," Int. Jl. Of Document Analysis and Recogntion, Vol. 3, pp.169-80, 2001.
[5] Basabi Chakraborty, Goutam Chakraborty, "A new feature extraction technique for on-line recognition of handwritten alphanumeric characters," Information Sciences, Vol. 148, pp. 55–70, 2002.
[6] Michael Blumenstein, XinYu Liu, Brijesh Verma, "An investigation of the modified direction feature for cursive character recognition," Pattern Recognition, Vol. 40, pp. 376-388, 2007.
[7] Wei Zeng, XiangXu Meng, ChengLei Yang, Lei Huang, "Feature extraction for online handwritten characters using Delaunay triangulation," Computers & Graphics, Vol. 30, pp. 779–786, 2006.
[8] Fengxi Song, Shuhai Liu, Jingyu Yang, "Orthogonilzed Fisher disriminant", Pattern Recognition, Vol. 38, Issue 2, pp. 311-313, 2005.
[9] M. Mahadeva Prasad, M. Sukumar, A. G. Ramakrishnan, "Divide and conquer technique in online handwritten Kannada character recognition," In the Proc. of Int. Workshop on Multilingual Optical Character Recognition Systems, 2009, Barcelona, Spain.
[10] M. Mahadeva Prasad, M. Sukumar and A.G. Ramakrishnan, "Orthogonal LDA in PCA transformed subspace," In the Proc. of 12th Conf. on Frontiers in Handwriting Recognition, pp. 172-175, 2010.
[11] M. Mahadeva Prasad, M Sukumar, "HMM based Two-Stage Classification Scheme to Improve Online Handwritten Kannada Numeral Recognition," Int. Jl. of Computer Science and Technology, Vol. 3, Issue 2, pp. 897-902, 2012.
[12] M. Mahadeva Prasad, "Writing Direction Feature based Online Handwritten Kannada Numeral Recognition," Int. Jl. of Computational Engineering Research, Vol. 9, No. 2, pp. 65-68, 2019.

Author Profile

Dr. M. Mahadeva Prasad is working as Assistant Professor in the Department of Studies in Electronics, Post Gradueate Cetnre, University of Mysore, Hassan, India. His research interests are signal processing, patterrn recogntion, audio and image processing, handwriting recognition. He has about 20 years of teaching and research experience. He is the the member of BoS and BoE of University of Mysore and other Universities of Karnataka. He is the life member of IETE, India and ISCA, India.

