International Journal on Recent and Innovation Trends in Computing and Communication ISSN: 2321-8169
Volume: 4 Issue: 9 108 - 111

Offline Handwritten Kannada Numerals Recognition

Sushritha S
Department of Computer Science & Enginerring
CIT, Gubbi
Tumkur, Karnataka, India
india.sush@gmail.com

N Lohitesh Kumar
Department of Computer Science & Enginerring
CIT, Gubbi
Tumkur, Karnataka, India
Lohiteshkn@gmail.com

***Abstract*—** Handwritten Character Recognition (HCR) is one of the essential aspect in academic and production fields. The recognition system can be either online or offline. There is a large scope for character recognition on hand written papers. India is a multilingual and multi script country, where eighteen official scripts are accepted and have over hundred regional languages. Recognition of unconstrained hand written Indian scripts is difficult because of the presence of numerals, vowels, consonants, vowel modifiers and compound characters. In this paper, recognition of handwritten Kannada numeral characters is implemented and the different Wavelet features are used as feature extraction in this paper. The zonal densities of different region of an image have been extracted in the database. The database consists of 50 samples of each Kannada numeral character. For classification, the K-Nearest Neighbor method is used. Recognition accuracy of 88% has been achieved.

Keywords-*Optical Character Recognition, Handwritten Kannada Numerals, Discrete Wavelet Transform, K-Nearest Neighbour, Zonal Densities*

*****

I. INTRODUCTION

Handwritten character recognition is one of the important research work in pattern recognition. Character recognition is a process of identifying a character from an image of a handwritten or printed text.

Handwritten character recognition can be classified into two types:

1. Offline recognition
2. Online recognition

In online recognition, the character has been recognized as soon as the character has been written or printed. On the other hand, in offline handwritten character recognition the character has been written or printed first, and later on recognition has been performed on the written documents, but the performance is directly dependent upon the quality of the input documents.

In this paper the offline character recognition has been performed for the Kannada numerals. The recognition of handwritten characters is very difficult compared to printed characters. There are many external and internal problems arises for an OCR system for handwritten characters.

The external problems are related to the variation in the slants, shapes of the characters and also different styles of writing. Due to similarity and dissimilarity between different handwritten characters, there is a chance of recognizing wrong characters. The internal problem in an OCR system is of because of the distortion in the character images during image acquisition and also due to noise disturbance and degraded and broken characters images which reduces the accuracy of the offline handwritten character recognition. The ten Kannada numerals are shown in Fig.1.

Decimal no.	0	1	2	3	4	5	6	7	8	9
Kannada Numeral	೦	೧	೨	೩	೪	೫	೬	೭	೮	೯

Fig.1: Kannada Numerals

The handwritten recognition of Kannada numerals finds application in automatic processing of various handwritten forms in various government departments and organizations, and also in digitization of old manuscripts etc.

II. RELATED WORK

Character recognition systems generally involve two steps, first is feature extraction which contains coefficients of the numeral called features and the second is classification method in which decision is made and the class value is assigned to the input numeral. The discussion on literature survey is presented in this section under different approaches for feature extraction and classification techniques.

In a work on two stage classification approach for Kannada numerals which was presented by Sandhya Arora [1], the first stage uses structural properties like shirorekha, spine in character and second stage exploits some intersection features of characters which are fed to a feed forward neural network. The Simple histogram based method which does not work well for finding shirorekha, vertical bar (Spine) in Devanagri characters. So a new technique which is based on differential distance is used to find a near straight line for shirorekha and spine is used. This method has been tested for 50000 samples and accuracy of 89.12% is achieved.

M.L.M Karunanayaka, C.A Marasinghe and N.D Kodikara [2] have presented the work on Thresholding, Noise Reduction and Skew correction of Sinhala Words. The paper introduces a

method called novel skew detection which is based on least square method and also robust indirect skew correction method of unconstrained cursive Sinhala words. The threshold selection is used with the combination of above three methods such as NIR, QIR, and analyzing gray level intensity distribution. The thresholding selection method is used to find accurate separation point of background and foreground of the gray level image. The proposed thresholding algorithm is a combination of selected threshold algorithms such as Native integral ratio technique (NIR), Quadratic integral ratio technique (QIR) and gray level intensity distribution of given image. To reduce the noise in Sinhala words, the Median filtering algorithm and connected component analysis method are used.

