Abstract

Segmentation of the text lines in an un-constrained handwritten documents still a challenging task because handwritten text lines are often un-uniformly skewed and curved, and the space between lines is not obvious. In this paper, we propose a text-line segmentation algorithm based on clustering using threshold. The connected components of document image are grouped, from which text-lines are extracted dynamically by coloring all the text-lines.
Text Line Segmentation of Handwritten Documents using Clustering Method based on Thresholding Approach

References

- Seni G., Cohen E. (1994), External word segmentation of off-line handwritten text line, pattern Recognition, 27, Issue 1, January, pp 41-52
- Marti U., Bunke H. (1999), A full English sentence database for off-line handwriting recognition, Proc. 5th
- F. Yin, C. L. Liu, (2007), Handwritten text line extraction based on minimal spanning tree clustering, Proc. 5th Int. Conf. on Wavelet Analysis and Pattern Recognition, Vol. 3, pp. 1123-1128.
- Fei Yin, Cheng-Lin Liu, Handwritten Text Line Segmentation by Clustering with Distance Metric Learning, National Laboratory of Pattern Recognition (NLPR), Institute of Automation, Chinese Academy of Sciences

Index Terms
Computer Science  Pattern Recognition

Keywords
Handwritten Document  Segmentation  Bounding Box  Threshold  Clustering