ISSN (e): 2250 – 3005 || Volume, 09 || Issue, 2 || February– 2019 ||
International Journal of Computational Engineering Research (IJCER)

Writing Direction Feature based Online Handwritten Kannada Numeral Recognition

M Mahadeva Prasad
Department of Studies in Electronics, Post Graduate Centre, University of Mysore, Hassan-573226, Karnataka, India

ABSTRACT: *The paper presents the work carried out in the design of recognition module for online handwritten Kannada numerals. The Kannada script uses the decimal number system and hence there are ten symbols corresponding to ten decimal numbers. The collected online handwritten numeral data consists of 140 samples per class. Out of this data, 70 samples per class are used for training and the remaining 70 samples are used for testing. In total, 700 data samples are used for training and the remaining 700 data samples are used for testing the designed recognition module. The collected data are subjected to preprocessing. The adopted preprocessing steps are; filtering, sampling and normalization. The cosine and sine angles subtended at each point corresponding to the horizontal and vertical coordinate values of handwriting trajectory are computed. The combined cosine and sine angles are used as writing direction feature of online handwriting. The extracted features are mapped to lower dimensional feature space by using OLDA. The nearest neighbor classifier is used for pattern classification. The experiments yielded the maximum average recognition accuracy of 98.57%.*
Keywords:*Kannada numerals; online verses offline handwriting; handwriting recognition; writing direction feature; orthogonal linear discriminant analysis;*

Date of Submission: 23-02-2019 Date of acceptance:14-03-2019

I. INTRODUCTION

The characters of a script can be grouped into; alphabets, numerals, and special symbols. While alphabets are used to convey messages, numerals are used for the measurement of quantity and counting. The decimal number system is used in most of the scripts. The decimal number system consists of ten numeral values corresponding to zero to nine. These ten numeral values are represented by ten different glyphs. The scripts of different languages have their own glyphs to denote the decimal number values. The Kannada script uses decimal number system. The Kannada language is the official and administrative language of Karnataka, one of the South Indian states of India. The script uses ten glyphs to represent ten decimal numbers.

Numeral data can be either machine printed or handwritten. Handwritten numeral data is further divided into offline handwritten data and online handwritten data. The machine printed and offline handwritten data are available in the image form. Therefore, optical character recognition (OCR) methods are used to recognize these numeral data types. The online handwritten data is in the form of two-dimensional coordinates of successive points of writing. The online handwritten data may contain one or more than one strokes. A stroke is a trajectory generated from the starting of writing to the end of writing. For the recognition of machine printed and offline handwritten numeral data, OCR methods are used. For online handwritten data, online handwritten character recognition methods are used.

Handwritten character recognition works have been the active research for few decades. Numerous efforts can be found in the literature on the design and development of recognition systems for machine printed, offline handwritten, and online handwritten characters. But the work carried out for Indian scripts is less when compared to other scripts. In addition, not enough research efforts work can be found for Kannada characters when compared to other Indian scripts. The work carried out to design the recognition modules for Kannada handwritten numerals can be grouped into offline handwritten numeral recognition modules and online handwritten recognition modules. There are some efforts carried out to design the offline handwritten recognition modules. In [1], the modified quadratic recognition modules are proposed to recognize the handwritten numerals of six Indian scripts. The six Indian scripts are; Devnagari, Bangla, Telugu, Oriya, Kannada and Tamil. Similarly, recognition systems for handwritten numerals for three South Indian scripts, Kannada, Tamil, and Telugu are proposed in [2] and the recognition systems for the Kannada and English