In the paper over 700 patterns in Sinhala [3] real postal addresses in NSF database is used for testing and the accuracy of 97.2% is achieved. The database consists of 500 patterns, which is written by undergraduate students. This pattern is used for testing and the accuracy of 99.6% is reported. The average accuracy of 98.4% was achieved using the above two types of patterns. In the proposed system two different styles of handwritings such as the real postal addresses (RPA) and words written by the students of the same education level (WSEL) is used for testing purposes. The Overall accuracy rate of 98.4% is achieved.

Mamatha H. R, Srikant Murthy K, Sudan S, Vinay G Raj and Sumukh S Joishas [4] discussed an OCR system to recognize Kannada Numerals. Fan Beam projection a variation of Radon transform is used for extracting the features of the Kannada numerals and the classification method used is Nearest neighbour classifier. The accuracy of recognition rate of 86.29% is achieved

An Offline Handwritten Kannada Vowels Recognition using KNN Classifier is presented by Rakesh Rathi, Ravi Krishan Pandey & Vikas Chaturvedi, Mahesh Jangid [5]. Handwritten documents or papers or any surface which can be scanned using scanner will be used as input for offline handwritten recognition system. Before applying classifier, feature extraction is accomplished for extracting the feature points (FP) i.e. also known as division points (DP). In this paper, the recursive dub division technique is used as feature extraction method, which is first time implemented on Kannada vowels. K-NN classifier is functioned for the learning and the testing phases, through which the recognition is done and achieve the high performances in terms of recognition rate, pre-processing and classification speed.

In a research paper on recognition of handwritten Tamil characters by Jagadeesh and Prabhakar [6] proposes to use Hidden Markov Model which is suitable for recognition of handwritten character. In the paper, the features like ligatures and concavity are used to determine segmentation points after the segmentation of the line clustering technique which is used for segmenting the words. Vectorization process is performed on the segmented character after thinning has been done. Horizontal and vertical information about the character will be taken from Vectorization process. HMM is used to classify the characters.

III. NUMERAL RECOGNITION SYSTEM

Handwritten Optical Character Recognition system for Kannada numerals consists of 5 different stages which are shown below.

1. Image acquisition
2. Pre-processing
3. Feature extraction using different wavelet transforms
4. Classification
5. Recognized character

The block diagram of the proposed offline Kannada numeral recognition system is shown in Fig.2.

The Kannada isolated numerals written on paper is scanned using a flatbed scanner/2560x1536 resolution camera to capture the document as a digital image and is subjected to pre-processing. In pre-processing stage, the noise in the image is eliminated using median filter.

Further the pre-processed image is binarized using thresholding method**.** The image is then normalized to 32 * 64 pixels to make the image representation invariant to size.

For a given numeral, two sets of features: statistical features and wavelet features are extracted. The numeral image is divided into 16 zones of size 8 * 16. The mean zonal densities are determined to extract 16 features which forms the statistical feature set. Wavelet features are extracted by applying Discrete Wavelet Transform to the row count vector obtained from the normalized image. The combined features of structural and wavelet features serve as feature set for the image.

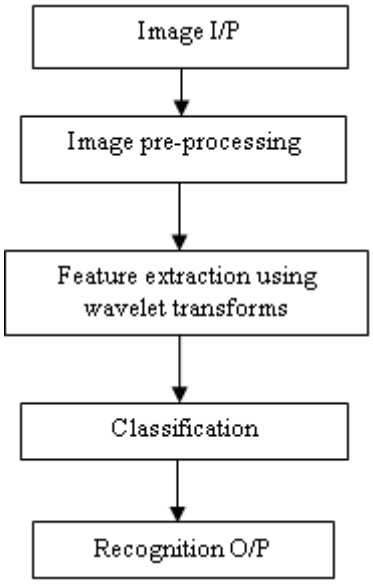

Fig.2: Block diagram of Handwritten Kannada Numeral Recognition

A brief overview of fundamentals of wavelet analysis and the wavelet transforms are described in the following section before introducing the method of wavelet based feature extraction method in detail.

A. Extraction of Wavelet Features

The pre-processed image is used as input to extract the feature. Wavelet transform is used as method for feature extraction in the proposed system. Wavelets are localized basis functions which are translated and dilated versions of some fixed wavelet. The image is decomposed into different frequency bands which are obtained by successive low-pass and high-pass filtering of the signal and down-sampling the coefficients after each filtering. In this work, the two families of discrete wavelet transforms i.e., Daubechies are used to extract features of an image.

