MAT JOURNALS

Journal of VLSI Design and Signal Processing
Volume 4 Issue 2

Recognition of Printed and Handwritten Kannada Characters using SVM Classifier

Urmila B. Basarkod [1], Shivanand Patil [2]
[1]P.G. Student, [2]Assistant Professor
Department of Computer Science and Engineering, KLE Dr. M. S. Sheshgiri College of Engineering and Technology, Udyambag, Belagavi, Karnataka, India
E-mail: urmilabasarkod@gmail.com

Abstract

The optical character recognition is the process of converting textual scanned image into a computer editable format but one of the major challenges faced is the recognition of character from the image. The proposed system is application software for Recognition of Kannada Printed and Handwritten Characters from an image. The input image is subjected for pre-processing to make the image noise free by using median filter and then it is converted to binary image. Segmentation process is carried out to extract one character from the image by performing horizontal segmentation followed by vertical segmentation. Co-relation coefficient is used for extracting the features from the image then the character is classified using SVM classifier finally the classified character is post-processed using its Unicode values to display the recognized character. We have obtained perfectness of 100% and 99% in recognition of Kannada Printed and Handwritten characters respectively.

Keywords: *Kannada Script; Optical Character Recognition; Median Filter; Local Binary Pattern, Co-relation Coefficient, SVM Classifier;*

INTRODUCTION

In the present scenario large amount of information is stored in images, thus character recognition has obtained significant popularity and it has emerged as an active area of research, Thus Image processing is mainly concerned about extraction of more valuable information from the images. The recognition of information begins by identifying the text from the image, this process of identifying and recognition text is called Optical Character Recognition (OCR).

Optical Character Recognition is mechanical conversion or electronic conversion of printed or hand written text images into a form that the computer can manipulate. Optical Character Recognition includes five stage of processing those include pre-processing, segmentation, feature extraction, classification, and post processing. Each stage result is depending on the result of previous stage.

Kannada is Dravidian language predominantly spoken by kannada people. It is the official language of Karnataka. The Kannada script is a combination of 49 letters and these letters are of 3 categories they are: swara includes 13 letters, vyanjanas include 34 letters and yogavaka includes 2 letters.

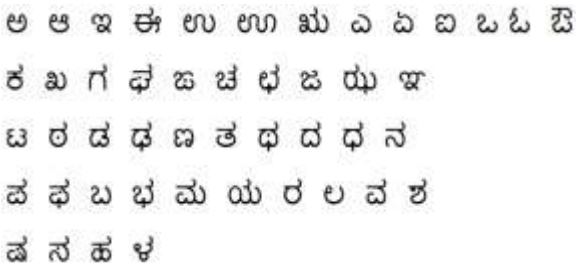

Fig. 1: *Kannada letters*

Some languages of Asia like Japanese and Chinese and many European languages

have used an OCR system. However less effort as put in the OCR system for Indian languages, particularly for South Indian languages such as Kannada. Thus this work concentrates on Recognition of Kannada Characters from an image.

