

## Literature Survey on Devanagari Character Recognition

Rajneesh Gupta<sup>#1</sup> , Tejas Bhatkal<sup>#2</sup> , Amit Yadav<sup>#3</sup> , Dinesh Chaudhari<sup>#4</sup>  
Department of Electronics and Telecommunication  
Xavier Institute of Engineering

Prof. Panil Jain  
Department of Electronics and Telecommunication  
Xavier Institute of Engineering

### Abstract

*In terms of character recognition there are several papers reported and most of them are for English character. This paper focused on Devanagari character recognition from images. Devanagari script is used for many languages such as Sanskrit, Marathi, Nepali and Hindi. Lot of work has been done in character recognition and lot of work is to be done. Devanagari script should be given a special attention so that analysis of this language can be done effectively. This paper presents an approach for recognition of handwritten Devanagari characters, Fifty Eighth and written characters each having(vowels=220, consonant=2000, digits=2000) resulting in 94640 images will be used for this experimentation. The handwritten characters is scanned and on every individual characters image transform is applied so as to get decomposed images of characters. Character recognition provides a alternative way of converting manual text into digital format and reduces the dependence of man power.*

**Keywords:**Devanagari character recognition, hand written character recognition,Machine Learning.

### Introduction

Handwritten character recognition is gaining popularity for many years and attracting researchers for the purpose of potential application development. These Potential applications reduce the cost of human efforts and save the time. In the last few centuries English Character Recognition has been comprehensively studied and progressed to a level, sufficient to produce technology driven applications. Unfortunately, this is not same case for Indian languages which are complex in terms of structure and computations. Nowadays the speedily growing computational power may provide a solution for implementation of Indian Character Recognition methods. Digital document processing is achieving popularity for various application to office and library automation, bank and postal services, publishing houses and communication technology. Devanagri is composed of two Sanskrit word “Deva” and “Nagri”. Deva means God and Nagri means city. The Devanagari script is used for over 120 languages, including Hindi, Marathi other languages and dialects, making it one of the most used and adopted writing systems in the world. The Devnagri script is also used for classical Sanskrit texts.

### Proposed Technology

The decision making depends upon the model used for training and the testdata. The testdata is generated by preprocessing the dataset (in this case devanagiri) and using feature extraction algorithms. Feature extraction/selection is an important step for the dimensionality reduction. The testdata is used as reference for recognition of the character. The design cycle used in this study is shown. The images obtained from processing dataset exist in two groups, namely, training and

testing. After preprocessing and feature extraction the classifier makes decision after comparing it to the model referred by the training algorithm.

## Literature Review

### A. Deep Learning Based Large Scale Handwriting Devanagari Character Recognition.

In this paper, Ashok Kumar Pant and his group worked on a new public image dataset for Devanagari script: Devanagari Character Dataset (DCD). Their dataset consisted of 92 thousand images of 46 different classes of characters of Devanagari script segmented from handwritten documents. They also explored the challenges in recognition of Devanagari characters. Along with the dataset, they also proposed a deep learning architecture for recognition of those characters. Deep Convolutional Neural Network had shown superior results to traditional shallow networks in many recognition tasks. Keeping distance with the regular approach of character recognition by Deep CNN, they were focused on the use of Dropout and dataset increment approach to improve test accuracy. By implementing these techniques in Deep CNN, they were able to increase test accuracy by nearly 1 percent. The proposed architecture scored highest test accuracy of 98:47% on their dataset.

### B. Review of Research on Devanagari Character Recognition by : Vikas Dongre

In this paper they studied and investigated the direction of the Devanagari Optical Character Recognition research (DOCR), analyzing the limitations of methodologies for the systems which can be classified based upon two major criteria: the data acquisition process (on-line or off-line) and the text type (machine-printed or hand-written). The paper was arranged to review the DOCR methodologies with respect to the stages of the CR systems, rather than surveying the complete solutions. Although the off-line and on-line character recognition techniques have different approaches, they share a lot of common problems and solutions. Since it is relatively more complex and requires more research compared to on-line and machine-printed recognition, off-line handwritten character recognition was selected as a focus of attention in this article.

### C. Effective Techniques for the Detection, Extraction and Conversion of Devanagari Text from Traffic Panels by : Miss. Gayatri H. Khobaragade

In this paper, they presented the techniques for the traffic panel's detection, extraction and not least for conversion also. In this paper we use color segmentation for panel detection either green or yellow from the background image which gave nearly 95% accuracy. Next was text detection for that they used object extraction and segmentation concept which smartly separate all the objects (text) and gives 90% accuracy. Text recognition and Extraction gave 99% result but using their own dictionary only. By using OCR trained dataset it gave 80% accuracy but by applying concept of additional dictionary it gave finally 99% result. For text conversion they used N-gram algorithm whose work was to check the overall probability of single letter throughout the whole dictionary, that gave 99% result. They proposed that in future scope dictionary data can be improved for more accurate result as the more efficient dictionary gives more correct results.

### D. A Research Proposal On Recognition of Degraded Devanagari Text by : Rajni Kumari

In the following paper they proposed method in which The scanned documents was in Binarization (digitization) form. Single column printed text was considered. Text and non-text separation was primarily defined. They Proposed new algorithms to be tested on 500 data. It was scanned from old books, newspapers and magazines. Textual Degradation (Broken characters , Touching characters: Line touching overlapping and Inter -Character touching , Bleed through text) and Non-textual degradation ( Underline or other lines marks on the document and Back side visible. The proposed method was

supposed to restore the highly broken characters. The restoration was to be done in two phases one before segmentation and second after character segmentations. The proposed method used hybrid approach to handle broken characters and other degradation. Some of the existing methods such as gabor filter and active contour model along with devanagari structural properties was also used. The proposed method was supposed to solve the line overlapping segmentation problem. This method made use of the projection profiles and CC with some heuristics to segment the overlapped text documents in a robust way.

