ISSN 2348-1196 (print)
International Journal of Computer Science and Information Technology Research ISSN 2348-120X (online)
Vol. 3, Issue 2, pp: (1107-1114), Month: April - June 2015, Available at: www.researchpublish.com

Offline Handwritten Kannada Character Segmentation and Recognition based on Zoning

[1]Ms. Roopa Tonashyal, [2]Mr. Y. C. Kiran

[1]M. Tech Student, Department of ISE, Dayananda Sagar College of Engineering, Bangalore, India
[2]Research Scholar, Jain University Bangalore, India
[2]Associate Professor, Department of ISE, Dayananda Sagar College of Engineering, Bangalore, India

Abstract: **Handwritten Kannada Character Recognition has been a challenging research domain due to its diverse applicable environment. Handwriting has always been and will possibly continue to be a means of communication. There is a need to convert these handwritten documents into an editable format which can be achieved by Handwritten Character Recognition Systems. This considerably reduces the storage space required. In this paper focus is on offline handwritten kannada characters. Feature extraction is performed using zoning method together with the concept of euler number. This increases accuracy and speed of recognition as the search space can be reduced.**

Keywords: **Aspect Ratio, Euler Number, End Points, Offline Handwritten Character Recognition, Zoning.**

I. INTRODUCTION

Handwritten character recognition has been one of the most challenging and fascinating areas in the field of image processing and pattern recognition. Character recognition is generally defined as machine simulation of human reading. It is also known as Optical Character Recognition. It contributes tremendously to the progress of an automation process and can enhance the interaction between man and machine in a number of applications. There are several research works that have been put forward with its complete focus on new methods and techniques with an aim to cut down the processing time to as less as possible while rendering higher recognition accuracy.

A lot of work has been done on the recognition of printed characters of Indian languages. On the other hand, attempts made on the recognition of handwritten characters are few. Most of the research in this area is mainly focusing on recognition of off-line handwritten characters for the scripts like kannada, Devanagari and Bangla . From the literature survey it can be seen that there is a lot of demand for character recognition systems for Indian scripts and an excellent review has been done on the Optical Character Recognition (OCR) for Indian languages.

Handwriting recognition is a process that needs training. Training can be carried out using a set of samples of handwriting taken from a group of writers(writer-independent training) or using samples of handwriting from an individual writer(writer-dependent training). From various experiments carried out by researchers it can be concluded that the accuracy of recognizing handwriting is higher in case of writer-dependant training because the text samples from an individual writer are used to train the recognizer.

Kannada is the official language of the South Indian state of Karnataka. It has its own script derived from Bramhi script. Kannada script has a base set of 52 characters, comprising 16 vowels and 36 consonants. Further there are distinct symbols that modify the base consonants, called consonant and vowel modifiers. The number of these modifiers is the same as that of the base characters. The characters called aksharas are formed by graphically combining the symbols corresponding to consonants, consonant modifiers (optional) and vowel modifiers using well defined rules of combination. Therefore, the number of theoretically possible combinations of Kannada characters is 16 vowels, 36*16=576 consonant-vowel combinations and 36*36*16=20736 consonant-consonant-vowel combinations.

