ORAN: A Basis For An Arabic OCR System

Zidouri, A.; King Fahd Univ. of Pet. & Miner., Dhahran, Saudi Arabia;

Intelligent Multimedia, Video and Speech Processing, 2004. Proceedings of 2004 International Symposium on;Publication Date: 20-22 Oct. 2004;ISBN: 0-7803-8687-6

Summary

We present a system called ORAN (offline recognition of Arabic characters and numerals). This system is based on a method called modified MCR (minimum covering run) expression for document images. Using the correspondence between binary images and bipartite graphs, the MCR expression can be found by constructing a minimum covering or maximum matching in the corresponding graph. We use the structural information obtained from this expression to describe the character strokes according to some extracted features. These are obtained after a zoning scheme, where the baseline is detected and the line of text divided into four zones. Reference prototypes for the system are built according to a structural description of characters in some model documents. By this method, we overcome the problem of segmentation that is inherent to Arabic characters, even when they are machine printed or typed. Simple matching of the candidate characters to reference prototypes is performed. A recognition rate of more than 97% is achieved.