Abstract

The character recognition comes under applications of image processing. The character samples are stored in a suitable image format in digital form. Each sample in the image form is properly preprocessed, segmented and the required features defining writer invariants are obtained. The word segmentation is one of the major components in document image analysis.

It provides crucial information for skew correction, zone segmentation, and character recognition. The word segmentation is an operation that seeks to decompose an image of a sequence of word into sub images of individual symbols. Its decision, that a pattern isolated from the image is that of a word, can be right or wrong. It is wrong sufficiently often to make a major contribution to the error rate of the system. In this paper we introduced Arabic word segmentation for document images are presented. We are using the bounding box regions to enclose the characters of the Arabic words and then the resulting letter spaces are progressively filled to merge the character bounding boxes to get the Arabic word bounding boxes. The proposed technique is completely avoiding the line segmentation process which normally precedes word segmentation in stuffy methods. We have tested appropriate method on documents of Arabic scripts and theirs have obtained encouraging results from proposed technique.

References

Index Terms

Computer Science  Image Processing

Keywords

Arabic Text Line Segmentation  Connected Component Labeling  Morphological Dilation
Arabic Scripts
Document Analysis
Preprocessing