International Journal of Image Processing and Vision Science

Volume 1 | Issue 1 Article 4

July 2012

Handwritten Kannada Vowels and English Character Recognition System

B. V. Dhandra
Department of P.G.Studies and Research in Computer Science, Gulbarga University, Gulbarga, Karnataka, dhandra_b_v@yahoo.co.in

Gururaj Mukarambi
Department of P.G.Studies and Research in Computer Science, Gulbarga University, Gulbarga, Karnataka, gmukarambi@gmail.com

MALLIKARJUN. HANGARGE
Department. of Computer Science, Karnatak Arts, Science and Commerce College, Bidar, India, mhangarge@yahoo.co.in

Recommended Citation

Dhandra, B. V.; Mukarambi, Gururaj; and HANGARGE, MALLIKARJUN. (2012) "Handwritten Kannada Vowels and English Character Recognition System," *International Journal of Image Processing and Vision Science*: Vol. 1 : Iss. 1 , Article 4.
DOI: 10.47893/IJIPVS.2012.1003
Available at: https://www.interscience.in/ijipvs/vol1/iss1/4

Handwritten Kannada Vowels and English Character Recognition System

B.V.Dhandra , Gururaj Mukarambi
Department of P.G.Studies and Research in Computer Science, Gulbarga University, Gulbarga, Karnataka
dhandra_b_v@yahoo.co.in
gmukarambi@gmail.com

Mallikarjun Hangarge
Department of Computer Science, Karnatak Arts, Science & Commerce College, Bidar, Karnataka.
mhangarge@yahoo.co.in

***Abstract*— In this paper, a zone based features are extracted from handwritten Kannada Vowels and English uppercase Character images for their recognition. A Total of 4,000 handwritten Kannada and English sample images are collected for classifications. The collected images are normalized into 32 x 32 dimensions. Then the normalized images are divided into 64 zones and their pixel densities are calculated, generating a total of 64 features. These 64 features are submitted to KNN and SVM classifiers with 2 fold cross validation for recognition of the said characters. The proposed algorithm works for individual Kannada vowels, English uppercase alphabets and mixture of both the characters. The recognition accuracy of 92.71% for KNN and 96.00% for SVM classifiers are achieved in case of handwritten Kannada vowels and 97.51% for KNN and 98.26% for SVM classifiers are obtained in case of handwritten English uppercase alphabets. Further, the recognition accuracy of 95.77% and 97.03% is obtained for mixed characters (i.e. Kannada Vowels and English uppercase alphabets). Hence, the proposed algorithm is efficient for the said characters recognition. The proposed algorithm is independent of thinning and slant of the characters and is the novelty of the proposed work.**

Keywords- DIA, OCR, KNN, SVM

I. INTRODUCTION

The document images can be classified into printed document images, handwritten document images and both. The analysis of such documents consists of number of paragraphs, columns, lines and words, scripts in document images and so on. The recognition of handwritten scripts of a document is one of the challenging tasks due to diversified style of writing. If a document contains only one script, then the corresponding OCR system can be used for document image processing. If a document contains more than one script then individual OCR systems may be used for the identification of the script/language for their processing. Now the effort is on to develop a multilingual OCR system for automatic reading and processing of the multilingual documents. In this direction a small amount of work is carried out in the Indian context. Hence, this has motivated to consider the study of handwritten Kannada vowels and English uppercase alphabets recognition system as the initial work to meet the objective of processing bi-lingual (Kannada & English) documents. In the literature many feature extraction methods are used at various levels for handwritten character recognition for various scripts/languages such as Fourier Descriptors, Shape descriptors, Spatial, Discrete Cosine Transform, Random Transform, Central Moments, Zernike Moments, Zone, Structural, Statistical, Optical Depth Decision tree etc. Hence in the proposed study an attempt is made to use the zone based features for the recognition of handwritten Kannada vowels and all the 26 uppercase characters of the English. In the following a brief account of the work carried out in the literature is presented.

Dhandra et al [1] have proposed spatial features for handwritten Kannada and English Character Recognition, and have achieved the recognition accuracy of 90.01% for handwritten Kannada vowels and 91.04% for handwritten English uppercase alphabets. Phokharatkul et al [2] have proposed Ant-Miner Algorithm for handwritten Thai Character Recognition and have reported the recognition accuracy of 82.07%. R.M. Suresh et al [3] have proposed a fuzzy technique for Tamil handwritten character recognition and he reported recognition accuracy of 94%. Teng Long et al [4] have used Dynamic Time Warping (DTW) algorithm for handwritten English uppercase alphabets collected from camera user interface and reported a recognition accuracy of 97.3%, 98.6% and 98.8% with 16, 32 and 256 normalized discrete angle values. Ertugrul Saatci et al [5] have proposed Multiscale Handwritten Character Recognition Using CNN Image Filters for handwritten English uppercase alphabets and he has exhibited a recognition accuracy of 93%. Velappa Ganapathy et al [6] have proposed a method of Multiscale Neural Network Training Technique for handwritten English uppercase alphabets and he reported recognition accuracy of 85%. Dayashankar Singh et al [7] have proposed a directional features for Hindi and English characters and he reported a recognition accuracy of 97% with 12 directional input. Dharamveer Sharma et al [8] have proposed zone based features for isolated handwritten Gurumukhi script and reported recognition accuracy of 72.54%. S.Arora et al [9] have proposed a chain code features for handwritten Devnagari characters and reported recognition accuracy of 98.03%. From the literature survey,

