# Abstract


## Radwa Gaber Fathalla


### Extraction of Arabic Text from Complex Color Images


Words have always been important carriers of information. They convey a lot of aspects about images in which they are embedded. In spite of the many approaches that have been proposed to separate text from images, very few of them have handled Arabic script. This paper presents a technique to extract Arabic words from a variety of colored images with complex backgrounds. In order to accomplish the task we have chosen the Connected Components (CC) approach. It starts with the breakdown of the RGB image into tiny homogeneous regions using the watershed transform, followed by region merging. The resulting CCs are aggregated into blocks, some of which are the candidate words. Each block is then condensed into a single vector holding the values of its features. The features generally describe the geometrical nature of the Arabic script, including a set of invariant moments. The final decision as to classify the blocks as Arabic words other was left up to a support vector machine (SVM) before passing them to an OCR software.