JES Journal of Engineering Sciences

Vol11,Issue2,FEB/2020

ISSN NO:0377-9254

GRAPH BASED PRINTED KANNADA CHARACTER RECOGNITION SYSTEM BY USING BACKPROPAGATION NEURAL NETWORK

RAVIKUMAR B. CHAWHAN[1], MARPUDI RAGHAVENDRA RAO[2]

1 .CSE Dept., Samskruti College of Engineering & Technology, Ghatkesar, Telangana.

ravismsk1@gmail.com

2. CSE Dept., Samskruti College of Engineering & Technology, Ghatkesar, Telangana.

raghavendramarpudi@gmail.com

Abstract-- An optical recognition system (OCR) scans the input printed Kannada character and identifies character resides in database model to outline a separate transcript, it's been pre-processed. There many existing OCR systems are available for handling the printed English report with sensible levels of accurateness. Those systems are existed for European languages and some of the Asian languages such as Japanese, Chinese, etc. As many OCR systems exists also for Kannada languages. Printed Kannada character can be recognized using neural network. This methodology uses three-layer model such as, input layer, hidden layer and output layer. The combination of graph based features and neural classifier gives the expected result.

Keywords-- character recognition, pre-processing, BPNN, training, testing.

I. INTRODUCTION

Now a day's in Karnataka more importance is given to utilize kannada in all the fields, hence the kannada language processing and recognition using computer machine has become vital. Presently so many OCR systems are available for handling the english documents with a satisfactory accuracy. To build kannada OCR systems, it is comprising of different challenges/issues since all kannada characters appear in similar shapes and containing different components in them. The recognition of printed kannada character is challenging field of image processing.

Kannada is the regional language of the South Indian states. It has its own script derived from Bramhi script. Modern Kannada alphabet has a base set of 52 characters, comprising of 16 vowels shown in fig 1.1 (called as swaragalu) and 36 consonants shown in fig 1.2 (vyanjanagalu).

ಅ ಆ ಇ ಈ ಉ ಊ ಋ ೠ ಎ ಏ ಐ ಒ ಓ ಔ ಅಂ ಅಃ

Figure 1.1 vowels in Kannada.

ಕ ಖ ಗ ಘ ಙ
ಚ ಛ ಜ ಝ ಞ
ಟ ಠ ಡ ಢ ಣ
ತ ಥ ದ ಧ ನ
ಪ ಫ ಬ ಭ ಮ
ಯ ರ ಲ ಳ
ವ ಶ ಷ ಸ ಹ

Figure 1.2 Consonants in Kannada

Since large number of researches are going on the development of an efficient and robust Kannada OCR system for different challenges such as scripts containing diverse font styles and sizes. Many methods have been proposed for OCR of Indian scripts like Bangla, Devanagiri, Telugu and Tamil and Kannada. In many of the existing system, Classical moment invariants were introduced by Hu (1962) which is invariant under transformation methods. Neural Networks have been used for character recognition systems. Neural networks have fast training/learning rate because of their local-tuned neurons (Moody & Darken 1989). They have also exhibited universal approximation property and have the better generalization ability (Park & Wsandberg 1991). Also the graph based techniques are used in many fields which lead efficient and satisfactory

results. But no OCR systems have used graph based techniques to recognize printed kannada characters.
Kannada characters comprise of more curves in the Kannada characters, similar in their shape. Some of the characters like na, sa are seems to be same. Printed Kannada may differ in font styles and sizes. By overcoming all these, machine has to recognize the printed Kannada character.
Printed kannada character recognition systems can be used in bank for processing cheques, postal documentation, government boards, advertisements, shop names, addresses, business cards etc. Extracting text from these require accurate recognition of the characters amidst different environmental conditions like luminosity, rotation, reflection, scaling among others.

A. Challenges / Issues

There exists many challenges in kannada character processing and recognition systems. No kannada OCR systems are complete in themselves because of these issues. Those are as follows.

- **Curves:** More curves in the Kannada numerals, most of the Kannada characters have the similar curves, so it is difficult to identify the individual Kannada character.
- **Similar shapes:** Most of the Kannada characters that are similar in shape to one another like Ra, Tha and BA, bha includes little variation among them This may leads to difficulties like identification and reduction in the performance of recognition system.
- **Sizes:** Kannada characters are less variation in sizes, especially for all printed Kannada character fixed in to the standard size
- **Font:** Each Kannada character has the various fonts and styles.