General Character Recognition System is shown below

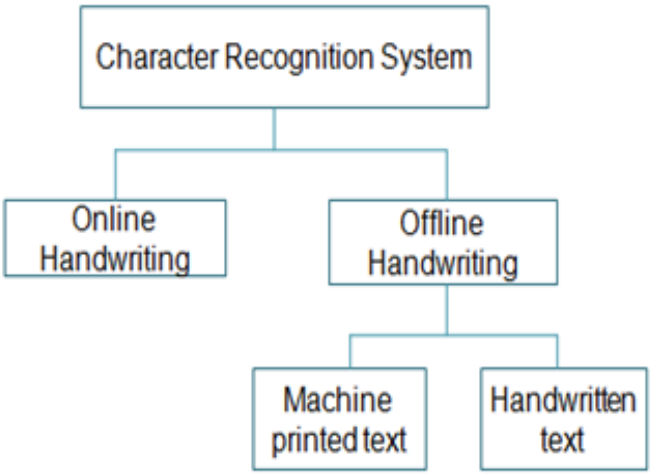


Fig 1.1General Character Recognition System

Major steps of Handwritten Kannada Character Recognition are shown in below fig

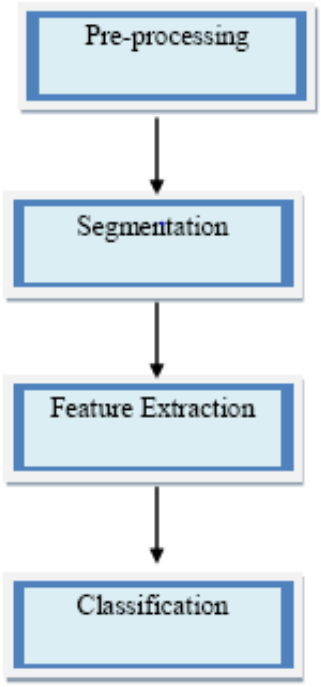


Fig 1.2 Stages of Character Segmentation and Recognition

II. PREPARATION OF DATASET

First we take the blank A4 size paper. On that blank paper we make 13 rows and 14 columns. And wrote the words on those blocks. These words are written from 10 different people. So totally we collected around 1000 words. These papers are then scanned by hp scanner setting DPI by 600. After scanning, each word is cropped and performed pre-processing.

ISSN 2348-1196 (print)
International Journal of Computer Science and Information Technology Research ISSN 2348-120X (online)
Vol. 3, Issue 2, pp: (1107-1114), Month: April - June 2015, Available at: www.researchpublish.com