LITERATURE SURVEY

In[1] author provides general method for Recognition of Character from the image. This process of recognising text from image is step by step process which includes pre-processing for basic processing input image like binarization for converting gray scale image into Binary image and also for removing noise from the image, segmentation is the processing step that divides image into line then into character, feature extraction is to calculate the characteristics of an image, classification is to compare characters from database, and at last it is post processing. This paper provides a brief survey on the extraction of characters from given images. In [3] usually images include some noise, because of presence of these noise quality of an image degrades. Therefore making an image noise free is too important in order to maximise the quality of an image. The process of removal of noise or improving the quality of image is called de-noising. In this an author has discussed about some list of noise that can be present in the image and the some de-noising techniques those can be used to minimise the noise in the image. In [5] Recognition of text from an image include 5 steps among these classification is one of the necessary step need to be carried out for recognising the text from the image. The efficiency and accuracy of recognising text highly depends on classification phase. In this paper an author provided briefly about classifiers, to the beginners of this field. In this they have explained about Support Vector Machine (SVM), Artificial Neural Network (ANN), Decision Tree (DT), and KNN classifier, which are highly popular classifiers for recognising text in the field of image processing. In [6] author used OCR system for handwritten Kannada and Telugu digits recognition they have used zoning features along with the K-nearest Neighbour classifier and Support Vector Machine classifier independently to identify handwritten Kannada and Telugu digits. They have considered an image consisting a digit which is then divided it into 64 zones. Computed each zone pixel density and adopted classifiers independently to classify the digits. Acquired exactness for handwritten Kannada and Telugu digits recognition of 95.50%, 96.22% and 99.83% and 99.80% respectively. In [7] author has considered an image containing Kannada vowels or English characters, and normalized it into 32*32, then divided into 64 zones and the pixel density is counted for each zone. These features are assigned to KNN and SVM classifiers. In case of handwritten Kannada character recognition, gained an average perfectness of 92.71% and 96.0% with KNN and SVM respectively. For uppercase English alphabet gained an average perfectness of 97.51% and 98.26% with KNN and SVM respectively. In[8] Recognised handwritten Kannada and English characters from an image based on spatial features. Characters are assigned to KNN classifier. The algorithm is dependent on the type and size of the structuring element used for feature extraction. It is independent of the image normalisation, thinning, noise and slant of the characters. Gained a perfectness of an average percentage 96.2% and 90.01% % in case of handwritten numerical and English uppercase alphabets respectively.

PROPOSED SYSTEM

Aim of the proposed system is to recognise the Kannada Printed and Handwritten Characters from the considered image using Optical Character Recognition. The process includes five steps

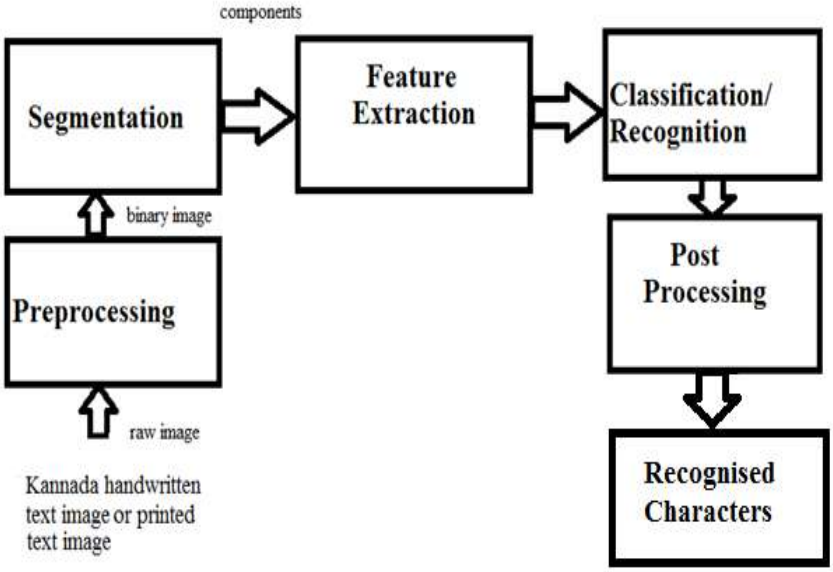

Fig. 2: *System Architecture*

Step 1: Pre-processing of the image:
An image is considered and it is converted from RGB - color image into gray scale image. Then median filter-2 is used to filter the image then it is converted to binary image of 0's and 1's form using thresholding. Then BWareaopen is used to remove noise from the binirized image and after completion, the image is passed for segmenting the image.

Step 2: Segmentation process
The pre-processed image is considered and it is segmented horizontally then followed by vertically to extract the single character from the image. It considers the image performs horizontal segmentation and divides into lines, Then it considers the first line, Performs vertical segmentation on the line and divide it into single characters which are independent of other characters in the line. Repeat this process for all the characters in the line, and for all the lines in the image.

Step 3: Feature Extraction
Feature extraction is one of the major steps, which describes the characteristics of an image. Even the accuracy of the classification / recognition is depending on the result of feature extraction. For extraction of potential features from considered image, the image is segmented by LBP and normalized to size of 42*24 and after normalizing correlation coefficient is used for extraction of text from image. The result of correlation coefficient lies between 0 to 1. If the result is 0, there are no similarity exits. If the result is 1 both images are exactly the same

Step 4: Classification
For classification of characters SVM (Support Vector Machine) classifier is used. . It provides separation between 2 linearly separable classes which is achieved by considering a hyper plane on all data points of considered classes. The hyper plane provides proper classification on all data points. Consider two classes A and B. All the points belongs to class A are labeled +1 and the points belongs to class B are labeled as -1. As shown in below figure.

