Sanskrit Word Extraction

Jitendra V. Bhalodia

ABSTRACT *Handwritten Sanskrit word using a gradient feature. However, most of the current work in these areas is limited to English and a few oriental languages. The lack of efficient solutions for Indic scripts and languages such as Sanskrit has hampered information extraction from a large body of documents of cultural and historical importance. This paper describe two things first, we use gradient feature for edge detection for Sanskrit handwritten documents that makes use of domain knowledge of the script. Second, we use neural network to train the system by which words are easily recognized by the system.*

1.INTRODUCTION

India is a multi-lingual and multi-script country comprising of eighteen official languages. One of the defining aspects of Indian script is the repertoire of sounds it has to support. Because there is typically a letter for each of the phonemes in Indian languages, the alphabet set tends to be quite large. Most of the Indian languages originated from Brahmin script. Sanskrit is the most popular script in India. It is an ancient language and no longer spoken but written materials still exist. It is very expressive language, which has been influenced and enriched by Dravidian, Turkish, Farsi, Arabic, Portuguese and English. It has 13 vowels and 33 consonants. This Technology has been improving much under the purview of pattern recognition and image processing since a few decades. Hence various soft computing methods involved in other types of pattern and image recognition can as well be used for DOCR (Devanagari optical character recognition). Neural Networks are recently being used in various kind of pattern recognition. It plays an important role in handwritten character recognition.

2. CHARACTER MODELING

The Sanskrit language consists of 53 characters (17 vowels, 36 consonants) and is written from left to right.

3. CHARACTER RECOGNITION SYSTEM

Character recognition by neural network Replacing the recognized characters by standard fonts.

Assembling all the separated characters in the same order as they appeared in the input image to give final output. The process of a HCR of Devanagari script can be divided into phases as shown in Figure 3.
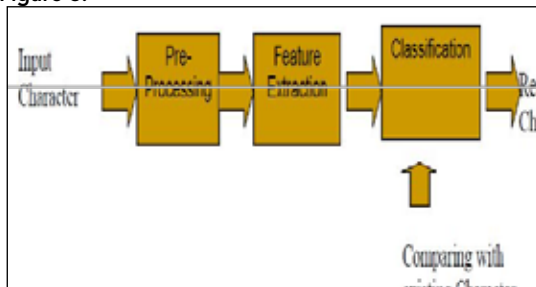
The Character Recognition System must first be created through a few simple steps in order to prepare it for presentation into MATLAB. The matrixes of each letter of the alphabet must be created along with the network structure. In addition, one must understand how to pull the Binary Input Code from the matrix, and how to interpret the Binary Output Code, which the computer ultimately produces the Character Matrixes. A character matrix is an array of black and white pixels; the vector of 1 represented by black, and 0 for white. The steps involved in recognition of character are: Matrix generation, Thinning of Binarised Image, Windowing, Feature Extraction, Training phase and database of trained neural networks for each character

4. FEATURE EXTRACTION

In this paper, we proposed a Sanskrit word recognition using sober operator Generally speaking, the gradient Feature represents local characteristic of a character image properly.

Gradient Calculation

Each image of character is normalized into 32x32 sizes. The gradient operator, named sobel operator is used to calculate the gradient. The Sober operator uses two templates to compute the gradient components in horizontal and vertical directions, respectively.

5. NEURAL NETWORK

Recognition of handwritten characters is a very complex problem. The characters could be written in different size, orientation, thickness, format and dimension. This will give infinite variations. The capability of neural network to generalize and insensitive to the missing data would be very beneficial in recognizing handwritten characters.

6. CONCLUSION

Handwritings of different person are different; therefore it is very difficult to recognize the handwritten characters. Handwritten Character recognition is an area of pattern recognition that has become the subject of research during the last some decades. In this paper, a system for recognizing handwritten Sanskrit Word has been developed. To develop this system we use sobel operator for edge detection The Sobel operator is used in image processing, particularly with edge detection algorithms. Technically, it is a discrete differentiation operator, computing an approximation of the gradient of the image intensity function. A lot of efforts to developing new feature extraction techniques or modifying the existing feature extraction techniques.

REFERENCE [1] Verma B.K, "Handwritten Hindi Character Recognition Using Multilayer Perceptron and Radial Basis | | [2] D.S. Yeung, "A Neural Network Recognition System for Handwritten Chinese Character Using Structure Approach," . | [3] Dayashankar Singh, Maitreyee Dutta and Sarvpal Harpal Singh, "Comparative Analysis of Handwritten Hindi Character Recognition Technique". | [4] Dayashankar Singh, Maitreyee Dutta and Sarvpal H. Singh, "Neural Network Based Handwritten Hindi. | [5] Weipeng Zhang and Yuan Yan Tang,"Handwritten character recognition using combined gradient and wavelet", IEEE International Conference on Neural Network, 2006 | [6] DS Yeung, "A neural network recognition system for handwritten chinese character using structure approach"