Example for cropped image is shown in below figure

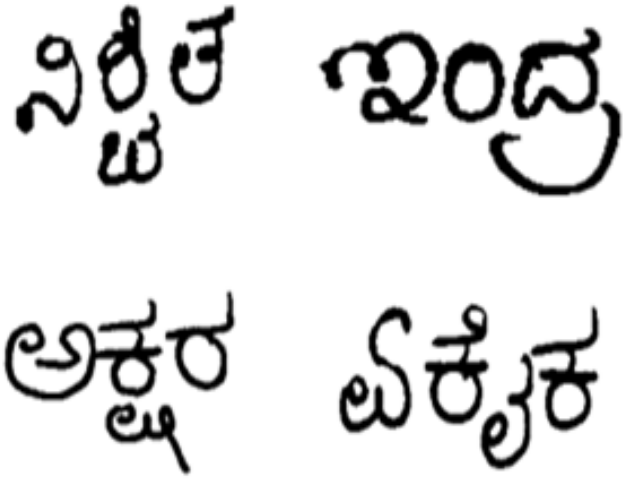


Fig 2.1 Cropped images from A4 Sheet

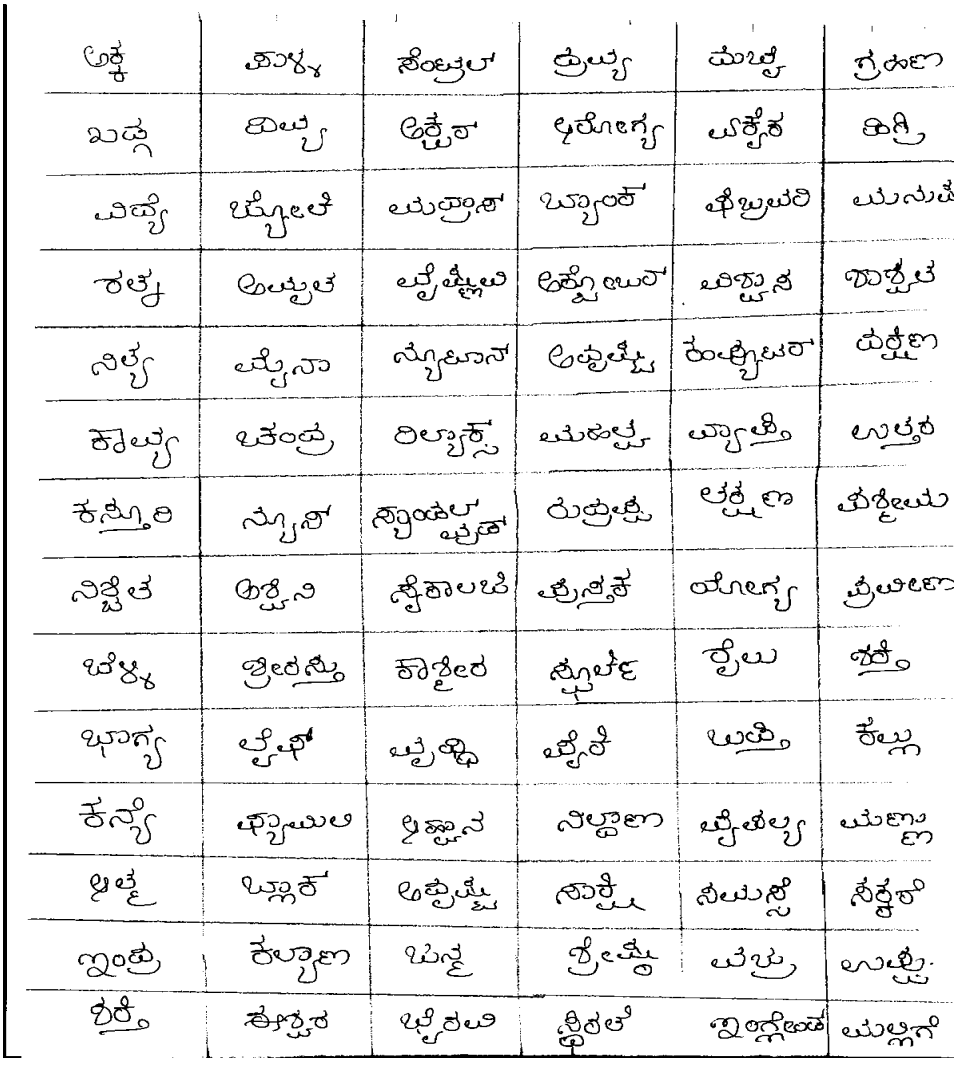


Fig 2.2 Cropped Words

ISSN 2348-1196 (print)
International Journal of Computer Science and Information Technology Research ISSN 2348-120X (online)
Vol. 3, Issue 2, pp: (1107-1114), Month: April - June 2015, Available at: www.researchpublish.com