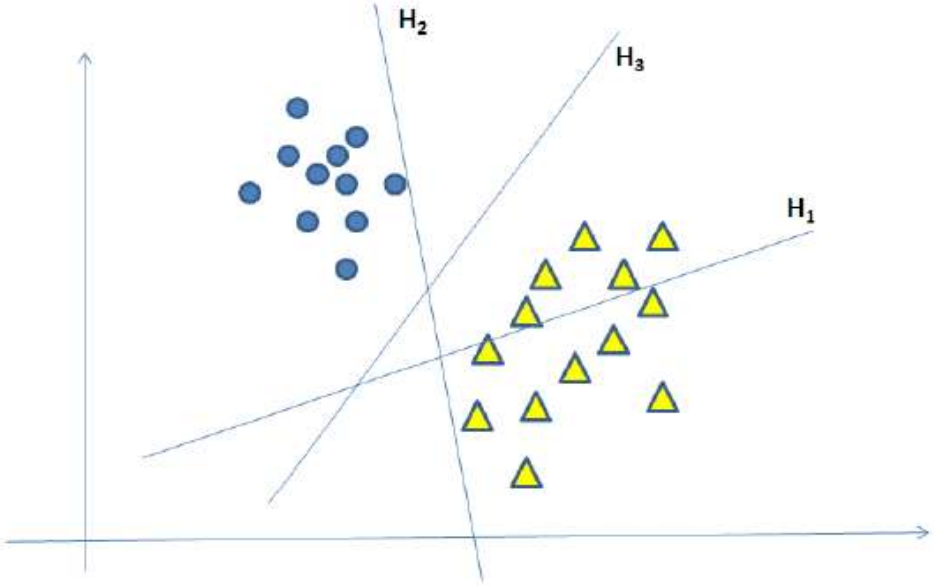

Fig 3: *SVM Classification*

Step 5:Post-Processing

Post - Processing is last step of this project which considers the image belongs to class-A which is classified by SVM classifier. Assign them unique ASCII value. In this the testing character is compared with the ASCII codes or Unicode's of the Kannada characters and matched character is displayed onto the screen.

Experimental Result

The experiment was carried out for Kannada Printed and Handwritten Characters by using SVM as a classifier. The result of the experiment is gained by using 118 Training examples, and tested by 118 examples. Below pictures show the result snapshots of the proposed system same procedure is followed for handwritten characters. Table 1 shows the comparison of different methods those were implemented by different authors with the proposed system.

