ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 5, May 2017

# Offline Kannada Handwritten Word Recognition Using Locality Preserving Projections (LPP)

Pooja K Gowda, Chethan S , Harsha J , Rakesh J, Tanushree K N

Student, Dept. of CSE, Global Academy of Technology, Bangalore, Karnataka, India

Student, Dept. of CSE, Global Academy of Technology, Bangalore, Karnataka, India

Student, Dept. of CSE, Global Academy of Technology, Bangalore, Karnataka, India

Student, Dept. of CSE, Global Academy of Technology, Bangalore, Karnataka, India

Assistant Professor, Dept. of CSE, Global Academy of Technology, Bangalore, Karnataka, India

**ABSTRACT**: Nodes Offline Kannada Handwritten Word Recognition is a highly researched topic in present OCR. The offline kannada handwritten word(HWR) plays a vital role in the field of image processing. The recognition of offline kannada handwritten words are complex in nature when compared to online, as there is a much variations in the handwriting styles, type of paper used, quality of the scanner etc. As there is a large number of characters present in kannada language, it makes a open source problem for the researchers, Pre-processing, feature extraction and classification are the important steps carried out in offline kannada handwritten word recognition. For the feature extraction Locality Preserving Projections( LPP) method is used. Support Vector Machine (SVM) method is used for classification.

**KEYWORDS:** Kannada handwritten word, Pre-processing, Feature extraction, Classification, Locality Preserving Projections, Support Vector Machine.

## I. INTRODUCTION

Kannada is one of the oldest and is the official language of the Karnataka state in the Indian constituency. Kannada has its own script derived from Bramhi script. Kannada script has a set of 49 characters. They are classified into three categories: Swara (vowels), vyanjana (consonants), and yogavahakas. There are 13 vowels, 34 consonants and 2 yogavahakas. Image processing is nothing but processing of images using mathematical operations by using any form of signal processing for which the input is an image, a series of images, the output of image processing may be either an image or a set of words or parameters related to the image. Locality Preserving Projections (LPP) is used for linear dimensionality reduction. The first step in Recognition is going back to the roots of the languages and studying the words which make up the language. Each character is unique in many ways and if extraction unique features of the character in the word is done, the computer will be trained about that particular word.

## II. RELATED WORK

In [1] the, Kannada character recognition with application to automatic form processing is presented. Only 57 characters are considered for processing. Using suitable pre-processing techniques only handwritten characters are extracted. PCA and HoG are used for feature extraction. Performance of features is compared using neural network classifier. Results of pre-processing implemented using Visual Studio using OpenCV library are presented. HoG is found to have better recognition accuracy than PCA for large number of classes.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 5, May 2017

The proposed directional spatial features showed quite encouraging performance with respect to handwritten Kannada characters. Based on the KNN classifier 1000 numerals and 1400 vowels of Kannada character are classified. They have obtained 91.04% with KNN classifier. It is discussed in paper [2].
The edge segmentation for Kannada characters with probabilistic neural network referred in [3] can give very high efficiency than other existing Methods .The proposed approach is efficient in classifying different sizes of Characters and fonts. In this paper they obtained 99.17% accuracy.

The developed technique is tested with multiple images of Kannada Handwritten Text where the image undergoes the testing of each individual module.Once the testing is complete, the recognized characters are printed onto a text pad and the editable text file is checked to find the recognition rate.Accuracy rate is 80-85%.It is discussed in paper [4]
In paper proposed in [5] algorithm Normalized chain code and wavelet transform is executed on a database of 2800 Kannada vowels and 6800 isolated handwritten Kannada consonants images, with 200 images representing each character per class. Considering 50% for training and 50% for testing. Accuracy rate 90.14%.
In paper [6], feature extraction is performed by extracting features from wavelet transform. Also, features such as corner detection, correlation, quadrant density, aspect ratio and width feature are extracted as structural features. Artificial neural network classifier is used to recognize the handwritten Kannada characters and numerals. Training data set of 4350 samples is used to obtain consistent feature values. The proposed method is tested on a dataset of 1450 samples. The overall accuracy obtained for Kannada characters and numerals are 91.00% and 97.60%, respectively. In future, we will consider handwritten Kannada characters and numerals with more complexity.
In [7] paper comprehensive study of pre-processing, segmentation, feature extraction and classification techniques for character recognition is explored. In addition a novel idea and its efficacy for Kannada natural writer independent handwritten character recognition are explored. Various methods have been used in each phase of the recognition process, whereas each approach provides solution only for few character sets. Challenges still prevails in the recognition of normal as well as abnormal writing, slanting characters, similar shaped characters, joined characters, curves and so on during recognition process. Accuracy rate is 81.8%.

## III. PROBLEM STATMENT

Offline Kannada Handwritten Word Recognition (HWR) plays a major role in the field of image processing and pattern recognition. Compared to online recognition, handwritten words cannot be identified easily because of the variations in the handwriting styles, types of paper used, quality of scanner etc. Large number of characters present in kannada language makes it a open problem for the researchers.By extracting the density from binarized image, we will categorize and recognize the words.