The following steps are followed for feature extraction of a numeral:

1. Number of black pixels along each row of the binarized image has been determined to form a 32 sized vector.
2. 1D wavelet transform on row count vector has been applied to get 17 approximation coefficients (A) and detailed coefficients (D).
   [A,D] = dwt (Row_count_vector)
3. Number of black pixels along each column has been counted to form a 64 sized vector.
4. 1D wavelet transform on column count vector has been applied get 32approximation coefficients (M) and detailed coefficients (N).
   [M,N] = dwt (Column_count_vector);

Then the approximation coefficients A and M have been directly taken as wavelet features. Thus, from each numeral image 16 structural features and of 49 wavelet features are extracted to from a combined feature vector of 65 elements. Further these features elements are given as input to the KNN classifier for classification.

B. KNN Classifier

Nearest-Neighbor Classifier is an effective technique used for classification problems in which the pattern classes exhibits a reasonably small degree of variability. The k-NN classifier is based on the assumption that the classification of an instance which are most similar to the classification of other instances that are nearby in the vector space. It works by calculating the distances between the input patterns with the training pattern. A k-Nearest-Neighbor classifier takes into only the k nearest prototypes to the input pattern, and the majority of the class values of the k-Neighbors determine the decision. In the k-Nearest-Neighbor classification, the distance between the features of the test sample and the features of each and every training sample is computed. The class of majority among the k-Nearest training samples is based on the minimum distance criteria.

1. All instances correspond to points in an n-dimensional Euclidean space
2. Classification is delayed till a new instance arrives
3. Classification done by comparing feature vectors of the different points
4. Target function may be discrete or real-valued

IV. EXPERIMENTAL RESULTS

The developed OCR system was first tested with 10 numerals. Combined statistical and wavelet features of test patterns corresponding to 50 samples from each of 10 classes are extracted and presented to the input of pre-trained KNN classifier. KNN classifier was trained with 50 samples of each class of numerals. A good recognition rate 84.2% was obtained.

Owing to the successful recognition result by the developed OCR system, full set of 10 Kannada numerals were considered next. A KNN was trained with 50 samples of each of 10 numerals totalling of 500 numerals. Features extracted from the test data set were given input to the pre-trained Knn classifier. A maximum recognition rate of 84% was obtained as listed in Table 1.

TABLE I
RECOGNITION RATE OF KANNADA NUMERALS

Numeral Set	Recognition Rate in%
Numerals	84

To study the effect of choice of wavelet family used for representing the numeral two wavelet families (Haar and Daubechies) from orthogonal wavelet system are considered for experimentation. The recognition rate obtained from the system is given Table 2 and is shown in Fig.3. Features extracted from Daubechies wavelet out performed as these wavelets are compactly supported in the time domain and have good frequency domain decay.

The performance of the handwriting recognition system was analyzed by varying the number of training samples. The recognition rate of the system for samples varying from 10 to 40 is given in Table 3.

TABLE II
RECOGNITION RATE FOR DIFFERENT WAVELET FAMILIES

Wavelet Families	Recognition Rate in %
Haar	70
DB1	84

Fig.3: Recognition Rate for different Wavelet Families

TABLE III
RECOGNITION RATE FOR DIFFERENT NUMBER OF SAMPLES

No of samples	Recognition rate in %
10	74.1
20	78.4
30	84.2
40	84.4

It was found that the recognition rate increases with the increasing number of samples as shown in Fig.4. An optimal sample size for training was found to be 50. Further which by increasing the samples the recognition rate did not improved significantly.