B. Overview of Work with Result

In this work, printed Kannada character recognition problem is approached in three steps. Initially the scanned printed Kannada character image is preprocessed to be suitable for extracting features. The obtained image is used to extract suitable graph based parameters that can separate the individual character. The last step involves creating Back Propogation Neural Network (BPNN) and presenting extracted features to train it. During testing, the trained BPNN examines the features with respect to the knowledge available and recognize the character.

This system reached the expected results using 28 graph based features and BPNN. The system has given the 96.4% accuracy for 50 different characters. 8 samples for each printed Kannada character are used for the training the BPNN and 3 samples completely different from training samples are used for testing.

II. PROPOSED WORK

The problem defined after literature survey is **"Graph Based Printed Kannada Character Recognition System"**. Once Kannada text scanned by a system, it produces just an image file. The system cannot understand the letters from image file, hence it cannot search, edit or change the fonts, as in a word processor. It would use OCR software to convert it into a text or word processor file so that it could do those things. The result is much more flexible and compact than the original page image of Kannada text. The need for OCR arises in the context of digitizing the documents from the library, which helps in sharing the data through the Internet
The goal is to outline and develop a graph based technique for printed Kannada character recognition system, which takes the input as printed Kannada character image and recognize it. The proposed model for this work is indicated in fig 3.1. It comprises of two cases, training phase and testing phase It includes a few stages which are examined in further sections.

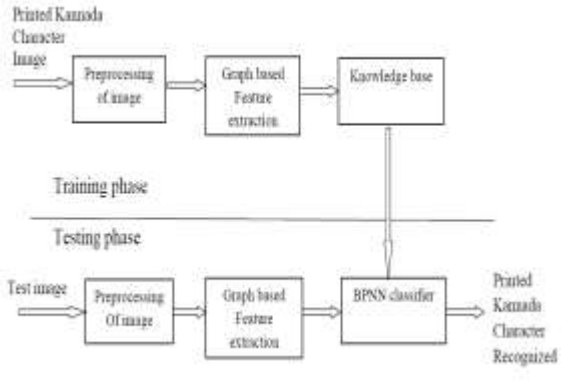


Figure 2.1 Block diagram of proposal model
The principle steps included in accomplishing OCR are as per the following:

- Preprocessing
- Feature Extraction
- Classifiers

A. Pre-Processing

The digital image containing printed complex Kannada character is given as an input for computer system. Examine the input character by utilizing a flatbed scanner or advanced camera. Pre-processing for the images generally includes the morphological operations. Such as evacuating low frequency background noise, normalizing the intensity of the individual particle images. The input will be in RBG format. The preprocessing techniques normally used in any text, image processing are shown in fig 3.2.

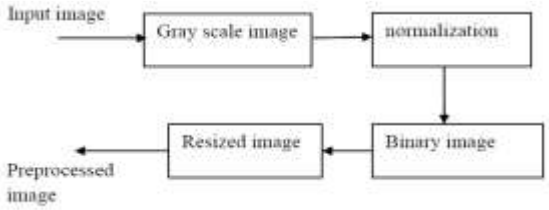


Figure 2.2 Block diagram of Preprocessing

The components of preprocessing are discussed bellow.

B. Gray scale image

The input printed Kannada image is in RGB format convert it into a grayscale image as shown in fig 3.3.

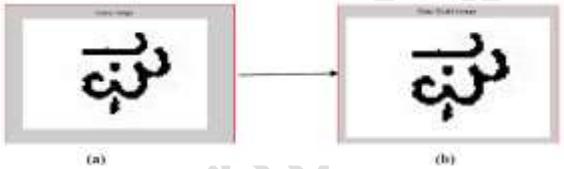


Figure 2.3 (a) grayscale image (b) normalized image

C. Normalization

Changing arbitrary sized images into a some standard size are known to be the normalization process. The Bicubic interpolation, linear sized normalization techniques and can be used for standard sized images as shown in fig 3.4