III. SYSTEM DESIGN

Training Phase

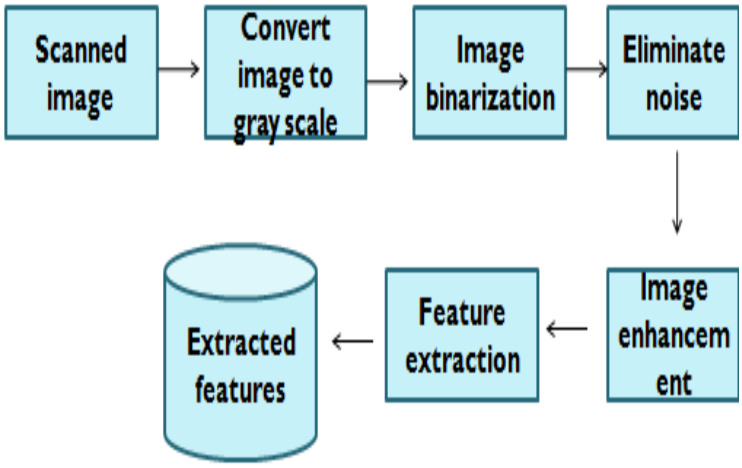


Fig 3.1 Training Phase

Testing Phase

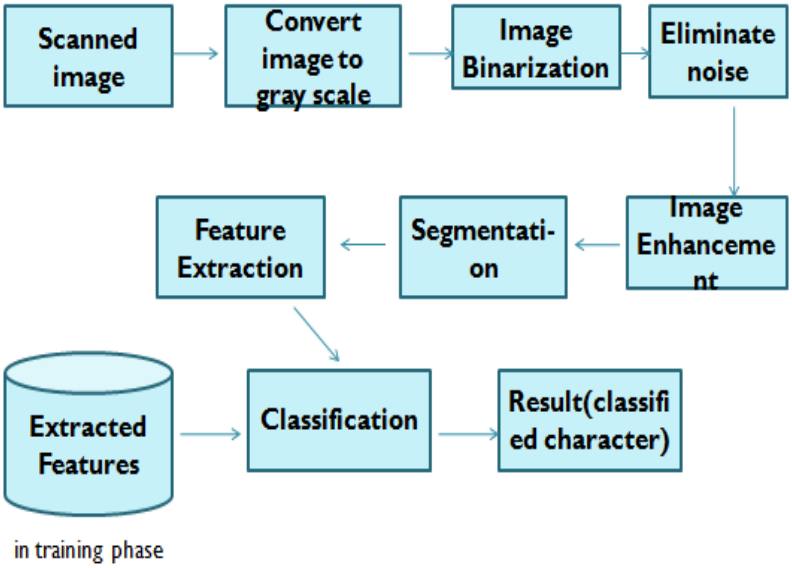


Fig 3.2 Testing Phase

IV. PRE-PROCESSING

The images are subject to certain pre-processing operations in order to get rid of unnecessary artefacts. Pre-processing of the image is carried out in the following way:-

Binarization:

This converts the gray-scale image into black and white image where in the pixel values of the image are either 0 or 1.

Cropping and Resizing:

The top-leftmost pixel, top-rightmost pixel, bottom-leftmost pixel and bottom-rightmost pixel of the images are identified and stored. These values are fed to the cropping function in order to extract only the character from the image. After cropping the character image, the image is resized to a standard size.

ISSN 2348-1196 (print)
International Journal of Computer Science and Information Technology Research ISSN 2348-120X (online)
Vol. 3, Issue 2, pp: (1107-1114), Month: April - June 2015, Available at: www.researchpublish.com

Noise removal:

In order to remove noise from the image median filter is used. To enhance the image quality dilation is used which takes one of its arguments as the structuring element.

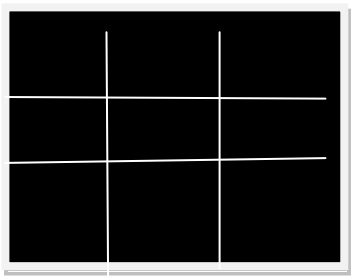
V. SEGMENTATION

1. Scan the BMP image vertically for the word, to find first ON pixel and remember that X coordinate as X1. Treat this as starting coordinate for the character.
2. Continue scanning the BMP image then we would find lots of ON pixel since the character would have started.
3. Finally we get the OFF pixel column and remember that X coordinate as X2
4. X1 to X2 is the character.
5. Repeat the steps above until we find all characters in the word.

VI. FEATURE EXTRACTION

Creation of zones

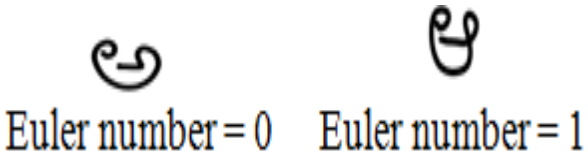
To create zones the pixels at 50^{th} row, 50^{th} column, 100^{th} row, 100^{th} column are set to 1 which will form 9 zones as follows



Euler number

A number obtained by subtracting the number of holes in the image from number of objects in the image

For example



End points

And thus the concept of finding the end points of characters was added to eliminate the problem arising from using only the euler number. End points of each character were noted down as to which zones they lie into.

VII. CLASSIFICATION

The SVM classifier is a two class classifier. There is a discriminant functions based on which it works. This discriminant function is hyper plane which represents a surface. This hyper plane separates the patterns as two classes. For OCR applications a number of two class classifiers are trained with each one distinguishing one class from the other. Each class label has an associated SVM and a test example is assigned to the label of the class whose SVM gives the largest positive output. The input image i.e. testing image is rejected if SVM does not gives positive output.

