Chapter 7

Conclusions and Future Scope

The techniques for the recognition of handwritten Hindi text by segmenting and classifying the characters have been proposed in this thesis work. The problems in handwritten Hindi text written by different persons are identified after carefully analyzing the text. To solve these problems new techniques have been developed for segmentation, feature extraction and recognition.

In the present work, four new segmentation algorithms have been proposed. These include text line segmentation, segmentation of half characters, segmentation of lower modifiers and segmentation of touching left modifier from consonant in the middle region of the word. A new technique based on header line and base line detection to segment the overlapped lines of text in handwritten Hindi text have been proposed. Determination of the header line is very tough. The position of the header line in particular line of text and header line in a particular word of the same line may vary.

Determining the presence of lower modifier, presence of half character, touching left modifier with consonant in the middle region or to determine the presence of touching characters is very arduous. The new threshold values for their presence in the word have been proposed. For segmentation of half characters from consonants structural properties of the text are considered. For segmentation of lower modifiers, a new technique based on shape of lower modifiers is proposed. A technique based on position and length of the left modifier is proposed for segmentation of left modifier from touching consonant in the middle region of the word. For the validity of the
algorithms, the proposed algorithms are also tested on printed Hindi text and obtained pleasing results.

After the segmentation of text, the features are extracted for recognition. A new feature set based on topological features or structural properties of the text has been proposed.

A new technique called merging of features for the feature extraction has been proposed in the present work. A particular feature of particular character may depend upon other feature of the character. In such cases, next feature is extracted only if previous feature is available otherwise not. It leads to reduce in number of features to be extracted. Further, the problems in feature extraction are identified and many heuristics are applied to solve those problems. The overall results obtained with proposed algorithms for segmentation and recognition of handwritten Hindi text is very challenging.

SVM and Rule based classifiers are used for the classification of characters of handwritten Hindi text.

7.1 Contributions of the Work

This thesis has made the following contributions in the field of handwritten Hindi text recognition:

i) To the best of researcher’s knowledge, this is the first attempt towards the development of OCR for handwritten Hindi text.

ii) New techniques are proposed for segmentation of overlapped line of handwritten Hindi text, segmentation of conjuncts (half characters), segmentation of lower modifiers and segmentation of touching modifiers or consonants in the middle region.
iii) A new feature set has been proposed which contains slant and size invariant features. The topological features are extracted which are very robust.

iv) A new technique is proposed for feature extraction to increase the speed of recognition. All the features are not extracted from each character, only main features and unique feature of that character are extracted.

v) A new technique is proposed for word recognition.

vi) Rule based classifier and SVM classifier is used for character recognition.

7.2 Future Scope

The proposed algorithms used for segmentation of handwritten Hindi text can be extended further for recognition of other Indian scripts.

- The proposed algorithms of segmentation can be modified further to improve accuracy of segmentation.
- New features can be added to improve the accuracy of recognition.
- These algorithms can be tried on large database of handwritten Hindi text.
- There is a need to develop the standard database for recognition of handwritten Hindi text.
- The proposed work can be extended to work on degraded text or broken characters.
- Recognition of digits in the text, half characters and compound characters can be done to improve the word recognition rate.
Bibliography


Bansal, V. (1999), “Integrating Knowledge Sources in Devanagari Text Recognition”,


Documents”, Proceedings of the Tenth International Workshop on Frontiers in Handwriting Recognition, pp. 515-520.


