

# DSP based Optical Character Recognition for Devnagari charaters.

Kajal Gade, Madhuri Mogal, Heena Sindhani, Prajakta Sonawane, Vaishali Khandve

1. Student , Computer Engineering, MET BKC IOE Nashik, Maharashtra, India
2. Student , Computer Engineering, MET BKC IOE Nashik, Maharashtra, India
3. Student , Computer Engineering, MET BKC IOE Nashik, Maharashtra, India
4. Student , Computer Engineering, MET BKC IOE Nashik, Maharashtra, India
5. Professor, Computer Engineering, MET BKC IOE Nashik, Maharashtra, India

## ABSTRACT

*Optical Character Recognition (OCR) is a research area that has attracted the interest of researchers for the past forty years. Character recognition is the mechanical or electronic conversion of scanned image of handwritten, typewritten or printed text into machine encoded text. Although the subject has been the center topic for many researchers for years, it remains one of the most challenging and exciting areas in pattern recognition. In our project, a robust DSP-based OCR is developed to recognize the Devnagari characters, display them as well as produce output in audio form.*

*There are millions of people in India who speak Hindi and write Devnagari script. Research in Optical Character Recognition (OCR) is popular for its application potential in banks, post offices, defense organizations and library automation etc. However most of the OCR systems are available now but they were only software based. In this work, a technique for OCR System for different five fonts and sizes of printed Devnagari script which will be hardware related. The recognition rate of the proposed OCR system with the document of image of Devnagari Script has been found to be quite high.*

---

## 1. Introduction

Machine simulation of human reading has become a major research topic since the introduction of digital computers. The motive for such an effort was not only the challenges in simulating human reading but also the possibility of efficient applications in which the data available on paper documents has to be translated into machine-readable format. Automatic recognition of printed and handwritten information found on documents like cheques, envelopes, forms, and other manuscripts has a variety of practical and commercial applications. Handwritten recognition of words is a system for converting the written text into usual words, which have an vital role in many human computer interface uses. Handwritten character recognition is an important and challenging field of Optical Character Recognition (OCR) Handwritten character recognition is a tough problem due to the great variations of writing styles, and so on so it is major area of research[5].

### i.) Alphabets

There are about 50 basic characters in the script. The combination of vowels and consonants is called Swaras and Vyanjanas respectively.

### ii) Vowels

Devanagari has 18 vowels out of which 11 are frequently used. Others can be seen in the Vedic and non-Vedic Sanskrit text.

### iii) Consonants

There are around 33 consonants in Devanagari script which are grouped phonetically.

## 2. Literature Survey

Following are the existing systems:

- 1) Ashutosh Aggarwal et.al. proposed system on Handwritten Devanagari Character Recognition Using Gradient Features.

### Gradient Features:

The gradient measures the magnitude and direction of the greatest change in intensity in a small neighborhood of each pixel. (In what follows, "gradient" refers to both the gradient magnitude and direction). Gradients are computed by means of the Sobel operator. The Sobel templates used to compute the horizontal (X) & vertical (Y) components of the gradient [1].

- 2) Neeraj Pratap and Dr. Shwetank Arya by applying technique of Template Matching, Support Vector Machine Classifier, Combination Classifier.

**Template matching:** It is one of the Optical Character Recognition techniques. Template matching is the process of detecting the location of a sub image called a template inside an image. Once a number of corresponding templates is found their centers are used as corresponding points to determine the registration parameters. Template matching determines the similarities between a given template and windows of the same size in an image and identifying the window that produces the highest similarity measure. The recognition rate of Template Matching is mainly depending on noise and image deformation. For improved classification Deformable Templates and Elastic Matching are used.

**Support Vector Machine Classifier:** For Data Classification, SVMs (Support Vector Machines) are a useful technique. A classification process generally involves separating the data into two sets, training and testing sets. Each instance in the training set contains one target value (i.e. the class labels) and several Attributes (i.e. the features or observed variables). The goal of SVM is to produce a model (based on the training data) which predicts the target values of the test data given only the test data attributes. Many researchers used SVM successfully [6].

- 3) Haidar Almohri, John S. Gray, Hisham Alnajjar proposed the technique of ART Neural Network for generating the character database and matching characters with that database. The technique is described below:

**ART Neural Network:** A training set database has to be generated for the network to be trained 700 sample characters chosen from the most popular Arabic fonts and sizes are used to generate the database. The 14 features described previously are extracted from this set of characters using MATLAB and the results are saved in a text file which could be used by Professional II/PLUS as the training set.

## 3. System Analysis and Design

### 3.1 System Analysis

This section contains control flow i.e. start to end process which is done by user to use the system and also contains data flow i.e. how document's quality is improved and how we get a clear as well as speech output.

### Control Flow

First the user will provide the scanned document to the system, then system takes it as an input and also interfaces with the DSP processor which performs various preprocessing tasks on the input document such as binarization, segmentation, feature extraction, pattern matching from the devnagari database, after then it will perform the post-processing tasks and also generates the readable output. After that the control passes to the speech generation module which converts the text in the audible format and then control flows to the another system on which the generated output is displayed and also speech out the the output.

### Data flow

The user can read or study any ancient literature or old documents which are originated from the devnagari script. User can read it without any difficulty while interpreting because of optical character recognition method which results into high quality output And also for blind user the system is beneficial ,it speech out the text which is generated as output.

