A Novel Approach For Skew Estimation Of Document Images In OCR System

Sarfraz, M. Zidouri, A. Shahab, S.A.; Dept. of Inf. & Comput. Sci., KFUPM, Dhahran, Saudi Arabia;

Computer Graphics, Imaging and Vision: New Trends, 2005. International conference;Publication Date: 26-29 July 2005; ISBN: 0-7695-2392-7

King Fahd University of Petroleum & Minerals

http://www.kfupm.edu.sa

Summary

Optical character recognition (OCR) is an area which has always received special attention. OCR systems are typically built on the strategy of divide and conquer, rather than recognizing documents at one go. They utilize several stages during the course of recognition. There have been many stages in a typical OCR system, preprocessing stage in considered to be indispensable. An input image or information need to be normalized and converted into format acceptable by OCR system. OCR systems typically assume that documents were printed with a single direction of the text and that the acquisition process did not introduce a relevant skew. Practically this assumption is not very strong and printed document could be skewed at some angle with horizontal axis. In this paper, we have proposed a new technique for skew estimation of image document. In the proposed scheme, multiscale properties of an image are utilized together with principal component analysis to estimate the orientation of principal axis of clustered data.