International Journal of Research
Available at https://edupediapublications.org/journals
e-ISSN: 2348-6848
p-ISSN: 2348-795X
Volume 05 Issue 12
April 2018

Trajectory Feature based Online Handwritten Kannada Basic Character Recognition System

M. Mahadeva Prasad
Dept. of Studies in Electronics
Hemagangotri, University of Mysore
HASSAN-573226, India
E-mail: prasada9@gmail.com

***Abstract*—The paper presents the design of online handwritten Kannada basic character recognition system using trajectory feature. The Kannada basic character set has fifty characters. The online handwritten character data is subjected to preprocessing and feature extraction. By aligning the preprocessed character to the Cartesian coordination system, the trajectory feature is extracted. The trajectory features is comprised of distance and angle features. The features are mapped to the lower dimensional feature space using 2D-LDA subspace algorithm. The recognition module is designed based on the nearest neighbor classifier. The simulation experiments are repeated for distance, angle, and the combined distance and angle features. The maximum value of average recognition accuracy of 85.03% is achieved for the twenty one number of eigen vectors.**

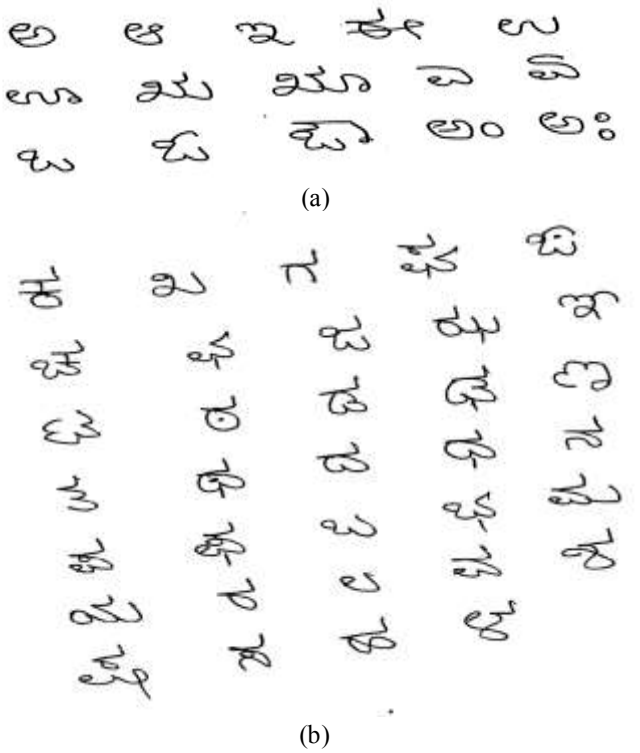
Keywords*—*Trajectory feature; online handwriting recognition; Kannada basic character recogntion;

I. INTRODUCTION

The design of an efficient online handwriting recognition system for Indic scripts is a challenging task. The huge sized character set and their complex shapes, almost similar shaped characters, and variability in writing among individuals are the constraints to realize the good recognition system. In spite of these constrains, many efforts have been made to improve the recognition performance of the recognition system. Among these efforts, relevant and discriminative feature selection [1] has been carried out to minimize the misclassification among characters. In addition to this, script depended feature selection has the added advantage to further improve the performance of the recognition system.

A trajectory or flight path is the path that a moving object follows through space as a function of time [2]. In online handwriting, the two-dimensional coordinates of successive points of writing are recorded as a function of time [3]. Therefore, the online handwriting can be treated as a trajectory traced by the pen from the starting of writing to the end of writing.

The Kannada character set is classified into basic characters, compound characters, and numerals. The Kannada basic character set consists of fifty characters. The basic character set is grouped into swaras (vowels) having sixteen characters, and vyanjanas (consonants) having thirty four characters [4]. A sample of handwritten Kannada basic characters is shown in Fig. 1.



(a) (b)

Fig. 1. Sample of Handwritten Kannada Varnamale Set. (a) Swaras (Vowels), (b) Vyanjanas (Consonants).

