In a multi-lingual country like India, script segmentation or script separation of the multi-script in an image of a document page is of primary importance for a script identification system. For script segmentation of such a document page, it is necessary to segment multi script forms before running individual OCR of the script. In this paper we present a technique for script segmentation of the individual text line for printed in Indian language document. Our line wise script segmentation approach is Horizontal Projection Profile based script segmentation. A prototype of the system has been tested on printed Indian language lines of script and an average accuracy of 99% has been achieved.

References

- Zhou L, Y Lu and C L Tan, "Bangla/English script Identification based on analysis of connected component Profiles," In Proc. 7th DAS, 2006

Index Terms

Computer Science
Text Processing
Keywords
Script line documents  OCR