Abstract

The segmentation of unconstrained handwritten text lines into words is an important stage in
word recognition systems. This paper addresses a methodology to overcome the challenges,
which are amplified by the non-uniform spaces between words and overlapping components by
using a few statistical approaches. The system was developed using Java 2 and ImageJ tool.
In this approach, a text line image is scanned vertically, holding only the spatial information. A
scheme based on distance metrics and gap classification into inter-word gap and intra-word gap is presented. The threshold value is determined by using arithmetic mean, inter-quartile mean or trimmed mean based on the variation in the text. A pre-processing of removal of noise and correction of skew angle and dominant slant angle were done to improve the recognition accuracy. The system was illustrated with a few cases. A quantitative analysis of the experiment done on the system by using 1100 text lines from IAM database achieved an accuracy of 96.72% and found the system faster and reliable. Further, the proposed method is compared with the contour based and non-contour based techniques.

References

Index Terms

Computer Science

Image Processing

Keywords

Inter-quartile Mean; Projection Profile; Connected Component; Distance Metrics