ISSN 2348-1196 (print)
International Journal of Computer Science and Information Technology Research ISSN 2348-120X (online)
Vol. 3, Issue 2, pp: (1107-1114), Month: April - June 2015, Available at: www.researchpublish.com

VIII. RESULTS

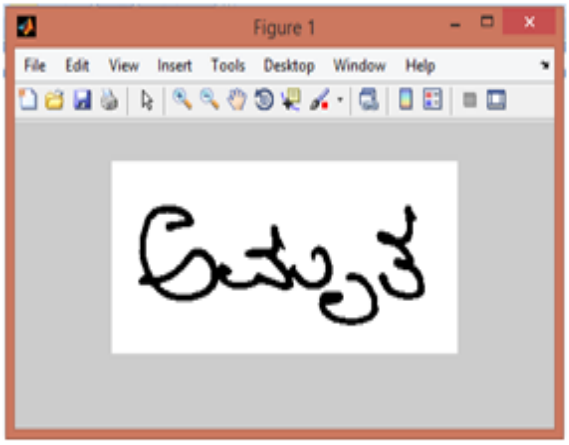


Fig 8.1 Input Image

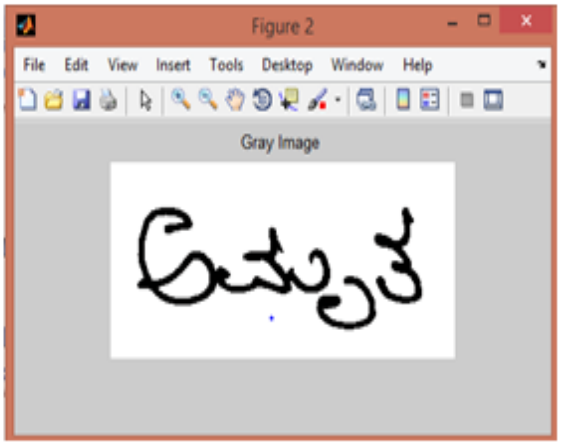


Fig 8.2 Gray Scale Image

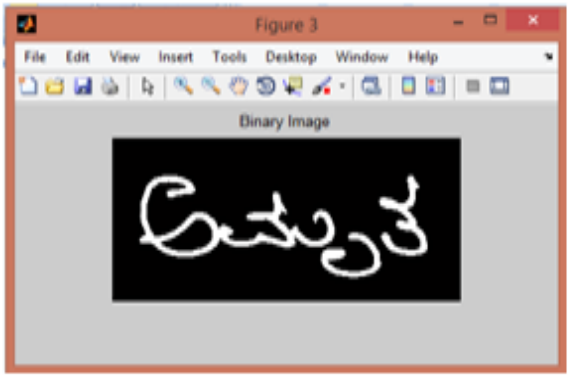


Fig 8.3 Complemented Image

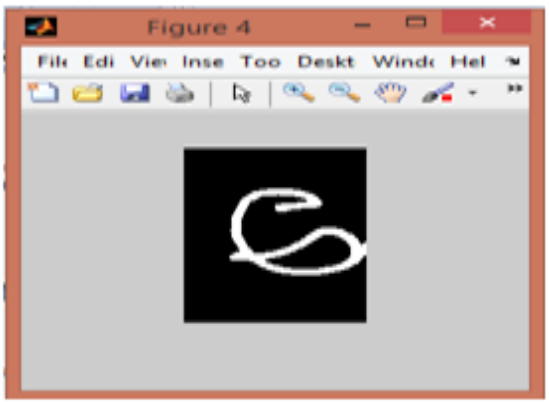


Fig 8.4 Segmented First Character

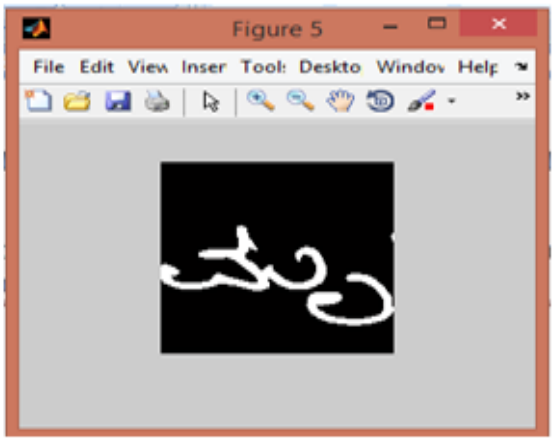


Fig 8.5 Segmented Second Character

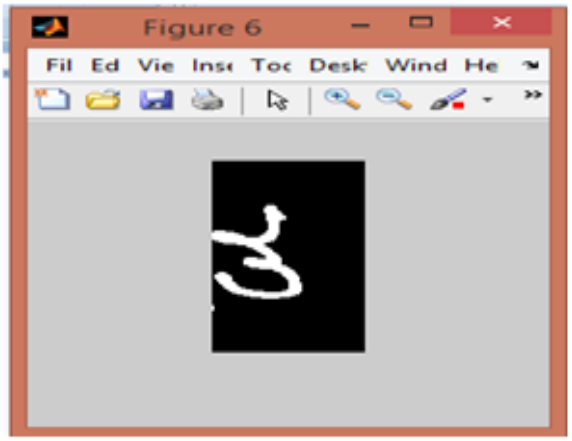


Fig 8.6 Segmented Third Character

ISSN 2348-1196 (print)
International Journal of Computer Science and Information Technology Research ISSN 2348-120X (online)
Vol. 3, Issue 2, pp: (1107-1114), Month: April - June 2015, Available at: www.researchpublish.com