#### E. **Recognition Of Handwritten Devanagri Characters Using Machine Learning**

##### **Approach By : Pankaj Kale**

The system proposed was designed to recognize 50 simple devanagri characters. The process was included with data collection, preprocessing, segmentation, array formation and classification. The handwritten database of 50 characters were created and stored in 64\*64 sizes, from various 10 subjects. Characters were written using black 0.5 mm to 1.0 mm lead pen. The dataset was created with style and shape variations. they had taken 10 handwritten sentences for testing. The first step was preprocessing, during the preprocessing phase images were filtered for salt and pepper noise present in scanned documents. This was followed by binarization of the image. The second step was segmentation in which exclusion of shirorekha line was done. The sum of all pixels in a row were found and were performed for all rows so as to get a column vector. The row containing maximum number belongs to 'shirorekha' line. Then they Deleted that 'shirorekha' line by replacing its pixels with background pixels. Bounding box was applied to the characters in which the character and kana, matra, velanti etc. were separated. The segmented image was bounded to a size of 64\*64 pixels. These were written as separate images. Images were resized to 35\*35 pixels and further converted to an array of 1225 elements. That array was given as input to the neural network. Training matrix consisted of 1225 rows and 10 columns for the 10 subjects of the first alphabet. Similarly for the rest of 50 characters "swaras" and "vyanjanas" were trained.

#### F. **Optical Character Recognition for Hindi Language Using a Neural-network**

##### **Approach BY : Divakar Yadav**

In this paper, they suggest an OCR for printed Hindi text in Devanagari` script, using Artificial Neural Network (ANN), which enhances its effectiveness. One of the chief reasons for the deprived acknowledgment rate is fault in character segmentation. Hindi is the spoken by a lot of people in India, with more than 300 million users. As there is no severance between the characters of texts printed in Hindi as here is in English, the Optical Character Recognition (OCR) systems urbanized for the Hindi language bear a very pitiable identification rate. Pre-processing, character segmentation, feature extraction, and finally, classification and recognition are the major steps which are followed by a general OCR. The occurrence of touching characters in the scanned documents supplementary set hurdles to the segmentation process, creating a serious difficulty when designing a striking character segmentation technique. The pre-processing tasks measured in the paper are adaptation of grayscale images to binary images and image rectification. The basic symbols, bagged as the basic unit from the segmentation progression, are accepted by the neural classifier. For progress of the neural classifier, a back-propagation neural network with two hidden layers is used. The classifier is skilled and veteran for printed Hindi texts. In this effort, three characteristic extraction techniques are stated as histogram of projection based on mean distance, histogram of projection based on pixel value, and vertical zero crossing, have been used to recover the pace of detection. These feature extraction techniques are commanding adequate to pull out features of even indistinct characters/symbols. By applying the OCR the outcome of performance of approximately 90% accurate recognition rate was achieved.

### **G. Handwritten Devanagari Numeral Recognition by Fusion of Classifiers By : PrabhanjanS**

The main aim is to Recognize handwritten Devanagari numerals which has many applications in the field of postal address, document processing and so on. Due to its vast applications, many researchers are working towards development of effective and efficient handwritten numeral recognition. In this paper, we have proposed a hybrid method to recognize handwritten devanagari numerals. The proposed method uses, stacking approach to fuse the confidence scores from four different classifiers viz., Naïve Bayes (NB), Instance Based Learner (IBK), Random Forest (RF), Sequential Minimal Optimization (SMO). Also, the proposed method extracts both local and global features from the handwritten numerals.

### **References**

- [1] Ashok Kumar Pant, Prashna Kumar Gyawali, Shailesh Acharya, "Deep Learning Based Large Scale Handwritten Devanagari Character", International high performance building conference at Prude, 2014.
- [2] Vikas J Dongre, Vijay H Mankar, " A Review of Research on Devnagari Character Recognition", International Journal of Computer Applications (0975 – 8887) Volume 12– No.2, November 2010.
- [3] Kasturi Upasani, P. V. Baviskar, " A REVIEW OF DIFFERENT TECHNIQUES FOR DEVNAGRI SCRIPT RECOGNITION USING IMAGE PROCESSING", Global Journal of Advanced Engineering Technologies Volume 5, Issue 1- 2016 ISSN (Online): 2277-6370 & ISSN (Print):2394-0921.
- [4] Miss. Gayatri H. Khobaragade, " Effective Techniques for the Detection, Extraction and Conversion of Devanagari Text from Traffic Panels", International Journal of Computer Science and Mobile Computing, Vol.4 Issue.5, May- 2015, pg. 314-323.
- [5] Rajni Kumari, " A Research Proposal on Recognition of Degraded Devanagari Text", Pankaj Kale, Arti V. Bang, " Recognition Of Handwritten Devanagari Characters Using Machine Learning Approach", International Journal of Industrial Electronics and Electrical Engineering, ISSN: 2347-6982 Volume-3, Issue-9, Sept.-2015.
- [6] Divakar Yadav, Sonia Sánchez-Cuadrado, "Optical Character Recognition for Hindi Language Using a Neural-network Approach", J Inf Process Syst, Vol.9, No.1, March 2013
- [7] A. Deepika, S. Shalini, M. Sheela, " Handwritten Devanagari Numeral Recognition by Fusion of Classifiers", J Comput Eng Inf Technol Vol: 4 Issue: 2, July 06 2015.