numerals are designed in [3]. Many researchers have worked on the design and development of the recognition modules exclusively for the offline handwritten Kannada numerals. Some of them are given in the next section.
In [4], directional chain code information of contour points of characters and quadratic classifiers are used to recognize the offline handwritten Kannada numerals. In [5], the projection distance metrics and general regression neural network are used to recognize the unconstrained handwritten Kannada numerals. Character recognition system for both machine printed and offline handwritten numerals is proposed in [6]. Similar work for the recognition of handwritten Kannada numerals can be found in[7-10].
As explained in the previous sections, there are some efforts made to design and develop the recognition modules for handwritten Kannada numerals. Based on the literature survey, there are only two efforts [11] and [12] made to design the recognition modules exclusively for online handwritten Kannada numerals. In [11], the features extracted from the online handwritten Kannada numerals are transformed to PCA subspace. Then the performance of subspace algorithms LDA and OLDA are evaluated in the PCA transformed subspace. In [12], HMM based two-stage classifier is designed to improve the recognition performance of online handwritten Kannada numeral recognition system. Since both of these recognition modules require two-level feature extraction and two-stage classification, the recognition modules demand more system resources and recognition time. To overcome these limitations, in the present work, writing direction feature with single-stage classification is designed.

II. ONLINE HANDWRITTEN DATABASE

The online handwritten numeral data used in the present work is collected from the native writers of Kannada by using Tablet PC. Figure 1 shows the sample of offline handwritten Kannada numerals and the corresponding Indo-Arabic numerals. The online handwritten Kannada numeral data sample corresponding to Indo-Arabic numeral 3 is shown in Fig. 2.

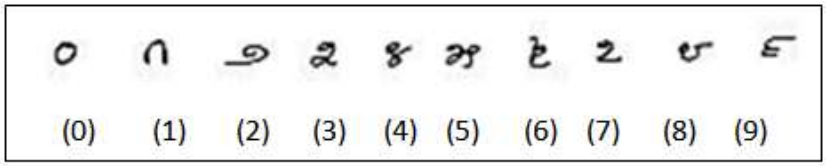

Fig. 1. Handwritten data sample of Kannada numerals and the corresponding Indo-Arabic numerals

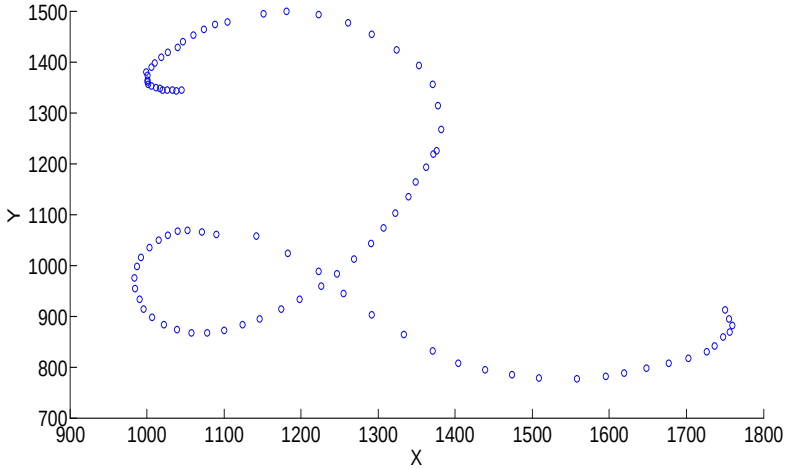

Fig. 2. Online handwritten Kannada numeral sample corresponding to Indo-Arabic numeral 3.

III. PREPROCESSING

The online handwritten Kannada numeral data consists of the x- and y-coordinate values as shown in Fig. 2. The collected data is subjected to various preprocessing stages. The data is filtered using a Gaussian low-pass filter. The filtered data is resampled in space by using linear interpolation technique so that all character samples have thirty numbers of samples. Then, they are size normalized so that every character fit into a bounding box of size [0, 1] in both horizontal and vertical directions.