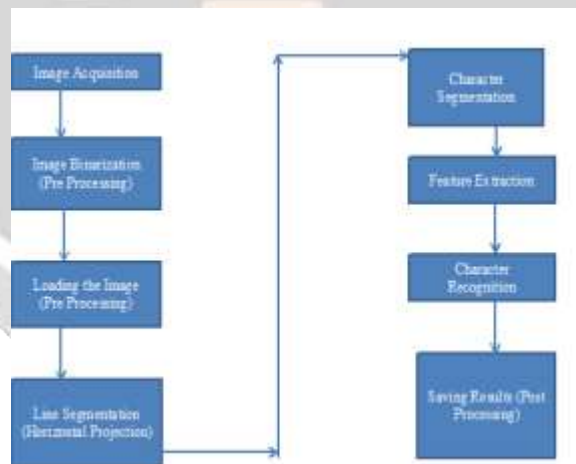
## 4. System Architecture

The system architecture is shown in Fig. . It consists the following methodology.

### Methodology

The proposed system will work in steps as follows:

1. Image Acquisition
2. Preprocessing ( Binarizing and Loading the image)
3. Line Segmentation
4. Character Segmentation
5. Feature Extraction
6. Recognition
7. Post Processing ( Saving the Results in a file)



**Fig a: Basic Flow Diagram**

### 1. Image Acquisition:

Handwritten Image is captured from optical scanner and are converted into digital images. The scanner is used is 300 dpi scanner. The image should have a specific format such as .jpeg, .bmp, .png etc. An Optical Character Recognition System for Indian Scripts

### 2.Pre-processing:

Pre-processing aim to produce data that are easy for OCR system. Pre-processing phase is applied to remove unwanted parts from the image by applying one or more method.

- 1) RGB to Gray
- 2) Threshold
- 3) Complement the Image
- 4) Morphological Operations like opening and closing
- 5) Binarization
- 6) Noise removal using filters

### **3. Segmentation:**

Segmentation of handwritten word is very important task, it is important to improve the accuracy of handwritten word since recognition system is heavily depend upon segmentation phase. Segmentation means to subdivide. technique involves Line, Word and Character segmentation .

### **4. Feature extraction:**

Feature extraction is one of the most important steps in developing a classification system. It is represented as a feature vector. Major goal of feature extraction to extract a set of feature which maximizes recognition.

### **5. Classification:**

The classification is nothing but matching of database characters with the input image characters. For classification purpose various classifier used like Support Vector Machine, K-Nearest Neighbors , Bayesian Classification, and Decision Tree Classification. Here used Template Matching for classification purpose.

### **Template matching:**

This is one of the simplest approaches to pattern recognition. In this approach a prototype of the pattern that is to be recognized is present. The given pattern that is to be recognized is compared with the stored patterns. The size and style of the patterns is not considered while matching.

### **6. Post Processing:**

Post-processing stage is the last stage of the proposed recognition system. Post-processing step involves grouping of symbols. The process prints the corresponding recognized characters in the structured text form. It generates the text file. It performs the association of symbols into strings which is referred to as group. In post processing other considerations like cost of errors are considered for the final decision.

### **Advantages**

- Computer systems equipped with such an OCR system improve the speed of input operation, decrease some possible human errors and enable compact storage, fast retrieval and other file manipulations.
- Using OCR, the handwritten and type written text could be stored into computers to generate databases of existing text without using the keyboard.

- It is an efficient to automate the process of digitizing the devnagari documents such as books, articles, etc. would be highly beneficial and commercially valuable.

## 5. CONCLUSION

Our system will present recognition of handwritten Devnagari characters free from normalization thereby giving flexibility and allowing size variation. The Devnagari script is used extensively at the grass root levels in many states of India. This Devnagari character recognition system can be used in developing Indian postal automation system, bank check processing system, as a reading aid to blind and it also has application in forensic science.

## 6. REFERENCES

- [1] Ashutosh Aggarwal, Rajneesh Rani, RenuDhir “Handwritten Devanagari Character Recognition Using Gradient Features” International Journal of Advanced Research in Computer Science and Software Engineering , Volume 2, Issue 5, May 2012
- [2] Ashwin S Ramteke, Milind E Rane “ A Survey on Offline Recognition of Handwritten Devanagari Script I
- [3] Divakar Yadav\*, Sonia Sánchez-Cuadrado\*\* and Jorge Morato\*\*\*“Optical Character Recognition for Hindi Language Using a Neural-network Approach”
- [4] Haidar Almohri, John S. Gray, Hisham Alnajjar “A Real-time DSP-Based Optical Character Recognition System for Isolated Arabic characters using the TI TMS320C6416T” Proceedings of The 2008 IAJC-IJME International Conference ISBN 978-1-60643-379-9
- [5] Naveen Shankaran, Aman Neelappa and C.V. Jawahar “ Devanagari Text Recognition: A Transcription Based Formulation” ICDAR, pp. 678-68, 2013
- [6] Neeraj Pratap and Dr. Shwetank Arya “A Review of Devnagari Character Recognition from Past to Future” International Journal of Computer Science and Telecommunications
- [7] Muhammad Arif Mohamad, Haswadi Hassan, Dewi Nasien., Habibollah Haron
- “A Review on Feature Extraction and Feature Selection for Handwritten Character Recognition” (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 6, No. 2, 2015 204 | P a g e w
- [8] Prerna Singh, “Radial Basis function for Handwritten Devanagari Numeral Recognition”, International Journal of Advanced Computer Science and Applications, pp. 126 -129, Vol. 2, No. 5, 2011.
- [9] R. Jayadevan, Satish R. Kolhe, Pradeep M. Patil, and Umapada Pal “Offline Recognition of Devanagari Script: A Survey” IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS—PART C: APPLICATIONS AND REVIEWS, VOL. 41, NO. 6, NOVEMBER 2011
- [10] Trupti R. Zalke, Prof. V.N. Bhonge “An Optical Character Recognition System for Indian Scripts” International Journal of Engineering and Technical Research (IJETR) ISSN: 2321-0869, Volume-3, Issue-5, May 2015.