International Journal of Research
Available at https://edupediapublications.org/journals
e-ISSN: 2348-6848
p-ISSN: 2348-795X
Volume 05 Issue 12
April 2018

II. REPROCESSING

The online handwritten Kannada character data consists of the x- and y-coordinate values corresponding to the trajectory traced by the pen-tip movement. Before extracting the trajectory features, the character data are subjected to different stages of preprocessing. The characters are low-pass filtered using Gaussian low-pass filter to remove the noise. To ensure that each characters have thirty number of data points, the characters are re-sampled in space along the arc length using linear interpolation. Then the data are shifted and size normalized to fit in a square box of side length one [5].

Let Q represents the preprocessed character data sequence vector as given in Eqn. (1).

$$Q = [q_1, q_2, q_3, \ldots\ldots\ldots\ldots, q_m] \quad (1)$$

Where, the vectors $q_i = (a_i, b_i)^T$ represent the x- and y-coordinate values of the preprocessed and normalized character.

III. TRAJECOTRY FEATURE EXTRACTION

A trajectory feature has the distance and angle feature. To extract the trajectory feature, the preprocessed character is aligned to the Cartesian coordinate system. The pictorial illustration of computing the trajectory feature is given in Fig. 2. A vector is defined from the origin of the coordinate system to the starting point of the character writing. The magnitude of this vector and the angle formed by this vector with the x-axis are computed using Eqns. (2) and (3). The process is repeated for all the points in the character.

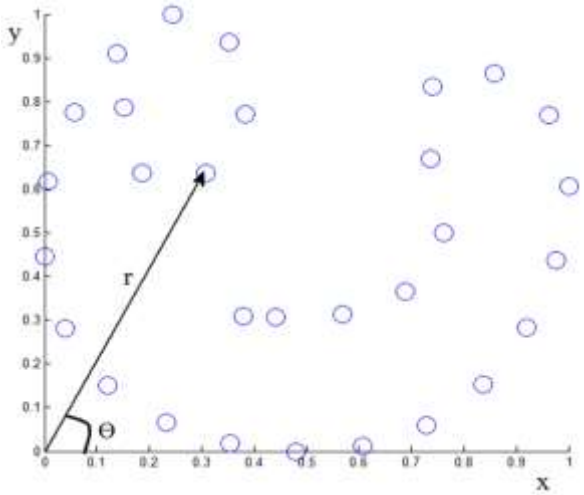


Figure 2 Trajectory feature extraction.

$$r_n = \sqrt{x_n^2 + y_n^2} \quad (2)$$

$$\theta_n = \cos^{-1}(x_n / r_n) \quad (3)$$

Let F_1, F_2 and F_3 be the distance, angle, and the combined distance and angle feature vectors.

$$F_1 = \{r_1, r_2, r_3, r_4, \ldots\ldots\ldots\ldots\ldots\ldots, r_N\} \quad (4)$$

$$F_2 = \{\theta_1, \theta_2, \theta_3, \theta_4, \ldots\ldots\ldots\ldots\ldots\ldots, \theta_N\} \quad (5)$$

$$F_3 = \{(r_1, \theta_1)^T, (r_2, \theta_2)^T, (r_3, \theta_3)^T, .., (r_N, \theta_N)^T\} \quad (6)$$

IV. EXPERIMENTS AND RESULTS

The online handwritten Kannada basic characters are collected using Tablet PC. The collected data is divided into two sets that are belonging to different and disjoint sets of writers. From the available 5300 total numbers of character data samples, 3750 samples are used for training the remaining 1550 samples are used for testing. Both training and test data samples are subjected to preprocessing. The total number of sampling points is fixed to thirty numbers. After preprocessing the characters, the trajectory features are extracted. The distance, angle, and the combined distance and angle features are mapped to lower dimensional feature space using 2D-LDA. The recognition module is trained with this subspace features. The test samples are classified using the nearest neighbor classifier.