IV. FEATURE EXTRACTION

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

$$P = [p_1, p_2, p_3, \ldots\ldots\ldots\ldots p_{30}] \quad (1)$$

Where the vectors, $p_i = (m_i, n_i)$ are given by,

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

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

In Equations (2) and (3), the values (x_{min}, x_{max}) and (y_{min}, y_{max}) correspond to minimum and maximum values of horizontal and vertical coordinate values. The direction feature value $(\theta_{cos}, \theta_{sin})$ is the cosine and sine angle subtended at a data point(m_i, n_i). These values of handwriting direction are computed by using relations given in Equations (4) and (5) respectively [13].

$$\theta_{cos} = \cos^{-1}\left(\frac{x_{i-1} - x_{i+1}}{d_i}\right) \quad (4)$$

$$\theta_{sin} = \sin^{-1}\left(\frac{y_{i-1} - y_{i+1}}{d_i}\right) \quad (5)$$

Where, $$d_i = \sqrt{(x_{i-1} - x_{i+1})^2 + (y_{i-1} - y_{i+1})^2} \quad (6)$$

By using the above computations, the direction feature vector F is formed as shown in Equation (7).

$$F_{23} = \{(\theta_{cos\,1}, \theta_{sin\,1})^T, (\theta_{cos\,2}, \theta_{sin\,2})^T, \ldots., (\theta_{cosN}, \theta_{sinN})^T\} \quad (7)$$

V. EXPERIMENTS AND DISCUSSION

The collected online handwritten numeral database consists of 140 handwritten data samples per numeral pattern. Therefore, the database comprises of 1400 handwritten data samples corresponding to ten Kannada numerals. To realize the writer independent online handwritten numeral recognition, the database is partitioned into a disjoint set of training and testing data sets. Out of the 140 data samples per class, first 70 data samples are used to train the recognition module and the remaining 70 data samples are used for testing the trained recognition module. Therefore, out of 1400 numeral data samples corresponding to ten numerals, 700 data samples are used for testing the remaining 700 data samples are used for testing. The simulation experiments are carried out by using MATLAB.

The collected handwritten numeral data samples are subjected to preprocessing. To eliminate the noises like jitter and repeated data points, the raw handwritten data are filtered by using Gaussian low pass filter. In order to have the common number of data points in each numeral data, the filtered data are resampling in space so that each data samples have 30 numbers of data points. To overcome the writing position variations of data samples, the resampled data sample are size normalized so that each data fits into a bounding box size of unit length. Then, using the relations given in the feature extraction section, the direction features are extracted. The extracted features are mapped to lower-dimensional feature space by using OLDA [14]. The subspace features extracted from the training data are used to train the recognition module. While testing, each of the test samples is subjected to preprocessing and feature extraction steps. The extracted features are mapped to OLDA subspace. The subspace mapped test sample is tested against the trained data models by using nearest neighbor classifier. The average recognition accuracies achieved for sine of handwriting direction feature, cosine of handwriting direction feature, and for the combination of these two features is given in Table 1. Table 1 reveals that the maximum average recognition accuracy of 98.6% is achieved with the proposed system.

In the earlier work reported in [12], the average recognition accuracy of 96.9% has been obtained. The recognition scheme requires the two-level subspace mapping. In the first level, the extracted features are mapped to PCA subspace. And in the second level, the PCA subspace features are mapped to OLDA subspace before training and testing the recognition module. In [13], to improve the recognition performance of online handwritten numeral recognition module, two-stage HMM classifier has been used and the maximum average recognition accuracy of 98.5% is achieved. In both of these modules, multiple features have been used. Compared to these two recognition modules, the numeral database used in the present work consists of more number of user data samples. In addition, the proposed system has yielded the better recognition accuracy with the single-stage classification by adopting the direction feature comprising of cosine and sine angle values of handwriting trajectory. Even though there is a marginal improvement in the average recognition accuracy in the proposed system when compared to the scheme in [13], the proposed recognition scheme is better because the system is a

single-stage classifier and it used single feature. In total, the performance of the proposed system is better when compared the earlier two schemes in terms of improved average recognition accuracy for the more number of user data samples and reduced recognition time and system resources requirements.

TABLE 1 AVERAGE RECOGNITION ACCURACY WITH 700 SAMPLES FORTRAINING AND 700 SAMPLES FOR TESTING.

