IJISET - International Journal of Innovative Science, Engineering & Technology, Vol. 1 Issue 4, June 2014.
www.ijiset.com

ISSN 2348 – 7968

A Standardized Frame work for Handwritten and Printed Kannada Numeral Recognition and Translation using Probabilistic Neural Networks

N. Shobha Rani[1], Ashwini P. S.[2]

[1]*Department of Computer Science, Amrita Vishwa Vidyapeetham, Mysore Campus Bogadi, Mysore, Karnataka*

[2]*Department of Computer Science, Amrita Vishwa Vidyapeetham, Mysore Campus Bogadi, Mysore, Karnataka*

Abstract

Numeral recognition is considered to be very prominent in most of the Character recognition researches. With respect to applications like number plate recognition and document processing the numerals are composed as a part of number plate images/application form type document images. This paper mainly focuses on eliminating language barriers that may arise while comprehending the regional language numerals by a non-regional user at the time of number plate recognition or other application form type document processing with special reference to Karnataka state. An algorithm is devised by incorporating the capabilities of functionalities of fast Fourier features and probabilistic Neural networks to recognize and convert the handwritten and printed Kannada numerals to Roman numerals.

Keywords— *Printed and handwritten numeral recognition, FFT's, Probabilistic Neural Network (PNN), application forms processing.*

1. Introduction

Handwritten and printed Roman numerals are integral part of most commonly used documents in real world. The recognition of printed and handwritten numeral recognition has been considered has an emerging research area spanned from Optical character recognition [1]. Many of the documents like application form for reservations, vehicle number plate images of various states, admission forms in Govt. Schools or colleges, historical documents pertaining to a state and all other types of documents may coexists the numerals with various other text. Especially Roman numerals can be observed in most of the real time documents to be used in various organisations for variety of their needs. Beginning with a simple literal like date, postal pin codes, identity card numbers, account numbers, register numbers, PAN card numbers, vehicle numbers, page numbers in a book or document and many other scenarios the recurrence of Roman numerals is high. Roman numerals are considered to be most standard uniform notation of numerals to be followed and understood by individuals belonging to different states/regions. Thus Roman numerals are significant portions that reveal the accurate details of particular Place/individual very importantly based on a Place/individual's ID card numbers, postal pin codes and vehicle number plate recognition. In a well civilized country like India there may be individuals belonging to varied states/ regions and working together in various Govt./Private Organizations without any regional differences but with some linguistic barriers. Any individual can commonly understand the Roman numerals but not the regional language numerals like Kannada/Telugu/Tamil/Malayalam numerals etc.

Along with the increasing popularity of the Roman numerals even states like Karnataka, Andhra Pradesh and various other states in South India has been using their regional language numerals in fulfilling their various job and communication needs. Especially the individuals residing in rural areas are exposed more towards the regional languages, linguistic admirers and norms of certain state to use regional language numerals in some job areas also raised the need of recognition and conversion of the regional language numerals to Roman numerals to eliminate the linguistic barriers among different state individual.

This paper intends to devise algorithms for segmentation, classification and recognition of handwritten and printed numerals in application with fast Fourier features and probabilistic neural networks that provides the interface to recognize and convert Kannada numerals to Roman numerals. Even though most of the research has been carried out in the area of handwritten/printed Kannada numeral recognition yet there is a need of experimenting it further to handle the various critics that may interrupt the faster and accurate recognition becoz recognition of a numeral plays a very typical role in identifying a particular individual. The current systems related to printed numerals have achieved almost 99% of accuracy where as in case of handwritten numerals it is still lagging below 90%. A generic numeral recognition system that can recognize both printed and handwritten Kannada numerals that can work with variety of datasets related to

ISSN 2348 – 7968

vehicle number plates or numerals that may occur in all types of general purpose documents etc.

