

# Jordan University of Science and Technology

## Automated System for Arabic Optical Character Recognition with Lookup Dictionary

**Authors:** Inad Aljarrah, Osama Al-Khaleel, Khaldoon Mhaidat,

**Abstract:** In this paper an Arabic Optical Character Recognition system is implemented. The system takes a scanned image of an Arabic text as an input and generates an editable text out of it. The system starts by segmenting the document which is presented as an image into lines, then each line is also segmented into separate words, after that each word is further segmented to sub-words. Each word or sub-word is segmented into separate characters, and then a features extraction process is applied on each character to calculate its features vector. The feature vector is then compared with templates of feature vectors for each of the Arabic alphabet with their variations. The minimum distance classifier is used in the classification stage. A recognition rate of 93.5% is attained. To improve the accuracy of the system, a lookup dictionary is employed to correct some of the misclassified characters. This resulted in improving the accuracy to 96.1%. The results achieved are promising regardless that Arabic Optical Character Recognition is considered many times harder to handle than its counterparts in other languages like English due to the continuity between the letters in the same word.