Abstract

The old manuscripts are a part of the richest cultural heritage and legacy of civilizations where the digitalization is a solution for the preservation of these manuscripts. The conception of handwriting recognition system knows today a great expansion and appears as a necessity in order to exploit the wealth of information contained in ancient manuscripts. In this paper, a holistic approach for spotting and searching query, especially, for images documents in handwritten Arabic is proposed. These operations need a lot of time and effort to do manual work. For this, we use in the first time text line segmentation of handwritten document image based on partial projection, where a sliding-window approach is used to locate the document regions that are most similar to the query. Histograms of Oriented Gradients (HOGs) are used as the feature vectors to represent the query and documents image, then Support Vector Machines (SVM) is used to produce a better representation of the query and to classify feature vectors. Finally, the application of the reclassification technique at the indexation stage, leads to better results.
References

Index Terms

Computer Science
Pattern Recognition

Keywords

Indexation, Classification, SVM, Segmentation, Arabic handwritten documents, Histograms of Oriented Gradients (HOG).