Abstract

A generic approach for the separation of overlapping and touching lines within handwritten text document images is proposed in this paper. Presence of touching or skewed that arises due to ascenders or descenders and style of writer makes text line extraction a difficult task. The approach is based on histogram and connected component analysis. The proposed method is a three stage approach wherein non overlapping lines are extracted during the first stage and separation of oriented and touching lines occurs during second and third stages respectively.
Average height of a text line computed using histogram profile forms the basis for text line segmentation. The proposed method has been evaluated on 120 handwritten documents written in English, Devanagari, Kannada, Telugu, and Malayalam scripts containing non-overlapping and overlapping or touching occurrences.

References

- Vikas J Dongre and Vijay H Mankar. DEVNAGARI DOCUMENT SEGMENTATION USING HISTOGRAM APPROACH. International Journal of Computer Science, Engineering and Information Technology (IJCSEIT), Vol. 1, No. 3, August 2011. 4
- Nazih Ouwayed, Abdel Belaid and Francois Auger. General Text-line Extraction


Transform based Technique for Text Segmentation"; JOURNAL OF COMPUTING, VOLUME 2, ISSUE 2, FEBRUARY 2010, ISSN 2151-9617.
- KALYAN TAKRU and GRAHAM LEEDHAM &quot;SEPARATION OF TOUCHING AND OVERLAPPING WORDS IN ADJACENT LINES OF HANDWRITTEN TEXT&quot;.
Proceedings of the Eighth International Workshop on Frontiers in Handwriting Recognition (IWFHR'02) 0-7695-1692-0/02 $17.00 © 2002 IEEE.

Index Terms

Computer Science
Image Processing

Keywords
Handwritten Document; Text-line; Segmentation; Histogram; Connected Component.