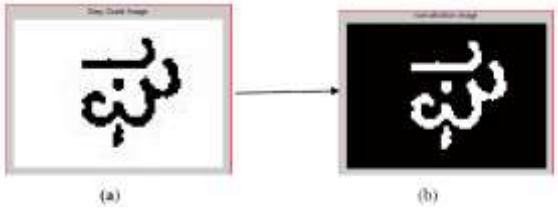


Figure 2.4 (a) grayscale image (b) normalized image

D. Binarization

Binarizing is a method of transforming a Kannada image into a gray scale Kannada image through threading. This technique suppresses background from the image. Binarized image shown in the fig 3.5

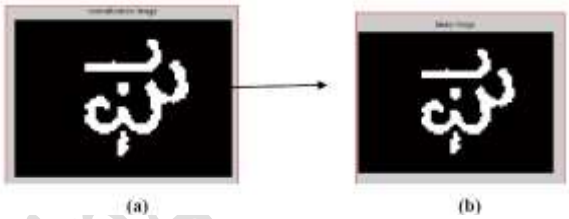
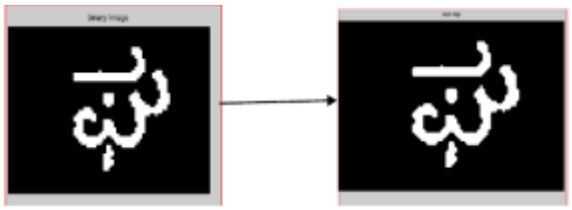


Figure 2.5 (a) normalized image (b) binary image

E. Resized image

Converting the printed Kannada character binary image data into some standard size as shown in the below fig 3.6.



(a) (b)

Figure 2.6 (a) binary image (b) resized image

E. Feature extraction

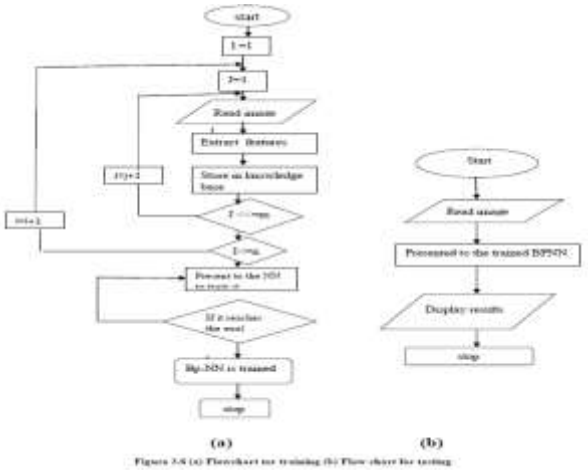
Extraction of the features from an image is important step in any recognition system, because the recognition accuracy totally depends on the features extracted. The main goal of a feature extraction method is to exactly get back the features. In the proposed system the global features such as number of nodes, number of lines, number of extreme points around an image, centroid along X and Y axis, top and bottom X and Y co-ordinate value are extracted.

F. Training

This includes developing a suitable neural network model (BPNN). Then the extracted features are input to the BPNN, which recognizes the different types of printed Kannada character images. The neural network architecture is trained itself accordingly. The training takes place such that the neural network learns that each entry in the input file has a corresponding entry in the output file. Flow chart for training BPNN is shown in the figure 3.6(a).

G. Testing

This is the step where we obtain appropriate result. In testing, input image from testing is selected and its features are extracted and given to the trained model, the trained BPNN model classifies given sample and produces output as type of recognized character and corresponding pattern. Flow chart for testing is shown in below figure.



III. EXPERIMENTAL RESULTS

A. FEATURES EMPLOYED

The feature extraction method step plays vital role in any recognition system because the recognition accuracy totally depends on the feature extracted. The heart of recognition for any character image is the extracting the features of that image. The main goal of feature extraction technique is to accurately regain the graph based features. The term "Feature Extraction" can be considered to encompass a very wide range of techniques and processes, ranging from simple ordinal / interval measurements derived from individual bands to generate, update and maintaining the discrete feature objects

Features extracted from the printed Kannada character image are the global features as follows.

B. Global Features

The graph based global features extracted in this work are as follows.

- **Nodes**: Number of nodes in a graph of printed Kannada character image.
- **Lines**: Number of lines between the nodes in a graph.
- **Extreme points:** x and y-co-ordinate values of 8 extreme points.
- **Centroid_x and Centroid_y:** Finding the Center location or centroid points (pixel value) of all regions in an image**.**
- **Xmax, Xmin**: these are minimum and maximum x-co-ordinate values among all the branch points.
- **Ymax, Ymin:** these are minimum and maximum y-co-ordinate values among all the branch points.