The simulation experiments are carried out using MATLAB installed in a PC with Intel Core i3 processor and 8 GB of RAM. By fixing the number of eigen vectors to thirty, independent experiments are carried out for distance, angle, and the combined distance and angle features. The average recognition accuracy and time taken by the system to classify 1550 test samples are given in Table I. The experimental results show that the maximum recognition accuracy of 84.19% is resulted for the combined distance and angle feature. To improve the performance of the recognition system, the experiments are repeated with the combined distance and angle feature by varying the number of eigen vectors in steps of five. During this series of experiments, the maximum average recognition accuracy of 84.9% is achieved for the twenty eigen vectors. The results of these experiments are given in Table II.

TABLE I PERFORAMNCE OF THE RECOGNTION SYSTEM FOR DIFFERENT FEATURES WITH THIRTY NUMBERS OF EIGEN VECTORS.

SL. NO.	FEATURE	AVERAGE RECOGNITION ACCURACY (%)	TEST TIME IN SECONDS
1.	DISTANCE	72.64	91.26
2.	ANGLE	71.87	91.10
3.	COMBINED DISTANCE AND ANGLE	84.19	101.28

TABLE II PERFORAMNCE OF THE RECOGNTRION SYSTEM FOR THE VARIED NUMBER OF EIGEN VECTORS WITH COMBINED DISTANCE AND ANGLE FEATURE.

SL. NO.	NUMBER OF EIGEN VECTORS	AVERAGE RECOGNITION ACCURACY (%)	TEST TIME IN SECONDS
1.	5	75.68	77.61
2.	10	83.16	84.08
3.	15	84.71	89.95
4.	20	84.90	94.94
5.	25	84.32	98.53
6.	30	84.19	101.28

The simulation results in Table II reveals that the average recognition accuracy of 84.71%, 84.9%, and 84.32% are obtained for fifteen, twenty, and twenty five eigen vectors respectively. In order to further tune the recognition engine for the number of eigen vectors, the experiments are repeated by varying the eigen vectors from fifteen to twenty five in steps of one. The results of this experiment are given in Table III. The results indicate that the maximum average recognition accuracy of 85.03% is obtained for the twenty one eigen vectors, and the system has taken 95.25 seconds to test 1550 character data samples.

TABLE III PERFORAMNCE OF THE RECOGNTRION SYSTEM FOR THE VARIED NUMBER OF EIGEN VECTORS FROM FIFTEEN TO TWENTY FIVE WITH THE COMBINED DISTANCE AND ANGLE FEATURE.

SL. NO.	NUMBER OF EIGEN VECTORS	AVERAGE RECOGNITION ACCURACY (%)	TEST TIME IN SECONDS
1.	15	84.71	89.95
2.	16	84.72	90.36
3.	17	84.77	91.57
4.	18	84.90	92.85
5.	19	84.83	93.75
6.	20	84.90	94.94
7.	21	85.03	95.25
8.	22	84.97	96.12
9.	23	84.77	97.22
10.	24	84.58	97.89
11.	25	84.32	98.53

V. CONCLUSIONS

In this paper, the design of the online handwritten Kannada basic character recognition system is disused. The recognition system is designed based the trajectory feature and 2D-LDA subspace algorithm. The simulation experiments are carried out for distance, angle, and the combined distance and angle feature. Among these three features, the combined distance and angle feature has given the maximum recognition accuracy of 84.1 %. With the combined distance and angle feature, and for the varied number of eigen vectors, the recognition system has yielded the maximum average recognition accuracy of 85.03% for twenty one eigen vectors.

VI. REFERENCES

[1] A Blum and P Langley, Selection of relevant features and examples in machine learning, Artificial Intelligence, Vol. 1-2, 1997, pp.245-271.

[2] https://en.wikipedia.org/wiki/Trajectory

[3] Plamondon, R and S N Srihari, On-line and off-line handwriting recognition: A comprehensive survey, IEEE Trans. on Pattern Analysis and Machine Intelligence, 22 (1), pp. 63-84, 2000.

[4] Ferdinand Kittel, A Grammar of the Kannada Language: Comprising the Three Dialects of the Language, Asian Educational Services, 1993.

[5] M. Mahadeva Prasad, M. Sukumar, A. G. Ramakrishnan, Divide and conquer technique in online handwritten Kannada character recognition, Int. Workshop on Multilingual Optical Character Recognition Systems (MOCR-2009), 2009, Barcelona, Spain.