## IV. METHODOLOGY

Image processing is a method to convert an image into digital form and perform some operations on . The main objective of the Recognizing the Kannada handwritten words using Locality Preserving Projection Method. Other objectives are, Collecting Handwritten Data samples. Pre-processing of Handwritten Dataset. Analyze structural features and patterns of Kannada characters and there by choosing the one with best accuracy. To propose Handwritten Character Recognition Model that recognises a single image of isolated Kannada words to the word document. Testing and performance evaluation by analyzing results of recognition engine.

## V. SYSTEM DESIGN

There are 5 important steps in offline Kannada Handwritten Word Recognition, they are shown in the figure 1

A. Data Acquisition.
B. Pre Processing.
C. Feature Extraction.
D. Classification.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 5, May 2017

E. Recognition.

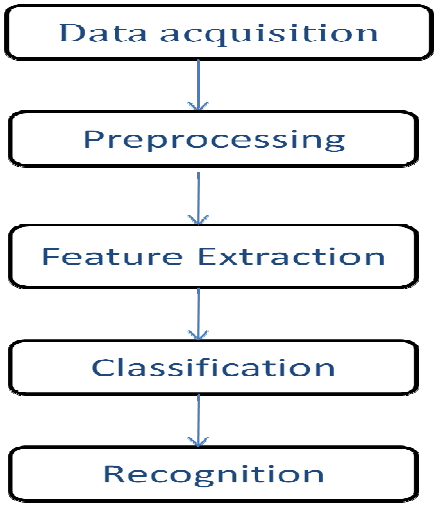


Figure 1: Five major steps in offline Kannada Handwritten Word Recognition

A. Data Acquisition :

The data collection is made for 47 kannada characters. In this we have our own dataset which consists of 30 districts of Karnataka. And those words sampling are taken from the 20 people. The handwritten words of 20 people are scanned, cropped saved in the format JPG, PDE.

B. Pre Processing :

The images of the handwritten word may have noise, disturbances, different color, or the word in the image might be small when compared to the background. Therefore we preprocess the image by removing noise, converting it to grayscale and binarizing the image.

**Step 1**: Noise removal

Digital images are prone to a variety of types of noise. Noise is the result of errors in the image acquisition process that result in pixel values that do not reflect the true intensities of the actual image. The below figure 2 represents the noise removal reduction.

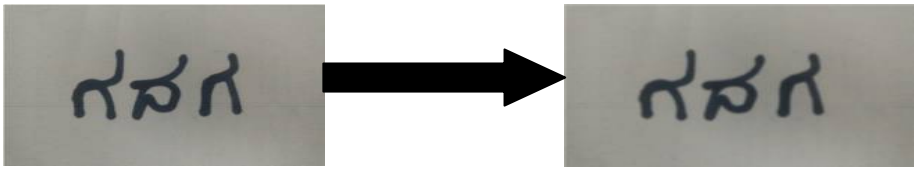

Figure 2: Noise removal

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 5, May 2017

After the removal of noise from the image, the next step is to convert the image to the grayscale. Since the dataset contains different pen color the image should be converted to grayscale. The below figure 3 represents the grayscale conversion.

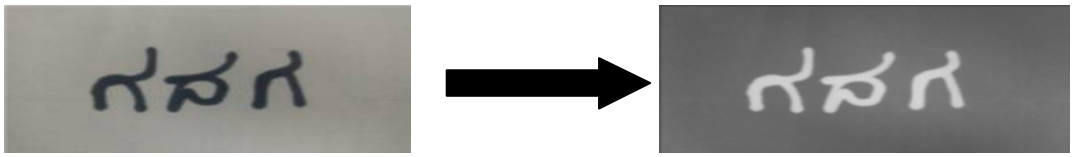

Fig 3: Gray Scale Conversion

**Step 3:** Binarization of the image

In this, the image is converted to the binary form that is black and white. The threshold constant will be set in between higher and lower values, where the higher and lower values will be set to white and black respectively. The below figure 4 represents the binary conversion.

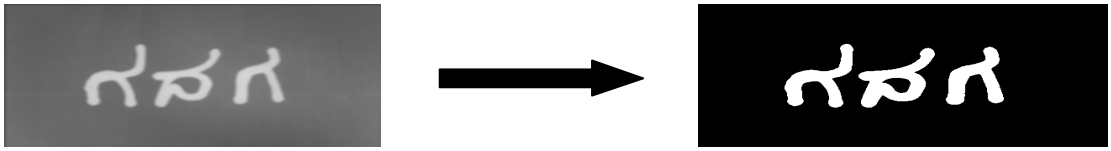

Fig 4: Binary conversion

C. Feature Extraction :

Feature extraction is the important stage in recognizing the handwritten word. The main objective of this is to abstract all the essential features of the scanned image. Locality Preserving Projections (LPP) is used to extract the required features of the image.

