Abstract

Baseline detection and line segmentation are essential preprocessing steps of any OCR system. In this paper we have proposed a robust and fast method for baseline detection based on projected pattern analysis of Radon Transform. The algorithm has been tested on more than 350 samples including both printed and handwriting of Persian/Arabic, English and also multilingual documents. Obtained results indicate that in spite of narrow interline spaces and noisy components our method is capable to extract baseline in documents precisely. In addition, in the case of multi-frequencies pattern, it has been shown that proposed method can reach its performance to accurate detection of base lines.
Journal of Computer Applications (IJCA) (0975 – 8887), Volume 7– No.3.


- Y. Li, Y. Zheng, D. Doermann, and S. Jaeger, “A new algorithm for detecting text line in

Index Terms

Computer Science  Image Processing

Keywords

Optical Character Recognition  Document Analysis  Multilingual Documents
Radon Transform

Neural Networks