The handwritten and printed Kannada numerals that are used in various types of documents are different from the one that are used in current vehicle number plate recognition systems. The existing systems also cannot convert the Kannada numerals to Roman numerals. Moreover, recognition of handwritten numerals is difficult because of the high Variability in writing styles of different Persons. The general purpose documents that are used in real time require a uniform conversion to one language. The numeral recognizer cum converter devised by us can be able to deal with all types of Kannada numerals that are used in different work environments. Especially, handwritten numeral recognition system that is used in postal pin code recognition requires this enhancement. All these various factors motivated us to develop this enhanced framework for Kannada printed and handwritten numeral recognition and translation system that can perform faster and accurate functioning.

2. Literature Review

There are numerous experimentations that are performed in the area of Kannada numeral recognition systems. Results of few of the experimentations are as discussed below.

Dinesh Acharya et al. [1] have used the 10 segment string, water reservoir, horizontal/vertical strokes, and end points as potential features for handwritten Kannada numerals. U. Pal et al. [2] have proposed zoning and directional chain code for Kannada numerals recognition. S.V. Rajashekararadhya et al. [3] have proposed zone centroid and image centroid based angle feature extraction system for isolated Kannada numerals recognition. Dhandra et al. [4] have proposed pixel density features for handwritten and printed Kannada mixed numerals recognition. Dhandra et al. [5] have proposed spatial features and considered a feature vector of length 13 for handwritten exclusively for Kannada numeral recognition. Rajput et al. [6] has proposed an approach for classifying Kannada,Telugu and Devanagari numerals using local and global structural features and probabilistic neural network classifier. Sivanandham et. al. [7] has discussed the various types of neural networks and its features. Dinesh Acharya et. al [8] has devised a method of recognizing isolated handwritten numerals using structural features and K-means clustering algorithm for classifying it. Anita pal et. al. [9] has proposed an algorithm to automatically recognize handwritten numerals using neural networks that can reduce the classification and experimentation time. Rajashekarardhya et. al. [10] has used the support vector machines and global features for the recognition of isolated handwritten numerals.

From the literature survey, it is evident that recognition of handwritten numerals is often done with respect to one type of application. Numerals recognition is a fascinating area of research, it is required to design a robust, accurate and faster handwritten and printed Kannada numeral recognition and translation system suitable for fulfilling varying needs in real time.

3. Proposed Methodology

The proposed system is intended to design a faster and accurate Kannada numeral recognition and translation using fast Fourier features and probabilistic neural networks [6]. The various steps involved in proposed numeral recognition system comprised of pre-processing, segmentation, feature extraction and classification, recognition and translation. Fig 1. Depicts the details of steps involved in proposed system..

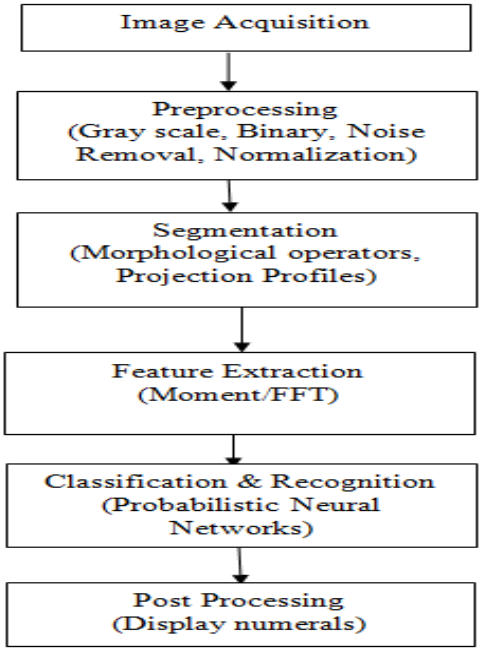

Fig. 1: Proposed System

A. Pre-Processing

The pre-processing contains a set of operations that are performed on scanned input image document. A handwritten document must be scanned and converted into a suitable format for further processing. Pre-processing consist of different functions to clean the image and make it appropriate for carrying out the recognition process accurately. The different functions are

ISSN 2348 – 7968

i. Grayscale

ii. Binarization

iii. Noise removal

iv. Normalization

B. Segmentation

Images containing text are of great use in the real world. Extracting or modifying any text in the image requires the text to be segmented out from the document image. Hence Segmentation is an imperative and significant stage of any character recognition system. It separates the document image's text into lines, words and numerals which is an input for further stages of character recognition [19].

