

Documents

Al-Dobais, M.A., Alrasheed, F.A.G., Latif, G., Alzubaidi, L.

Adoptive Thresholding and Geometric Features based Physical Layout Analysis of Scanned Arabic Books

(2018) *2nd IEEE International Workshop on Arabic and Derived Script Analysis and Recognition, ASAR 2018*, art. no. 8480378, pp. 171-176. Cited 1 time.

Abstract

In the digital age, developing an automated system to convert old printed books into digital form is a challenging task. In this paper we propose a novel technique for the recognition of Arabic scanned documents both with normal and complex layouts. The proposed algorithm is based on the local adaptive thresholding and geometric features which according to the author's knowledge is the first time it is applied to Arabic document image recognition based on the Physical Layout Analysis (PLA). The proposed method was applied to dataset consisting of 90 images collected from 700 books from various publishers and contains a total of 1112 zones; text zone, image zone, and graphic zone. The proposed algorithm achieved promising results with overall average recognition of 86.71% for Text and Image block regions for all three sets. The proposed novel algorithm outperforms the techniques mentioned in previous literature. © 2018 IEEE.

2-s2.0-85056191394

Document Type: Conference Paper

Publication Stage: Final

Source: Scopus