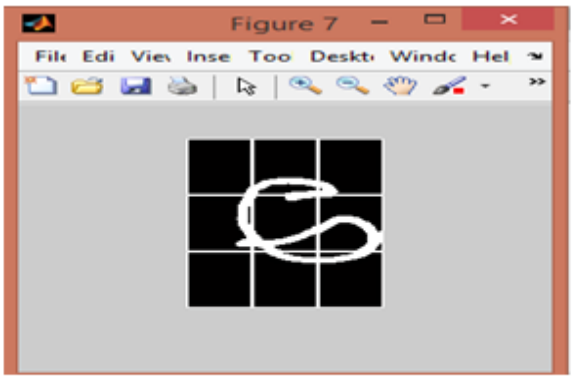


Fig 8.7 Feature Extraction for First Character

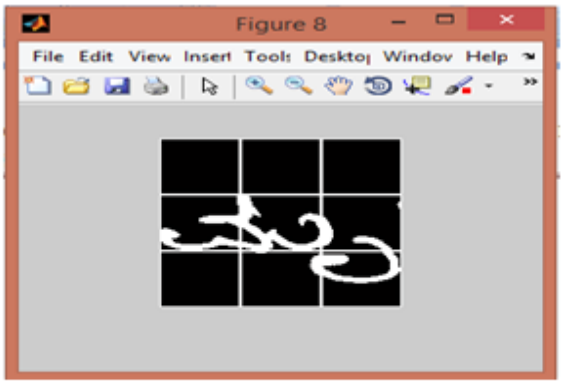


Fig 8.8 Feature Extraction for Second Character

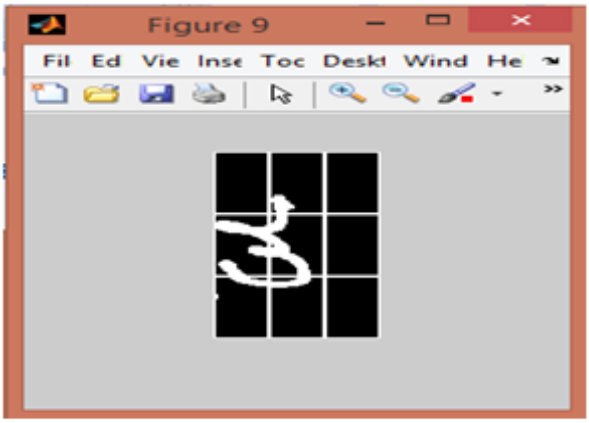


Fig 8.9 Feature Extraction for Third Character

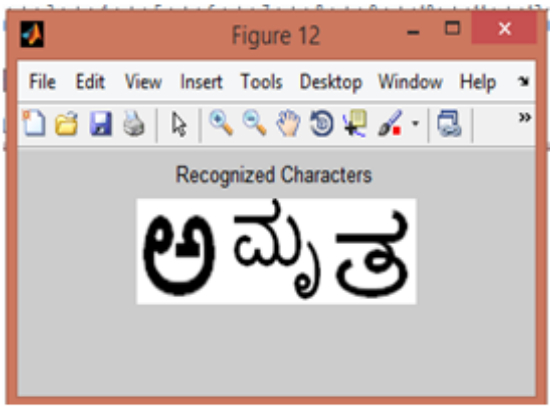


Fig 8.10 Recognized Characters

IX. CONCLUSION

Thus we have increase the accuracy of Handwritten Kannada Character recognition using zoning method for feature extraction and Support Vector Machine classifier for classification. It increases the speed of the accuracy and it requires less memory for storing the training samples.

REFERENCES

[1] Swapnil A. Vaidya, Balaji R. Bombade, "A Comprehensive Survey on Kannada Numerals and Character Recognition", ISSN: 2277 128X,March 2013.

[2] Rajashekararadhya S. V., Vanaja Ranjan P., Manjunath Aradhya V. N., "Isolated Handwritten Kannada and Tamil Numeral Recognition: A Novel Approach", First International Conference on Emerging Trends in Engineering and Technology- ICETET, pp.1192-1195,16-18 July 2008.

[3] Mamatha H. R., Karthik S., Srikanta Murthy K., "Feature Based Recognition of Handwritten Kannada Numerals – A Comparative Study", International Conference onComputing, Communication and Applications (ICCCA),22-24 Feb, 2012.

[4] S.V. Rajashekararadhya, p. Vanaja Ranjan, "Handwritten Numaral Recognition of kannada script",2009

[5] J.Pradeep, E.Srinivasan and S.Himavathi, "Diagonal based feature extraction for handwritten alphabets recognition system using neural network", International Journal of Computer Science & Information Technology (IJCSIT), Vol 3, No 1, Feb 2011.

[6] H. Imran Khan, Smitha U. V, Suresh Kumar D. S, "Isolated Kannada Character Recognition using Chain Code Features", International Journal of Science and Research (IJSR), India Online ISSN: 2319-7064.