Checking for printed Characters

Fig 4: *Selecting an image*

Fig 5: *Performing gray scale*

Fig 6: *performing Binarization*

Fig 7: *Removing noise from the image*

Fig 8: *Loading the training data set*

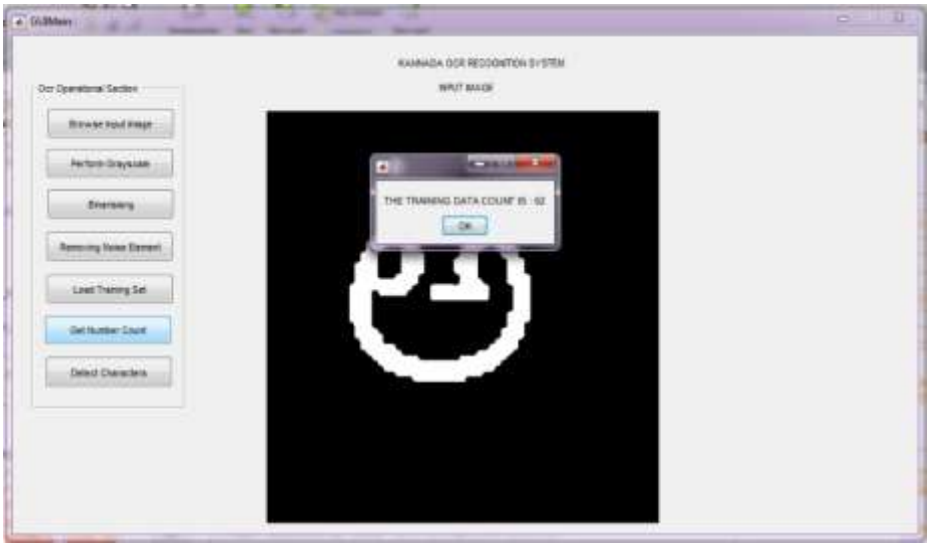

Fig 9: *Counting number of training data in the set*

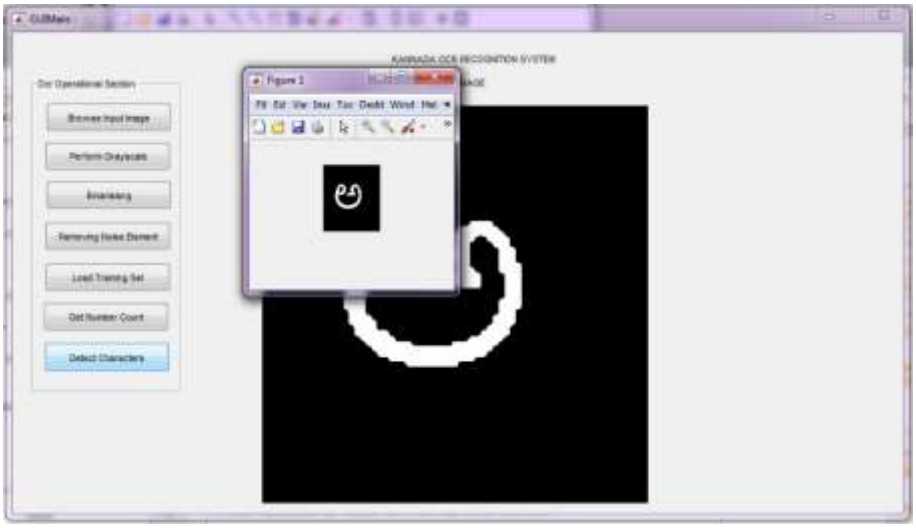

Fig 10: *Identifying the Characters from the image*

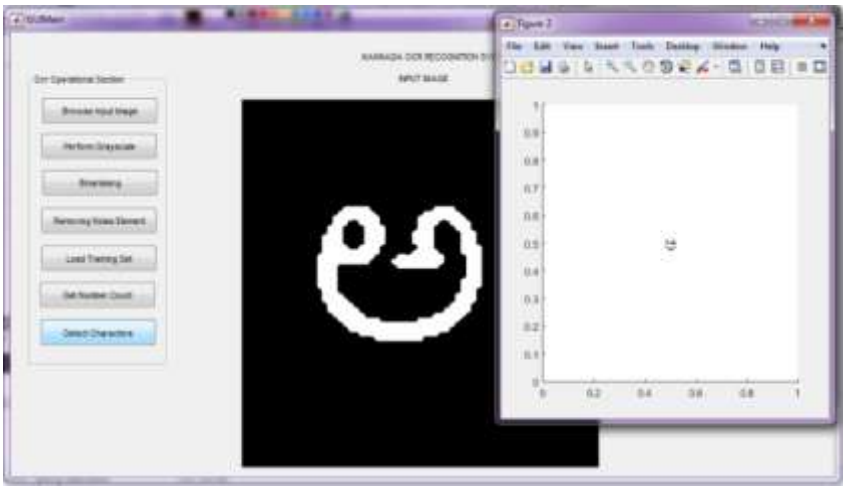

Fig 11: *Recognized Character*

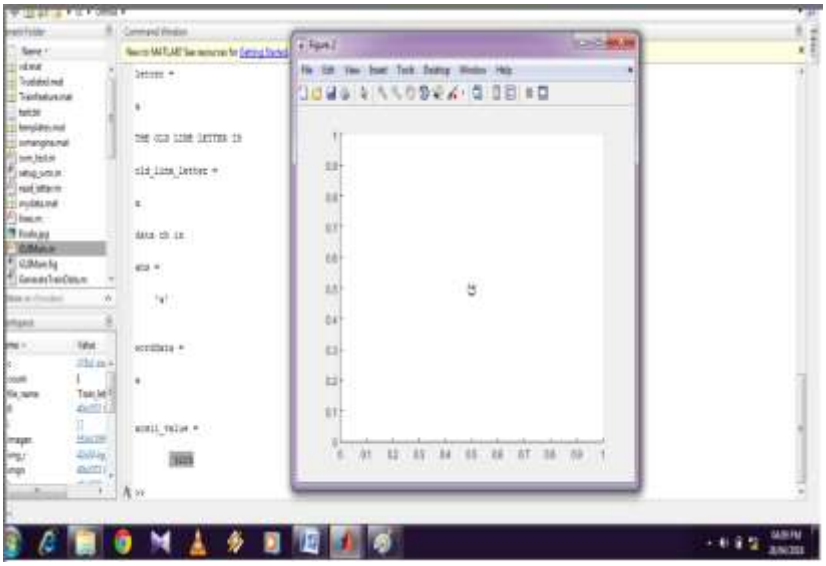

Fig 12: *Comparing Recognized Character with its Unicode*

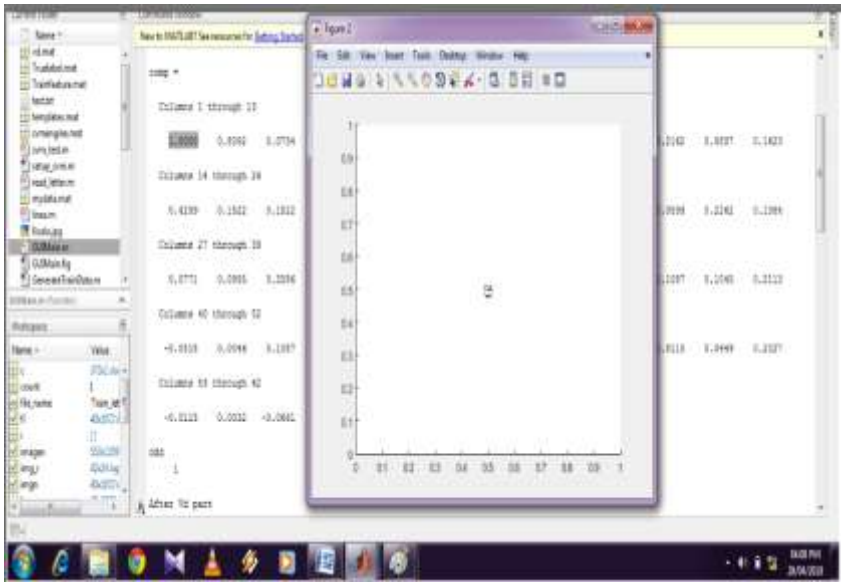

Fig 13: *Comparing Recognized Character with Character in trained set*

Table 1: *Comparison of Results*

Methods Proposed by	**Features and Classifier used**	**Percentage of Accuracy**
Tridib Chakraborty Chowdhury Md Mizan, and Suparna Karmaka		Not mentioned
gururaj mukarambi, mallikarjun hangarge, B. V. Dhandra	Zoning features K-Nearest Neighbor and SVM	95.50%
Gururaj Mukarambi, B.V.Dhandra , Mallikarjun Hangarge	density based zone features wih KNN and SVM	KNN 95.50% and 97.32% SVM 96.2% and 98.30%
B.V.Dhandra , Mallikarjun Hangarge, Gururaj Mukarambi	Zone based fetures with KNN and SVM	KNN and SVM Handwritten 95.50% , 96.22% Printed 100.0%, 100.0% Mixed(both) 97.32%, 98.30%
Shashikala Parameshwarappa, B.V.Dhandra	Spatial features considered for KNN	90.01%
Netravati Belagali, Shanmukhappa A. Angadi	Zoning based invariant moment feature with PNN	94.69%
Srikanta Murthy K Karthik S	HOG with SVM	95.02%
Proposed method	Co-relation coefficient with SVM classifier	Printed 100% Handwritten 99%

CONCLUSION

In current scenario most of the data stored in the form of an image, thus recognition of character plays a major role in extracting the data from an image. There exists a maximum work in regard of recognising Kannada Characters but there exists no standard solution to recognise Kannada characters from the image with reasonable perfectness. In this research work, presented a application software for Recognition of Kannada Printed and Handwritten Characters from an image. Firstly, extracted the image with character, then image is pre-processed to remove noise from the image and then Correlation Coefficient is used for extraction of features from image for Recognition of character. The proposed work has provided good perfectness on recognising characters from image by using SVM classifier. Gained a perfectness of 100% on Kannada Printed Characters and in case of Kannada Handwritten Characters gained perfectness of 99.0%.

