ABSTRACT

OCR can contribute immensely to the advancement of an automation process and can improve the interface between man and machine in many application like reading aid for the blind, automatic reading.Optical character recognition (OCR) refers to a process of generating a character input by optical means,for recognition in subsequent stages by which a printed or handwritten text can be converted to a form which a computer can understand and manipulate.Segmentation of lines is a preprocessing step for OCR. Segmentation of Kannada script poses challenges due to additional modifier characters,inter and intra word gaps.

A single column printed document may contain different font sizes and font styles like the headings with different font sizes,italicized sentences and sentences with different font sizes. Single column different documents with different font styles and different font sizes are considered for segmentation of lines.Both foreground and background information are used for accurate line segmentation.Segmentation methodologies such as morphology based approach,Horizontal projection profile and bounding box is used for extracting individual lines from printed Kannada documents,such that the proposed method takes care of the challenges posed by Kannada scripts accurately.