Feature	Average Recognition Accuracy in %
Cosine of writing direction feature	90.7%
Sine of writing direction feature	95.9%
Combined Cosine and Sine of writing direction features	98.6%

VI. CONCLUSIONS

In this paper, the work carried out to design the recognition module for online handwritten Kannada numerals is discussed. The collected online handwritten numeral data samples are subjected to preprocessing and feature extraction. The direction features are extracted from the preprocessed numeral data samples. The extracted direction features are mapped to lower-dimensional feature space using OLDA. The nearest neighbor classifier is used to classify the test data sample. The maximum average recognition accuracy 98.57% is obtained.

ACKNOWLEDGEMENTS

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]. U. Pal, N. Sharma, T. Wakabayashi, F. Kimura, "Handwritten Numeral Recognition of Six Popular Indian Scripts," In Proc.of 9th Int. conf. on Document Analysis and Recognition ICDAR07, Vol. 2, pp.749-753, 2007.
[2]. S.V. Rajashekararadhya, and P. Vanaja Ranjan," Handwritten Numeral Recognition of Three Popular South Indian Scripts: A Novel Approach," In the Proc. of the 2nd Int. Conf. on Information Processing, pp.162-167, 2008.
[3]. B.V.Dhandra, Gururaj Mukarambi, Mallikarjun Hangarge, "Kannada and English numeral recogntion," Int. Jl. of Computer Applications, Vol. 26, No.9, pp. 17-22, 2011.
[4]. N. Sharma, U. Pal, and F. Kimura, "Recognition of Handwritten Kannada Numerals, In the Proc. of 9th Int. Conf. on Information Technology, 2006.
[5]. Basappa B. Kodada and Shivakumar K. M., "Unconstrained Handwritten Kannada Numeral Recognition," Int. Jl. of Information and Electronics Engineering, Vol. 3, No. 2, pp. 230-232, 2013.
[6]. G. G. Rajput, Rajeswari Horakeri, Sidramappa Chandrakant, "Printed and Handwritten Mixed Kannada Numerals Recognition Using SVM," Int. Jl. on Computer Science and Engineering, Vol. 02, No. 05, pp. 1622-1626, 2010.
[7]. V. N. Manjunath Aradhya, G. Hemanth Kumar and S. Noushath, "Robust Unconstrained Handwritten Digit Recognition Using Radon Transform," In the Proc. of IEEEICSCN, pp-626-629, 2007.
[8]. S.V. Rajashekararadhya, P. Vanaja Ranjan, "Support Vector Machine based Handwritten Numeral Recognition of Kannada Script", In the Proc. of IEEE Int. Conf. on Advance Computing, pp. 6-7, 2009.
[9]. Rajput, G.G., Mallikarjun Hangarge, "Recognition of Isolated Handwritten Kannada Numeral based on Image fusion method", In the Proc. of Int. Conf. on Pattern Recogntion and Machine Intelligence, Vol. 4815, pp153-160, 2007.
[10]. Dinesh Acharya U., N. V. Subba Reddy, and Krishnamoorthi Makkithaya, "Multilevel Classifiers in Recognition of Handwritten Kannada Numerals," Int. Journal of Computer and Information Engineering, Vol. 2, No. 6, pp. 1908-1913, 2008.
[11]. 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, 172-175, 2010.
[12]. M. Mahadeva Prasad, M Sukumar, "HMM based Two-Stage Classification Scheme to ImproveOnline Handwritten Kannada Numeral Recognition," Int. Jl. Computer Science and Technology, Vol. 3, Issue 2, pp. 897-902, 2012.
[13]. I. Guyon,P . Albrecht,Y. Le Cun,J. Denker,W. Hubbard. "Design of a Neural Network Character Recognizer for a Touch Terminal," Pattern Recognition, Vol.24, Issue 2, pp. 105–119,1991.
[14]. Fengxi Song, Shuhai Liu, Jingyu Yang, "Orthogonilzed Fisher disriminant", Pattern Recognition, Vol. 38, Issue 2, pp. 311-313, 2005.

M Mahadeva Prasad" Writing Direction Feature based Online Handwritten Kannada Numeral Recognition" International Journal of Computational Engineering Research (IJCER), vol. 09, no. 2, 2019, pp 65-68