Segmentation of handwritten Kannada numerals is very difficult because of its structural complexity. Not only the structural complexity but also the user is given a liberty to write the text in his/her own way which also introduces certain level of complexity in recognizing numeral shapes accurately. Rewritten text, scribbled and scratched text in the image document also increases the complexity. The proposed system has employed the horizontal and vertical projection profiles for the numeral segmentation [15] as it deals with only the isolated handwritten/printed numerals.

C. Feature Extraction

Feature extraction is very typical and important in any numeral recognition research. The distinguished features of the numerals are mainly used by the classifier in order to identify and classify the numerals. The features of numerals are extracted and converted into a vector form, through which the neural network can recognize the appropriate numerals.

In this section, we have devised a feature extraction algorithm used for extracting the features from the segmented input data which leads towards efficient classification and recognition.

The moment features are extracted from each segmented numeral and which is been trained to a probabilistic neural network. Around 50 references of the same type of numeral is trained to the neural network in order to have an accurate and efficient numeral recognition. The neural network then classifies the numerals based upon the appropriate match of the trained features with the tested features of concerned segmented input numeral.

Algorithm: Fast Fourier Features Extraction

Input: Pre-Processed and segmented handwritten numeral Image
Output: Feature vector for Classification and recognition

Step 1: Compute the fast fourier features of each column of the numeral image.
Step 2: Store all the moment value of that numeral in a vector
Step 3: Repeat 1 and 2 to extract the features of the image and store it into a vector.

D. Classification

A Probabilistic neural network is used for classifying and recognizing the handwritten numerals. The Extracted features are given as the input to neural net classification algorithm. There are around 20 classes starting from 0 to 9 including with printed and handwritten Kannada numerals. The neural classifier consists of two hidden layers besides an input layer and an output layer. The total number of neurons in the output layer is 20 as the proposed system is designed to recognize both printed and handwritten Kannada Numerals. The following sub section explicates the more details of probabilistic neural networks

Probabilistic Neural Networks (PNN):

Probabilistic neural network is a kind of radial basis network suitable for classification problems. The PNN classifier provides a good generalization ability and fast learning capability, which are crucial for handwritten numeral recognition system. This networks uses a radial basis transfer function to calculate a layers output from its net input. The architecture of Probabilistic Neural Net is made up of four units viz. input units, Pattern units[Class 0…9], Summation units and Output units as shown in the Fig. 2.

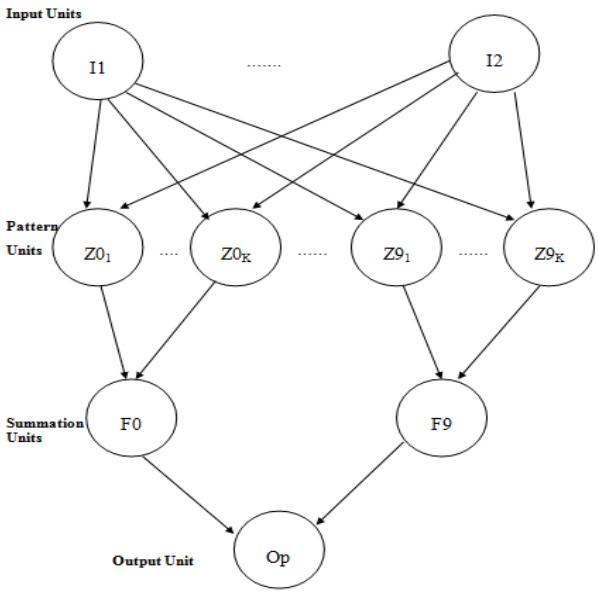

Fig. 2 Architecture of Probabilistic Neural Net

ISSN 2348 – 7968

When an input is presented, the first layer computes distances from the input vector to the training vector, and produces a vector whose elements indicate how close the input is to a training input. The second layer sums these contributions for each class of inputs to produce as its net output vector of probabilities. Finally, a complete transfer function on the output of the second layer picks the maximum of these probabilities, and produces a 1 for that class and a 0 for the other classes.

