Abstract

A script independent character segmentation from word images technique has been reported here. Word to character segmentation is an important preprocessing step of optical character recognition process. But in case of handwritten text, presence of touching characters decreases the accuracy of the technique of the segmentation of the characters from the word. In this paper, segmentation of handwritten word of four different scripts namely, Bangla, Devanagri, Gurmukhi and Syloti are considered as the test samples. All these scripts are characterized by the presence of a distinct line along the top of the most of the characters forming the words, called the headline or Matra. Unlike English script, the characters of these handwritten scripts and its components often encircle the main character, making the conventional segmentation methodologies inapplicable. For the segmentation technique two fuzzy features, to identify the Matra region and potential segmentation point, are used here.
Experimental results, using the proposed segmentation technique, on sample of 400 handwritten word images containing all the above mentioned scripts of Bangla, Devanagri, Gurmukhi and Syloti show a success rate of 95.41%, 93.61%, 91.23% and 92.37% respectively.

Reference


Index Terms

Computer Science  Pattern Recognition
A Script Independent Technique for Extraction of Characters from Handwritten Word Images

Key words

Character segmentation
handwritten word images

Script independent technique

Fuzzy features