

Off-Line Text-Independent Arabic Writer Identification Using Optimum Codebooks

Authors : Ahmed Abdullah Ahmed

Abstract : The task of recognizing the writer of a handwritten text has been an attractive research problem in the document analysis and recognition community with applications in handwriting forensics, paleography, document examination and handwriting recognition. This research presents an automatic method for writer recognition from digitized images of unconstrained writings. Although a great effort has been made by previous studies to come out with various methods, their performances, especially in terms of accuracy, are fallen short, and room for improvements is still wide open. The proposed technique employs optimal codebook based writer characterization where each writing sample is represented by a set of features computed from two codebooks, beginning and ending. Unlike most of the classical codebook based approaches which segment the writing into graphemes, this study is based on fragmenting a particular area of writing which are beginning and ending strokes. The proposed method starting with contour detection to extract significant information from the handwriting and the curve fragmentation is then employed to categorize the handwriting into Beginning and Ending zones into small fragments. The similar fragments of beginning strokes are grouped together to create Beginning cluster, and similarly, the ending strokes are grouped to create the ending cluster. These two clusters lead to the development of two codebooks (beginning and ending) by choosing the center of every similar fragments group. Writings under study are then represented by computing the probability of occurrence of codebook patterns. The probability distribution is used to characterize each writer. Two writings are then compared by computing distances between their respective probability distribution. The evaluations carried out on ICFHR standard dataset of 206 writers using Beginning and Ending codebooks separately. Finally, the Ending codebook achieved the highest identification rate of 98.23%, which is the best result so far on ICFHR dataset.

Keywords : off-line text-independent writer identification, feature extraction, codebook, fragments

Conference Title : ICPRIP 2015 : International Conference on Pattern Recognition and Image Processing

Conference Location : Sydney, Australia

Conference Dates : December 10-11, 2015