Algorithm: Probabilistic Neural Network Classifier
Input: Feature Vector of the segmented numerals.
Output: Recognition of the Numeral.

Step 1: Create the neural network.
Step 2: Configure the neural network by assigning weights or layers.
Step 3: Pass the feature vector with class index to the neural network
Step 4: Train the PNN classifier with feature vector.
Step 5: Classify the test images to its appropriate class label using PNN classifier with various radial values.
Step 6: Use the network to recognize the numerals.

4. Experimental results and discussion

In order to evaluate the performance of the proposed method, samples of around 200 printed and handwritten Kannada numerals were collected from different users. Datasets contains 4000 Kannada numerals references including both printed and handwritten numerals. These Numeral samples are used for training and testing purpose in the proposed system. The Fig. 2a and fig. 2b shows the user interface snapshots designed for the proposed system

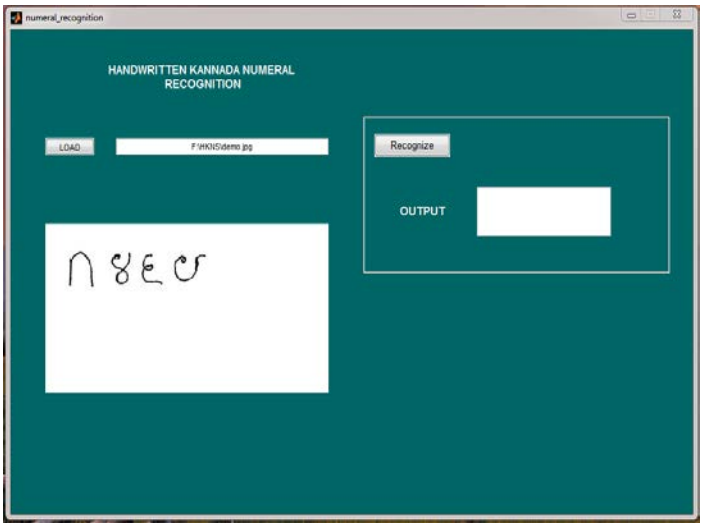

Fig. 2a: User interface of Numeral Recognition

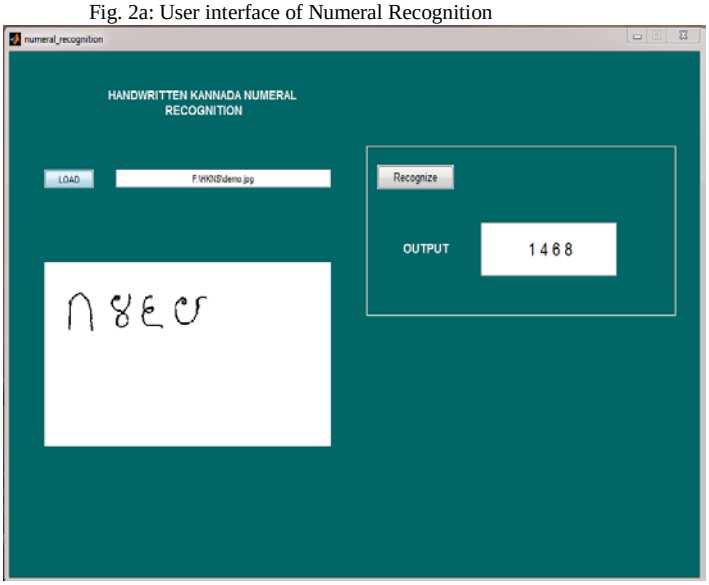

Fig. 2b: Output of recognized and converted numeral

5. Conclusions

Recognition approaches heavily depend on the nature of the data to be recognized. Even though most of researches are existing in the literature on Kannada Handwritten numeral recognition, However, there is no standard solution to identify all Kannada numerals with reasonable accuracy. Different approaches has been used in each phase of recognition process, where as each approach provides solution only for few numeral sets. Challenges still prevails in the recognition of normal as abnormal writing, slanting numerals, similar shaped numerals, curves and so on during recognition process.

However the proposed methodology proves to be effective and provides a standard solution for both printed and handwritten Kannada numeral recognition and translation.