it is clear that, still handwritten character recognition is an active area of research to develop multilingual OCR System for the Indian languages. This addresses the need for development of single algorithm for handwritten Kannada and English Character recognition system with reference to Karnataka state. Hence, proposed work in this paper is the step towards it.

This paper is organized as follows, Section I contains the introduction part and Section II describes data collection and preprocessing details. Section III deals with the feature extraction procedure. Experimental results and discussion are the subject matter of Section IV and Section V contains conclusion.

II. DATA COLLECTION/PREPROCESSING

To validate and verify the results of the proposed algorithm the standard database for Kannada script is not available; hence we have created our own Handwritten Kannada and English data sets from different professionals belonging to Schools, Colleges, Government Departments, Lawyers, Doctors and also aged persons in Cities and Villages. The collected data set contains a multiple lines of Kannada and English characters. These Kannada and English characters are scanned and manually segmented, then stored in bmp file format. These segmented samples may contain noise due to print quality, scanner etc. The noises are removed by using morphological opening operations. The noise free Kannada and English sample images are normalized into 32 x 32 dimensions. These normalized images are used for the feature extraction. A total of 1400 handwritten Kannada Vowel samples and 2600 handwritten English uppercase alphabet samples are used for the experiment. The Fig.1 and Fig.2 shows the sample data set of handwritten Kannada Vowels and English uppercase alphabets and Fig.3 shows the overall methodology.

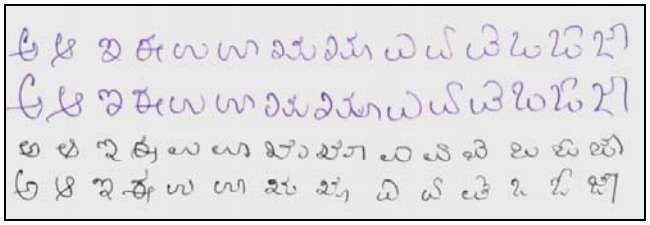

Fig.1 Sample dataset of Kannada Vowels

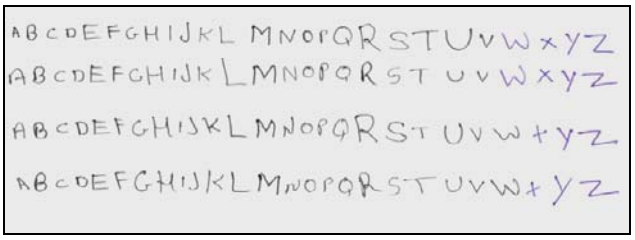

Fig. 2 Sample dataset of English uppercase Alphabets

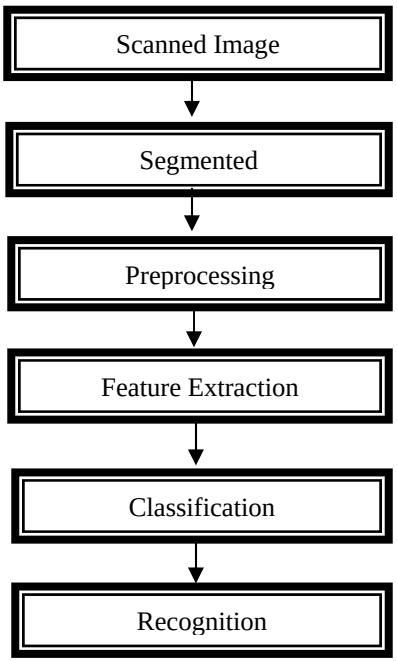

Fig.3 Methodology for the proposed problem

III. FEATURE EXTRACTION

The feature extraction is the important component of any recognition process, since the classification/recognition accuracy is depending on the features. Here the density based zone features are used for handwritten Kannada Vowels and English uppercase alphabets for their recognition. All preprocessed images are normalized into 32 x 32 dimensions. The normalization is carried out by bilinear technique. Then normalized image is divided into 64 zones and pixel density is calculated for each zone, there by generating 64 features. These 64 features are submitted for classification of images. The experiment is also carried out by considering large size zones, but it failed to capture the local information of the image. Hence we have fixed the size as 8 x 8 for experimentation as the optimal size. It provides the better results as compared to the large zone size results (4 x 4).

