2009

Text line and word segmentation of handwritten documents

Louloudis, Georgios

Elsevier

http://hdl.handle.net/11728/6429

Downloaded from HEPHAESTUS Repository, Neapolis University institutional repository
Title: Text line and word segmentation of handwritten documents

Year: 2009

Author: G. Louloudis, B. Gatos, I. Pratikakis, C. Halatsis

Abstract: In this paper, we present a segmentation methodology of handwritten documents in their distinct entities, namely, text lines and words. Text line segmentation is achieved by applying Hough transform on a subset of the document image connected components. A post-processing step includes the correction of possible false alarms, the detection of text lines that Hough transform failed to create and finally the efficient separation of vertically connected characters using a novel method based on skeletonization. Word segmentation is addressed as a two class problem. The distances between adjacent overlapped components in a text line are calculated using the combination of two distance metrics and each of them is categorized either as an inter- or an intra-word distance in a Gaussian mixture modeling framework. The performance of the proposed methodology is based on a consistent and concrete evaluation methodology that uses suitable performance measures in order to compare the text line segmentation and word segmentation results against the corresponding ground truth annotation. The efficiency of the proposed methodology is demonstrated by experimentation conducted on two different datasets: (a) on the test set of the ICDAR2007 handwriting segmentation competition and (b) on a set of historical handwritten documents.