C. Graph Based Technique

A graph is a popular representation method in the data structure. A graph has main two factors Vertex and edges. After the design of graphical delegation of printed Kannada character the analogy between the two graphs is found. With the captured data, it can build up a mathematical graph of data. The graph is a major approach of the graph theory thus undirected weighted graphs are utilized. An undirected mathematical graph G is a systemized pair (V, E) in which V act as a set of vertices (nodes) and E subset V * V is a set of unordered pairs form V called set of edges of the graph G or it is a set of vertices that are associated with links called edges. For every edge links there should be only two vertices, and for every two adjacent vertices and they are connected with an edge are known to be a neighbor. To an every edge E it will assign some non-negative number w which is considered as the weight of an edge E. finally if every edges present in graph G have weights assigned to them, then the graph G is said as weighted graph.

By applying the graph based techniques on the printed Kannada character image 28 features have been extracted which are listed in the table 4.1. All the features extracted from the graph based techniques are used as inputs to the BPNN classifier to recognize a given character. A detailed description of the BPNN is given in the next chapter.

Table 4.1 List of all features

Si. no	Features	Si. No	Features
1	E1	15	E15
2	E2	16	E16
3	E3	17	B1
4	E4	18	B2
5	E5	19	C
6	E6	20	L1
7	E7	21	L2
8	E8	22	T1
9	E9	23	T2
10	E10	24	Np1
11	E11	25	Np2
12	E12	26	Np3
13	E13	27	Nl1
14	E14	28	Nl2

Where,
E1-E16: Eigen features of extreme points around character image.
C : Centroid pixel value of the image.

L1, L2: These are minimum and maximum x-co-ordinate values among all the branchpoints.
T1, T2: These are minimum and maximum y-co-ordinate values among all the branchpoints.
Np, Nl : number of points and the number of lines drown on an image.

D. Back prapogation Neural Network

Neural Networks

Artificial neural networks are used to finding the solution for the computational complex task, such as computer vision, image processing, and pattern recognition. Neural network classical applies to the low level image processing, image segmentation n-clustering techniques for image coding, face and object recognition, signature and character recognition, document examination, medical imaging, radar imaging, target identification, nonlinear image filtering, reconstruction, and image restoration. It can also apply in the three-dimensional motion estimation, object recognition, stereo vision and expert systems.

E. Backprapogation Neural Network

A backpropogation neural system which is a decent learning technique, and is a speculation of the delta standard. It requires a dataset of the desired output for many inputs, making up the training set. It is extremely used for feed-forward networks. Back propagation it relies on the activation function, which is used by the artificial nodes be differentiable

Creating BPNN

The neurons in the first layer or input node n=28, in this the number n indicates the features used in the input nodes. The number of neurons used in the output layer is 6 which is equal to the number of pattern classes.

The number of nodes in the hidden layer is calculated using the formula:

$$N = \frac{I + O}{2} + Y^{0.5}$$

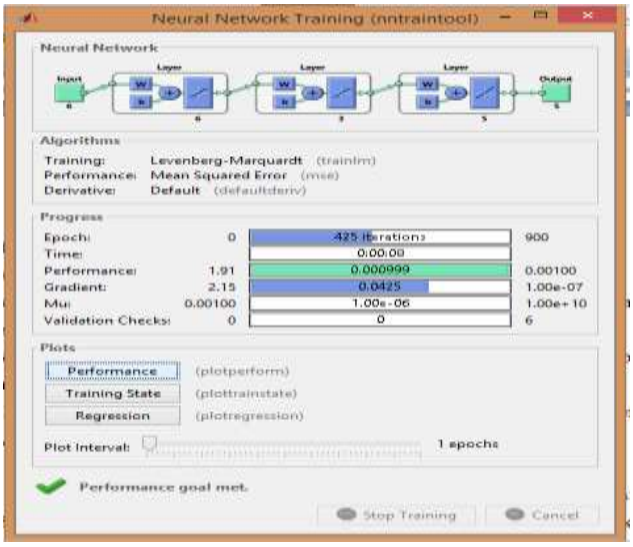
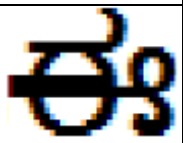