D. Classification :

During the classification phase the images are grouped according to their density .Support Vector Machine (SVM) is used for classifying the images.

E. Recognized Image :

Recognization is the last step of Offline Kannada handwritten Word. After the classification of the images according to their density, the accurate image of the input word is recognized and displayed.

## VI. EXPERIMENTAL RESULTS

The experiment was conducted on set of images as described in the table above, the dataset was prepared on the word images before the conduction of the experiment in the sampling phase all possible ways are recorded in 20 samples in average per word where all recorded in the data base against unique identifications. The experiment was conducted by writing the words on the white sheet in wide visible range later on some images were scanned using scanners and some were captured as images from the mobile camera later on those will be used as the input images in the format (jpeg, png, pdf). The system had identified the words and recognition rate was observed 80% in average for all the possible variations.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 5, May 2017

| Sl No | WORDS | NUMBER OF IMAGES | RECOGNIZED IMAGES | RECOGNIZATION RATE | AVERAGE |
|---|---|---|---|---|---|
| 1 | Bagalkote | 20 | 17 | 85% | 80% |
| 2 | Bangalore | 20 | 18 | 90% | |
| 3 | Belagavi | 20 | 15 | 75% | |
| 4 | Dharwad | 15 | 11 | 77% | |
| 5 | Haveri | 15 | 13 | 86% | |
| 6 | Mysore | 15 | 10 | 66% | |
| 7 | Dhavanageri | 15 | 12 | 80% | |

## VII. CONCLUSION

The simulation results showed that the proposed algorithm performs better with the total transmission energy metric than the maximum number of hops metric. The proposed algorithm provides energy efficient path for data transmission and maximizes the lifetime of entire network. As the performance of the proposed algorithm is analyzed between two metrics in future with some modifications in design considerations the performance of the proposed algorithm can be compared with other energy efficient algorithm. We have used very small network of 5 nodes, as number of nodes increases the complexity will increase. We can increase the number of nodes and analyze the performance.

## VIII. FUTURE WORK

In further, the concentration will be more on the data accuracy and extraction of the feature as density and classifications are concentrated more on efficient recognization. The more sampling results in more accuracy hence need to explore in automatic way of getting all tested images on live environment to the dataset for the further usage. Need to automate the sampling process by reducing the amount of manual interactions. By extracting the feature in form of density in number we can further add many more algorithms for classification and recognition where these algorithms needs to tuned for more accuracy. The same work need to bring on the mobile platform for offline recognition.

## REFERENCES

[1]. Shruthi Agarwal, Dr. Naveen Hemarjani, “Offline Handwritten Character Recognition with Devnagari Script”, IOSR Journal of Computer Engineering (IOSR-JCE), 2013.

[2]. Saniya Ansari, Dr. Udaysingh Sutar, "Study of Different Feature Extraction Techniques for Devnagari Handwritten Word Recognition", International Journal of Advanced Research in Computer Science and Software Engineering, 2015.

[3]. Vincent Neiger, “Handwritten digits recognition using OpenCV” Machine Learning in Computer Vision (CS9840), Western University, 2015.

[4]. S.A.Angadi and Sharanabasavaraj.H.Angadi, "Structural Features for Recognition Of Hand Written Kannada Character Based On Svm", International Journal of Computer Science, Engineering and Information Technology (IJCSEIT), Vol. 5, No.2, April 2015.

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

[6]. Dewi Nasien, Habibollah Haron, Siti Sophiayati Yuhaniz, “The Study of Handwriting Character Recognition (HCR) and Support Vector Machine (SVM)”, 2013.

[7]. George Margulis, “Optical Character Recognition: Classification of Handwritten Digits and Computer Fonts”, 2011.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 5, May 2017

## BIOGRAPHY

Pooja k Gowda, Currently pursuing BE degree in Computer science and engineering in Global Academy of Technology, Visveswaraya Technological University Belgavi. My research interests include Image Processing ,Cloud computing security and network security .

Chethan S Currently pursuing BE degree in Computer science and engineering in Global Academy of Technology, Visveswaraya Technological University Belgavi. My research interests include Image Processing ,Cloud computing security and network security .

Harsha J Currently pursuing BE degree in Computer science and engineering in Global Academy of Technology, Visveswaraya Technological University Belgavi. My research interests include Image Processing ,Cloud computing security and network security

Rakesh J Currently pursuing BE degree in Computer science and engineering in Global Academy of Technology, Visveswaraya Technological University Belgavi. My research interests include Image Processing ,Cloud computing security and network security .

Tanushree K N received Bachelor‘s degree in Computer Science and Engineering from Government Engineering College Visveswaraya Technological University Belgaviin 2011. Received Master degree in Computer Science and Engineering from East West Institute of Technology Visveswaraya Technological University, Belgavi in 2014 Currently Working as Assistant professor in the Department of Computer science and engineering in Global Academy of Technology Visveswaraya Technological University Belgavi. My research interests include Image Processing, Cloud computing security and network security and member of Computer Society of India (CSI)