Algorithm: Density based Zone Features for handwritten Kannada Vowels and handwritten English uppercase Alphabets.

Begin

1. The preprocessed image is divided into 64 zones.
2. The pixel density is calculated for each of 64 zones.
3. For empty zones, zero value are assigned.
4. Generated features are submitted for KNN and SVM Classifiers with 2 Fold cross validation method to recognize the underlying image as the Kannada vowel and English alphabet.

End

The Fig.4 shows the sample Kannada Vowel image divided into 8 x 8 zones

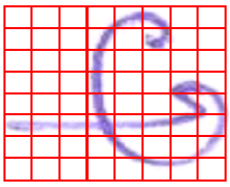

Fig. 4 Sample Image of Zone Size 8 x 8

IV. Experimental Results and Discussion

The KNN and SVM classifiers are used to classify 4,000 sample images of handwritten Kannada Vowels and English uppercase alphabets. The Table 1 shows the recognition accuracy (in %) of handwritten Kannada vowels with input of handwritten Kannada vowel and Table 2 shows the recognition accuracy of handwritten English uppercase alphabets with input of handwritten English uppercase alphabet and Table 3 shows the recognition accuracy for mixture input of handwritten Kannada vowels and handwritten English uppercase alphabets. The results exhibited in the table are encouraging for the handwritten Kannada and English characters recognition.

Table 1 Average Percentage of Recognition accuracy for handwritten Kannada Vowels with KNN (K=1) and SVM Classifiers with input of Kannada Vowel

Training samples =700, Test samples =700 Number of features = 64				
Handwritten Kannada Vowel	No. of sample Trained	No. of Sample Tested	Percentage of Recognition Accuracy With KNN	Percentage of Recognition Accuracy With SVM
	50	50	92.00	92 .00
	50	50	94.00	96.00
	50	50	96.00	98.00
	50	50	91.00	97.00
	50	50	94.00	99.00
	50	50	97.00	97.00
	50	50	98.00	98.00
	50	50	94.00	98.00
	50	50	85.00	94.00
	50	50	89.00	93.00
	50	50	87.00	94.00
	50	50	96.00	99.00
	50	50	88.00	94.00
	50	50	97.00	95.00
Average Percentage of Recognition accuracy			92.71	96.00

From Table 1, it is observed that the classification accuracy for vowels , and are little less. It is due to the fact that these three vowels are similar in shape.

Table 2 Average Percentage of Recognition accuracy for handwritten English uppercase alphabets with KNN (K=1) and SVM Classifiers with input of English alphabet

Training samples =1300, Test samples =1300, Number of features = 64				
Handwritten English Alphabet	No. of Sample Trained	No. of Sample Tested	Percentage of Recognition Accuracy With KNN	Percentage of Recognition Accuracy With SVM
A	50	50	89.00	97.00
B	50	50	95.55	97.77
C	50	50	100.00	99.00
D	50	50	95.00	96.00
E	50	50	96.00	98.00
F	50	50	96.66	97.77
G	50	50	100.00	100.00
H	50	50	100.00	100.00
I	50	50	100.00	100.00
J	50	50	100.00	100.00
K	50	50	97.00	95.00
L	50	50	99.00	99.00
M	50	50	98.99	98.99
N	50	50	94.79	97.92
O	50	50	99.00	99.00
P	50	50	99.00	99.00
Q	50	50	95.65	95.65
R	50	50	99.00	99.00
S	50	50	100.00	100.00
T	50	50	97.92	98.96

U	50	50	97.98	97.98
V	50	50	98.00	99.00
W	50	50	97.98	97.98
X	50	50	94.00	100.00
Y	50	50	98.68	98.68
Z	50	50	96.00	93.00
Average Percentage of Recognition accuracy			97.51	98.26

Table 3 Recognition accuracy for handwritten Kannada vowels and English uppercase alphabets of mixed characters input

