Abstract

The paper attempts to create a handwritten document retrieval system suitable for Tamil language, with a view to record traditional literature content for future reference. It projects a search mechanism to access the query word images using a statistical model based methodology. The scheme revolves around a well defined procedure which results in word
models from where the search word can be recognised and the relevant documents retrieved. The approach involves the use of hidden Markov models (HMM) to characterize the features of the dynamically varying strokes of handwritten characters. The strategy is investigated for a sample document set over a commonly used literature. The results reveal that the system yields a reasonable performance with considerable accuracy. The highlight of this procedure is that it can effectively segment differently written words from text lines in a document and imbibes in it a flexibility to cover a wide range of tilts in the strokes that are attached to the different words.

Reference


Index Terms

Computer Science
Pattern Recognition
Key words

Handwritten document retrieval
Profile features
Segmentation
Hidden Markov models