[7] Mamatha H.R, Karthik S, Srikanta Murthy K, "Classifier Fusion Method to Recognize Handwritten Kannada Numerals".

[8] Thungamani.M, Dr Ramakhanth Kumar P, Keshava Prasanna, Shravani Krishna Rau, "Off-line Handwritten Kannada Text Recognition using Support Vector Machine using Zernike Moments", IJCSNS International Journal of Computer Science and Network Security, VOL.11 No.7, July 2011.

[9] Mamatha.H.R, Sucharitha Srirangaprasad, Srikantamurthy K, "Data fusion based framework for the recognition of Isolated Handwritten Kannada Numerals", (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 4, No. 6, 2013.

[10] Jomy John, Pramod K. V, Kannan Balakrishnan, "Handwritten Character Recognition of South Indian Scripts: A Review", National Conference on Indian Language Computing, Kochi, Feb 19-20, 2011.

[11] M. Miciak, "Character recognition using Radon Transformation and Principal Component Analysis in postal applications", Proc. of International Multi conference on Computer Science and Information Technology, (2008) October 20-22, pp. 495-500.

[12] H. R. Mamatha, K. Srikanta Murthy, P. Vishwanath, T. S. Savitha, A. S. Sahana and S. Suma Shankari, "Evaluation of Similarity Measures for Recognition of Handwritten Kannada Numerals", CiiT International Journal of Digital Image Processing, ISSN 0974-9691 and Online: ISSN 0974-9586, DOI: DIP102011018, vol. 3,no. 16, (2011) October, pp. 1025-1029.

[13] A. Fitzgibbon and A. Zisserman, "On Affine Invariant Clustering and Automatic Cast Listing in Movies", Proceedings of 7th European Conference on Computer Vision, ECCV, vol. 3, (2002), pp. 304-320.

[14] G. G. Rajput, R. Horakeri and S. Chandrakant, "Printed and Handwritten Kannada Numeral Recognition Using Crack Codes and Fourier Descriptors Plate", IJCA Special Issue "Recent Trends in Image Processing and Pattern Recognition", RTIPPR, (2010), pp. 53-58.