Handwritten Training samples = 2000, Handwritten Test samples = 2000, Number of features = 64				
Mixed Kannada & English Chars.	No. of Sample Trained	No. of Sample Tested	Recognition Accuracy with Knn (k=1) in %	Recognition Accuracy With SVM in %
	50	50	89.11	92.08
	50	50	88.12	97.03
	50	50	97.03	99.00
	50	50	90.01	94.05
	50	50	94.06	96.04
	50	50	98.02	96.04
	50	50	97.03	99.00
	50	50	96.00	98.00
	50	50	93.00	94.00
	50	50	89.11	91.09
	50	50	90.01	88.12
	50	50	96.04	98.02
	50	50	91.09	90.01
	50	50	92.08	96.04
A	50	50	88.00	93.00
B	50	50	95.00	97.78
C	50	50	100.00	100.00
D	50	50	94.00	96.00
E	50	50	99.00	99.00
F	50	50	94.44	97.78
G	50	50	100.00	100.00
H	50	50	98.88	100.00
I	50	50	100.00	100.00
J	50	50	100.00	100.00
K	50	50	98.00	99.00
L	50	50	99.00	99.00
M	50	50	100.00	100.00
N	50	50	98.96	96.88
O	50	50	98.00	99.00
P	50	50	99.00	100.00
Q	50	50	96.74	94.57
R	50	50	97.00	99.00
S	50	50	99.00	100.00
T	50	50	97.92	98.96
U	50	50	98.99	96.97
V	50	50	98.00	98.00
W	50	50	96.97	97.98
X	50	50	88.00	97.00
Y	50	50	98.68	98.68
Z	50	50	96.00	94.00
Average Percentage of Recognition Accuracy			95.77	97.03

V. CONCLUSION

In this paper, the single OCR algorithm works for both the characters. The proposed features (i.e. zone based) are the potential features for recognition of handwritten Kannada vowels and English uppercase characters. The recognition accuracy is 92.71% for KNN and 96% for SVM Classifiers with 2 fold cross validation for handwritten Kannada vowels and 97.51% for KNN and 98.26% for SVM Classifier with 2 fold cross validation for uppercase English characters. The average recognition accuracy is 95.77% for KNN and 97.03% for SVM Classifiers for an input from the mixed images. The recognition accuracy is considerably high and the proposed method is independent of thinning and slant of the characters. The extension of the proposed work is in progress to meet the objective of Bilingual (Kannada and English) OCR System.

ACKNOWLEDGEMENT

This work is supported by UGC, New Delhi under Major Research Project grant in Science and Technology, (F.No-F33-64/2007 (SR) dated 28-02-2008). Authors are grateful to UGC for their financial support. We are also thankful to Dr.P.S.Hiremath and Dr.G.G.Rajput.

REFERENCES

[1] B.V.Dhandra, Mallikarjun Hangarge, Gururaj Mukarambi, "Spatial Features for Handwritten Kannada and English Character Recognition System", Special Issue on RTIPPR-10, International Journal of Computer Applications (IJCA), pp.146-150, Aug-2010.

[2] P. Phokharatkul, K. Sankhuangaw, S. Somkuarnpanit, S. Phaiboon, and C. Kimpan, "Off-Line Hand Written Thai Character Recognition using Ant-Miner Algorithm", World Academy of Science, Engineering and Technology 8, Year-2005.

[3] R.M. Suresh, "Printed and Handwritten Tamil Characters Recognition Using Fuzzy Technique", Proceedings of the International Multi Conference of Engineers and Computer Scientists 2008 Vol II, MECS 2008, 19-21 March, 2008, Hong Kong.

[4] Teng Long1, Lian-Wen Jin, Li-Xin Zhen , Jian-Cheng Huang, "one Stroke Cursive Character Recognition Using Combination of Directional and Positional Features", ICASSP-2005,IEEE,P.No v 449 - 452.

[5] Ertugrul Saatci and Vedat Tavsanoglu, "Multiscale Handwritten Character Recognition Using CNN Image Filters", IEEE-2002.

[6] Velappa Ganapathy, and Kok Leong Liew, "Handwritten Character Recognition Using Multiscale Neural Network Training Technique", World Academy of Science, Engineering and Technology, 39 2008.

[7] Dayashankar Singh, Sanjay Kr.Singh, Maitreyee Dutta, "Hand Written Character Recognition Using Twelve Directional Feature Input and Neural Network", International Journal of Computer Applications (0975 – 8887)Volume 1 – No. 3,2010.

[8] Dharamveer Sharma, Puneet Jhajj, "Recognition of Isolated Handwritten Characters in Gurumukhi Script", International Journal of Computer Applications, Volume 4, No.8, August 2010.

[9] S.Arora, D.Bhattacharjee, M.Nasipuri , D.K.Basu , M.Kundu, "Application of Statistical Features in Handwritten Devnagari Character Recognition", International Journal of Recent Trends in Engineering, Vol 2, No. 2, Academy Publisher, November 2009.

[10] B.V.Dhandra, Mallikarjun Hangarge, Gururaj Mukarambi, "Spatial Features for Multi-Font/Multi-Size Kannada Numerals Recognition", International Conference on Communication, Computation, Control and Nano Technology (ICN-2010), Bhalki, Bidar, Karnataka.

[11] B.V.Dhandra, Gururaj Mukarambi, Mallikarjun Hangarge," Zone Based Features for Handwritten and Printed Mixed Kannada Digits Recognition", International Conference on VLSI, COMMUNICATION & INSTRUMENTATION, ICVCI – 2011, Held at Kottayam, Kerala, India.