This procedure proved to be very fast and accurate by the application of probabilistic neural networks to perform Kannada numeral recognition due to its capability of high noise tolerance.

References

[1] B. V. Dhandra, R. G. Benne, Mallikarjun Hangarge, “Kannada, Telugu and Devanagari Handwritten Numeral Recognition with Probabilistic Neural Network: A Novel Approach”, IJCA Special Issue on “Recent Trends in Image Processing and Pattern Recognition” RTIPPR, 2010.

ISSN 2348 – 7968

[2] N. Sharma, U. Pal, and F. Kimur, “Recognition of Handwritten Kannada Numerals”, 9th International Conference on Information Technology (ICIT'06), 2006.
[3] S.V.Rajashekararadhya and P.V.Vanaja Ranjan,”Neural network based handwritten numeral recognition of Kannada and Telugu scripts”,TENCON 2008 .
[4] B. V. Dhandra, Mallikarjun Hangarge, Gururaj Mukarambi, ”Spatial Features for Handwritten Kannada and English Character Recognition”, Special Issue on RTIPPR-10, International Journal of Computer Applications, pp.146-150, Aug-2010.
[5] 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 Proceedings published by International Journal of Computer
[6] Rajput, G.G., Mallikarjun Hangarge, “Recognition of Isolated Handwritten Kannada Numeral based on Image fusion method”, PReMI07,LNCS, Vol. 4815, Springer Kolkatta, pp153-160, 2007.
[7] S.N.Sivanandam, S.Sumathi and S.N.Deepa,” Introduction to Neural Networks”, The McGraw-Hill publication.
[8] Dinesh Acharya U, N. V. Subba Reddy and Krishnamurthy, “Isolated handwritten Kannada numeral recognition using structural feature and K-means cluster”, pp.125 - 129, IISN-2007.
[9] Anita Pal, Dayashankar Singh, “Handwritten English Character Recognition Using Neural Network”, International Journal of Computer Science & Communication Vol. 1, No. 2, pp. 141-144, July-December 2010.
[10] S.V. Rajashekararadhya, p. Vanaja Ranjan, “Support Vector Machine based Handwritten Numeral Recognition of Kannada Script”, 2009 IEEE International Advance Computing Conference (IACC 2009) Patiala, India,pp. 6-7, March 2009.
[11] Rafael C. Gonzalez Gonzalez, Richard E. Woods, Digital Image Processing, Pearson Education Asia, 2nd Edition, 2002.
[12] S.V. Rajashekararadhya, P. Vanaja Ranjan, and V.N. Manjunath Aradhya, "Isolated handwritten Kannada and Tamil numeral recognition: A novel approach", First International Conference.
[13] oufu Wu, Yongwu Wu, Gang Zhou, Jing Wu, "Recognizing Characters Based on Gaussian-Hermite Moments and BP Neural Networks", International Conference on Intelligent Computation Technology and Automation, ISBN 978-0-7695-4077-1, 2010.
[14] Birijesh K. Verma, "Handwritten Hindi Character Recognition Using Multilayer Perceptron and: Radial Basis Function Neural Networks", IEEE International conference on neural networks, Vol. 4, pp. 2111-2115, Nov. 1995.
[15] Janusz A. Starzyk and Nasser Ansari, "Feedforward Neural Network for Handwritten Character Recognition", IEEE symposium on circuit and systems, 1992.
[16] J. SuthaA, N. Ramaraj," Neural Network Based Offline Tamil Handwritten Character Recognition System", International Conference on Computational Intelligence and Multimedia Applications, pp. 446-450, IEEE 2007.
[17] Deepayan Sarkar, Report on Optical Character Recognition using Neural Network, http://homepages. cae. wisc. edu/~ece539/project/f03/sarkar-rpt. pdf
[18] Wikipedia on Normalization (Image Processing), http://en. wikipedia. org/wiki/Normalization_(image_processing)
[19] Wikipedia on Optical Character Recognition, http//en. wikipedia. org/wiki/Optical_character_recognition
[20] Diego Orlando, Optical Character Recognition, http://www. mathworks. com/matlabcentral/fileexchange/18169-optical-character-recognition-ocr.