Abstract

Characterization of matrices for efficient classification has several options. There are various alternatives depending on the structure of the matrix. Different features can be adapted in different situations. Image recognition and in particular character recognition is an excellent example where large number of image matrices need to be stored and retrieved often at high speed, at the same time performing computational tasks, resulting in requirements of huge memory and computation time. Near 100% character segmentation accuracy is achieved based on a novel segmentation technique. Here feature extraction is based on the distinctive structural features of machine-printed text lines in these scripts. The final recognition is achieved through Support Vector Machine (SVM) classifiers. The proposed algorithms have been tested on a variety of printed Malayalam documents. Recognition rates between 97.72% and 98.78% have resulted.

Reference


Index Terms

Computer Science
Segmentation

Techniques
Key words

Pattern recognition

character classification

segmentation

Optical character recognition

Singular value decomposition

marginal frequency

Support vector machine classifier

Malayalam OCR