Figure 3.1 Run snapshot of BPNN

Output Pattern for Recognition

SI No.	Printed Kannada Character	Corresponding Pattern	Recognized Character
1		0 0 0 0 0 1	A
2		0 0 0 0 1 0	Aa
3		0 0 0 0 1 1	I
4		0 0 0 1 0 0	Ie
5		0 0 0 1 0 1	U

In this aspect, a various printed Kannada character which is not utilized in the training sets, are used to calculate the accuracy of perception. The process of recognition is repeated for various images which admit the trained and untrained images. Classification accuracy is calculated as.

$$\text{Classification accuracy} = \frac{\text{Number of correctly recognized printed Kannada characters}}{\text{Total number of testing printed Kannada characters}}$$

IV. CONCLUSION AND FUTURE WORK

The proposed system used graph based technique to extract features. Graph based methods have been used in many applications such as face recognition, hand recognition, etc. But not used for the printed Kannada character recognition. In this proposed work the graphical based features such as nodes, edges and Eigen vector of extreme points are used for recognizing the printed Kannada character. The Back Propagation Neural Network is used as classifier to recognize the printed Kannada character. The combination of graph based features and neural classifier gives the expected result. The experiment is carried out for the 50 different printed kannada characters and 46 characters are correctly recognized using graph based features. BPNN takes more time for training as the number of Kannada character increased in database. The accuracy of recognition can be increased by increasing the number of training images of each character.

Future work of this project includes an analysis of the new graph based features of Kannada character and combined with present feature vector to get higher accuracy. Directed graph can also be used to represent a kannada character. The proposed basic kannada character recognition can be used for kannada sentence/ language processing and understanding. Also this system can be used in handwritten kannada character recognition.

REFERENCES

[1] R Sanjeev Kunte and R D Sudhaker Samuel, "**A simple and efficient optical character recognition system for basic symbols in printed Kannada text,**" *S¯adhan¯a* Vol. 32, Part 5, October 2007, pp. 521–533. © Printed in India

[2] Thungamani.M1 Dr Ramakhanth Kumar P2 Keshava Prasanna3 Shravani Krishna Rau4" **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

[3] Mr.Nithya.E and Dr. Ramesh Babu D R" **OCR System for Complex Printed Kannada Characters**" ISSN: 2277 128X Available online at: www.ijarcsse. com

[4] Mamatha H.R, Sucharitha S and Srikanta Murthy K "**Multi-font and Multi-size Kannada Character Recognition based on the Curvelets and Standard Deviation**" International Journal of Computer Applications (0975 – 8887) Volume 35– No.11, December 2011

[5] Vishweshwarayya C. Hallur, Avinash A. Malawade, Seema G. Itagi " **Survey on Kannada Digits Recognition Using OCR Technique**" International Journal of Advanced Research in Computer Engineering & Technology (IJARCET) Volume 1, Issue 10, December 2012

[6] Umesh R S, Peeta Basa Pati and A G Ramakrishnan "**Set theoretic line segmentation and**

graph based strategy for bilingual Kannada-English OCR"

[7] G. G. Rajput, Rajeswari Horakeri, Sidramappa Chandrakant "**Printed and Handwritten Mixed Kannada Numerals Recognition Using SVM**" G.G. Rajput et. al. / (IJCSE) International Journal on Computer Science and Engineering Vol. 02, No. 05, 2010, 1622-1626

[8] Mamatha H R and Srikantamurthy K "**Morphological Operations and Projection Profiles based Segmentation of Handwritten Kannada Document**" International Journal of Applied Information Systems (IJAIS) – ISSN : 2249-0868 Foundation of Computer Science FCS, New York, USA Volume 4– No.5,October 2012 – www.ijais.org

[9] Niranjan S.K1, 3, Vijaya Kumar2,3, Hemantha Kumar G4, and Manjunath Aradhya V N5 **"FLD based Unconstrained Handwritten Kannada Character Recognition**" International Journal of Database Theory and Application Vol. 2, No. 3, September 2009

[10] Anjali Chandavale, Suruchi Dedgaonkar, Dr. Ashok Sapkal "**An Approach for Character Recognition Using Pattern Matching with ANN**" INTERNATIONAL JOURNAL OF SCIENTIFIC & ENGINEERING RESEARCH, VOLUME 3, ISSUE 10, OCTOBER-2012 1 ISSN 2229-5518.