Sampratilipi

CHAPTER 1: INTRODUCTION

Problem Definition

This project focuses on recognising handwritten characters of Devnagari script written in different styles and fonts.

Overview

Character recognition is the mechanical conversion of images of typed, handwritten or printed text into machine-encoded, with the from a scanned document or a photo of a document. Devnagari is an ancient and widely used script. The Devnagari script is used for over 120 languages In the last half century, the English character recognition was studied and the results were of such type that's it can produce technology driven applications. But the same approach cannot be used in case of Indian languages due to the nature of complication in terms of structure and computation. Now days there are different methodologies which are growing fastly in the area of Indian languages and character recognition. The offices, banks, schools and other organisations are working in the field of digital document processing. Devnagari is the national language of India and generally spoken by 600 million people in India. Devnagari should be given more special consideration for analysis and document retrieval due to its popularity. This paper is mainly concern for the people who are working in the Devnagari Optical character and it provides an over view about Devnagari character recognition system.

This method takes the input image from the user and applies various transformation. Then density based clustering is performed on this new image to identify the border points. Later matching of the input is done with the training data set.

The Devnagari Character Recognition System will be for the end user. This system will be designed to maximize the system's reliability by providing a means to assist humans in the

character recognition process which would otherwise have to be performed manually. By maximizing the efficiency and production, the system will meet the user's needs while remaining easy to understand and use.

More specifically, this system is designed to recognise the Devnagari characters written in different styles, fonts and sizes. The system also contains a database that comprises of a training set containing the variations in each character of the Devnagari script.

Different Phases of handwritten character recognition :

- Creation of Database
- Image Processing
  - Pixel processing
  - Thinning
  - Scaling
  - Noise reduction
- Density Based Clustering
  - Identify relevant border objects in four frames of a character image
  - Combine multiple clusters for further mapping with dataset
- Chain coding Technique
- Identification of characters – by comparing and Mapping to standard character database