

A method for Arabic handwritten diacritics characters

ABSTRACT

An Optical Character Recognition (OCR) is the process of converting an image representation of a document into an editable format. In addition, people have the ability to recognize characters without difficulty as reading papers or books. However, developing an OCR system that has the ability to read and recognized Arabic diacritics characters as human still, remain a problem. More, specifically, poor recognition rate in most of optical diacritics characters recognition is mainly attributed to failing in segmenting a handwritten text correctly. To overcome this problem, we perform develop a method based on seven operations; it starts with searching the text-line height followed by reading words from the line. Then identify the diacritics regions. The segmentation is also applied during this operation by converting the text-line into a grayscale and binary image. Moreover, we introduced a new model based on k-nearest neighbors (KNN) algorithm to identify diacritics and characters segmentation. KNN is trained to directly predict the diacritic from the text-line. Finally, we offer an evaluation discussion on optical diacritics characters recognition.

Keyword: Diacritics characters; Handwritten; KNN; Image recognition