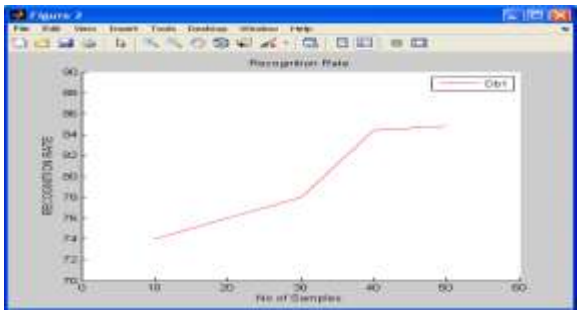

Fig.4: Number of Samples V/S Recognition Rate

V. CONCLUSION

The proposed work explored the successful development and implementation of an efficient off-line Kannada numeral recognition system based on wavelet features for the recognition of Kannada numerals. A good recognition rate of 84% for kannada numerals was obtained. The success is due to robust feature extraction scheme and the Knn Classification generalization capability. Following are the important outcome of our system development. As wavelet features are selected for the representation of Kannada numerals has turned out to be novel features with which it is sufficient to use a small set of training patterns. Here features are insensitive to the shape variations caused by the writing styles of different persons. The recognition reliability is increased with wavelet recognition methodology. It was observed that, the method reduced the rate of substitute errors.

REFERENCES

[1] Basappa B. Kodada and Shivakumar K. M-". Unconstrained Kannada Numeral Recognition", *International Journal of Information and Electronics Engineering*, Vol. 3, No. 2, March 2013.

[2] Santosh K.C., CholwichNattee, A Comprehensive survey on On-line handwriting recognition technology and its real application to the Nepalese natural handwriting, Kathmandu university, In*proceedings of Journal of Science, Engineering and Technology*, Vol. 5, No. I, January 2009, pp 31-55.

[3] Pritpal Singh, SumitBudhiraja-" Offline Gurmukhi Numeral Recognition using Wavelet Transforms",*I.J.Modern Education and Computer Science*, 2012, 8, 34-39 Published Online August 2012 in MECS.

[4] Satish Kumar, "Kannada Hand-printed Character Recognition using Multiple Features and Multi-stage Classifier", In *proceedings of International Journal of Computer Information Systems and Industrial Management Applications*, Vol.2, 2010, pp.039-055.

[5] SrinivasRaoKunte and Sudhakar Samuel R.D "Wavelet Features Based On-line Recogntion of Handwritten Kannada Characters", In *proceedings of Journal of the Visualization Society of Japan*, Vol. 20, No. 1, pp 417-420, 2000.

[6] L. M. Lorigo and V. Govindaraju, "Offline Arabic Handwriting Recognition: A Survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2006, vol. 22, no. 5, pp. 712-724, 2006

[7] S.V Rajashekarardhya and P. VanajaRanjan – "Handwritten Numeral RecogonitionKannada Script", Proceedings of the International Workshop on Machine Intelligence Reaserch, 2009 MIR Labs

[8] Ashwin T. V and P. S. Sastry, "Font and size independent OCR for printed Kannada documents using SVM classifier", In *proceedings of Open Research Forum, Fifth ICDAR*, Vol. 27, pp. 14-18, 1999.

[9] W. Niblack, An Introduction to Digital Image Processing", *Englewood Cliffs, N.J. Prentice Hall*, pp. 115-116. 1986.

[10] O. D. Trier, A. K. Jain, T. Taxt, "Features Extraction Methods for Character Recognitio- A Survey", *Pattern Recogntion*, Vol. 29, pp. 641-662, 1996.

[11] Mamatha H. R, Srikant Murthy K, Sudan S, Vinay G Raj and Sumukh S Jois- "Fan Beam projection based features to recogonizehandwritten kannada numerals", 2011 International Conference on Software and Computer Applicaions IPCSIT vol.9(2011),IACSIT Press,Singapore.

[12] G. G. Rajput, RajeswariHorakeri, SidramappaChandrakanth, Printed and Handwritten Kannada Numeral Recognition Using Crack Codes and Fourier Descriptors Plate", In *Proceedings of IJCA Special Issues on Recent Trends in Image Processing and Patten Recognition*, pp. 53-57, 2010.

[13] M. Abdul Rahiman, M.S. Rajasree, "A Wavelet Based Recognition System for Printed Malayalam Characters", In *Proceedings of International Journal of Recent Trends in Engineering*, Vol. 2, pp. 17-22, 2009

[14] Yang Song, Jian Huang, Ding Zhou, HongyuanZha, and C. Lee Giles, "IKNN: Informative K-Nearest Neighbor Pattern Classification", *Springer Verlag Berlin Heidelberg*, pp.248-264, 2007.

[15] Kemal Koche, " Comparison of Neural Network and Template Matching Technique for Identification of Characters in License Plate", In *Proceedings of the Int. Conf. on Information science and Applications*, pp. 186-190, 2010.