Abstract

Digital documents are easy to handle, share and store than hard copy of documents. These made people to prefer digital document over hard copy of documents. Digital documents are nothing but scanned images of a document or natural images of notice boards, traffic signs. Text detection is an important process required to extract text from images. Text from images can be extracted using Optical Character Recognition (OCR). OCR works in three phases as pre-processing, segmentation, character recognition. Pre-processing is the first phase which uses different techniques for making text easy to extract from images. In segmentation phase, each character is isolated. Then this will be given as input to OCR recognition phase which will compare it with training data-set and will recognize character. In this survey paper, different techniques for OCR are discussed.
5. Julinda Gllavata, Ralph Ewerth and Bernd Freisleben, A Robust Algorithm for Text Detection in Images, University of Marburg, D-35032 Marburg, Germany.
11. Adam Coates, Blake Carpenter, Carl Case, Sanjeev Satheesh, Bipin Suresh, Tao Wang, David J. Wu, Andrew Y. Ng, Text Detection and Character Recognition in Scene Images with Unsupervised Feature Learning, ICDAR.

Index Terms

Computer Science
Pattern Recognition

Keywords
Optical Character Recognition (OCR), Pattern Recognition, Automatic Number Plate Recognition (ANPR)