Abstract

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. Hence OCR has been a topic of interest for researchers all around the globe in the past decade and research paper involving OCR is increasing day by day. It is seen that efficient algorithms have increased the speed and accuracy of character segmentation and recognition. A substantial amount of work has been done on foreign languages such as English, Chinese etc. but very few paper are there for Indian languages baring a few for Hindi and Bengali. Hence our research work was directed towards development of a novel algorithm for Devanagari character segmentation for Hindi. Hindi is one of the eighteen languages recognized by the Indian constituency. It is also one of the oldest languages and is spoken by millions of people in India. Segmentation and Recognition of this
particular language is difficult because of the presence of complex connected characters and presence of shirorekha. A novel approach has been proposed for the segmentation of the connected character.

References


Index Terms

Computer Science | Pattern Recognition

Keywords

OCR, Shirorekha, Devanagari character segmentation.