REFERENCE

1. Chowdhury Md Mizan, Tridib Chakraborty and Suparna Karmakar, " Text Recognition using Image Processing", International Journal of Advanced Research in Computer Science, Volume 8, No. 5,Volume 8, No. 5, May – June 2017
2. Sukhjinder Kaur "Noise Types and Various Removal Techniques". IJARECE, Volume 4, Issue
3. Manshi Shukla and Barjinder Kaur," Image De-noising and its Methods: A Survey"
4. International Journal of Science and Research (IJSR), Volume 6 Issue 3,Abdalla Mohamed Hambal, Dr. Zhijun Pei, and Faustini Libent Ishabailu, "Image Noise Reduction and Filtering Techniques ".
5. Rama Gaur, Dr. V.S. Chouhan, "Classifiers in Image processing", IJFRCSCE Volume: 3 Issue: 6.
6. Mallikarjun Hangarge, Gururaj Mukarambi, and B. V. Dhandra 2011,

" A script independent approach for handwritten bilingual kannada And telugu digits recognition " International Journal of Machine Intelligence, ISSN: 0975–2927 & E-ISSN: 0975–9166, Volume 3, Issue 3, 2011, pp-155-159.
7. Mallikarjun Hangarge, Gururaj Mukarambi, and B. V. Dhandra ," Handwritten Kannada Vowels and English Character Recognition System ", IJIPVS, Volume-1 Issue-1 ,2012.
8. Mallikarjun Hangarge, Gururaj Mukarambi, and B. V. Dhandra," Spatial Features for Handwritten Kannada and English Character Recognition", RTIPPR, 2010.
9. Vishweshwarayya C. Hallur, Avinash A. Malawade, Seema G. Itagi, "Survey on Kannada Digits Recognition Using OCR Technique", IJARCET, Volume 1, Issue 10, December 2012.
10. B.V.Dhandra, Gururaj Mukarambi, Mallikarjun Hangarge, "Zone Based Features for Handwritten and Printed Mixed Kannada Digits Recognition" ICVCI 2011, Proceedings published by IJCA.
11. http://en.wikipedia.com
12. Shashikala Parameshwarappa, and B.V.Dhandra, "Basic Kannada Handwritten Character Recognition System using Shape Based and Transform Domain Features", International Journal of Advanced Research in Computer and Communication Engineering Vol. 4, Issue 7, July 2015.
13. Faouzi Bouchareb, Rachid Hamdi, and Mouldi Bedda, " Handwritten Arabic character recognition based on SVM Classifier"
14. Tamim Ahmed Khan, Syed Hassan Tanvir, Abu Bakar Yamin," Evaluation of Optical Character Recognition Algorithms and Feature Extraction Techniques" 2016 IEEE.
15. http://mathworks.com

BIOGRAPHIES

Urmila B. Basarkod is a M.Tech student in the Department of Computer Science & Engg., KLE Dr. M.S. Sheshgiri College of Engg. & Tech., Belagavi. She received B.E. degree in Computer Science& Engineering from S. G. B. I. T College of Engg. & Tech., Belagavi in 2016. Her research interests include Image Processing and Big Data.

Prof. Shivanand M. Patil is a faculty in the Department of Computer Science & Engg., KLE Dr. M.S. Sheshgiri College of Engg. &Tech., Belagavi. He did his Bachelors in Computer Science & Engg. from MMEC, Belagavi and M.Tech in Computer Science& Engg. from Basaveshwar Engineering College, Bagalkot. His research interests include Image Processing and Pattern Recognition. He has number of publications in to his credit in peer reviewed international journals and conferences.