Abstract

Today, Handwritten Character recognition is one of the challenging computational processes. Some computational fields like artificial intelligence, expert systems have provided an important role in recognition of handwritten characters. There is competition between the speed and efficiency. The human mind can easily decipher these handwritten characters easily, accurately and speedily. The human mind can do it because of the presence of densely neural network in his mind.

The world is fast moving towards digitalization. In the age of super-fast computational capabilities, everything has to be made digitalized so as to make the computer understand and thereby process the given information. Optical character recognition is a method by which the computer is made to learn, understand and interpret the languages used and written by the human beings. It provides us a whole new way by which computer can interact with human beings, in their own languages. Such time consumed and error free processes have been enlightened by the invention of OCR (optical character recognition) systems that read handwritten or printed data by recognizing them at high speed, by reading one character at a time. However, the current OCR systems capabilities are still inadequate and only small fraction of data can be entered into the computer by them. And so lots of efforts are still needed to enable them to read printed and handwritten characters more accurately. A substantial amount of work has been done on foreign languages such as English, Chinese etc. however on Indian script, the research work is comparatively lagging; most of the research work available is mainly on Devanagari and Bangla script. The work on other Indian script is in beginning stage.

In this thesis work we have proposed offline recognition of isolated handwritten character of Marathi derived from Devanagari. We have also extended our work by applying the same methodology to recognize handwritten Marathi Compound Character. For the proposed work we considered 36 basic characters and 45 compound characters. All these character are assumed to be isolated and having header lines on the top of every character.
The collection of handwritten samples of characters comprises of 250 samples of each 36 basic character and 45 compound character forming total size of 21600 samples. Samples of 36 basic characters and 45 compound characters were used to write by writer of different professions. These data samples were recorded on special design paper sheet. The dataset are scanned and store in .jpg , .gif and also in .tiff format.

After scanning, in preprocessing stage, the samples are converted to gray scale images. Then gray scale image is converted into binary image. We preprocessed these samples of characters by applying many preprocessing techniques like median filtration, image dilation and some morphological operation to remove different types of irregularities and enhance the sample image. We have used 60 x 60 pixel sized normalized images of characters to extract features.

Before extracted features, the character is pre-classified into proper category. The pre-classification is based on global and local structural features. The global feature consists of presence of vertical line, position of vertical bar in the character and enclosed region in the character. The local feature consists of the end points and junction position in the character. On the basis of global feature, the character is classified into three major categories based on the presence of vertical bar i.e. a) character with vertical bar at right (VEB: Vertical End Bar), b) character with vertical bar at middle (VMB: Vertical Mid Bar), and c) the character with absence of vertical bar (NB: No Bar). Vertical bar at right are further classified into two categories based on whether the vertical bar and rest of the character are connected or not to the bar.

After preprocessing and pre-classification we have extracted the relevant features. Different methods of feature extraction are generally used and many issue needed to consider while extracting features. Basically three types of moment features are considered Geometric Moment, Zernike Moment and Legendre Moment.

To select a set of appropriate attributes of features from the interested object for the purpose of classification, it has been among the fundamental problem in the design of pattern recognition system. One of the solutions, the utilization of moments for character recognition has received considerable attention in recent years. In this research, the new techniques derived to increase the accuracy and the efficiency in moment computing is addressed. Based on these developments, the significant improvement on image recognition using Zernike moments and Legendre moments has been achieved. This feature has not been explored for the use of offline
handwritten character recognition especially for Marathi derived from Devanagari character.

In this proposed work we have used Geometric Moment, Zernike Moment and Legendre Moment features which are used to form feature set. These feature set are used for classification of basic and compound character.

Character sample are classified in specified category on the basis of structural feature with help of classifiers. For classification purpose we have used three types of classifiers Support Vector Machine (SVM), Multilayer Perceptron (MLP) and k-Nearest Neighbour (k-NN) to recognize Handwritten Marathi derived from Devanagari character.

The experiment that is done was possible to enhance the performance of the system by placing the character in particular category and then creating the feature set. SVM have proved that as a better classifier than MLP (64.92%, 53.35%) and k-NN (47.73%, 48.96%) for Geometric Moment as a feature extraction technique with recognition rate of 67.32% for Basic Character and 57.10% for Compound Character. For Zernike Moment as feature extraction technique, SVM has proved to be best than MLP (74.03%, 73.13%) and k-NN (59.87%, 60.94%) with recognition rate of 98.46% for Basic Character and 98.32% for Compound character. For Legendre Moment as feature extraction technique, SVM has proved to be better than MLP (70.96%, 70.76%) and k-NN (58.35%, 57.96%) with the recognition rate of 96.36% for Basic Character and 91.75% for Compound character.