Abstract

In a country like India where different scripts are in use, automatic identification of handwritten script facilitates many important applications such as automatic transcription of multilingual documents and for the selection of script specific OCR in a multilingual environment. Existing script identification techniques depend on various features extracted from document images at character, word, text line or block level. In this paper, we propose a novel method towards multi-script identification at block level. The recognition is based upon features extracted using Discrete Cosine Transform (DCT) and Wavelets of Daubechies family.
The proposed method is experimented on handwritten documents of eight Indian scripts that include English script and yielded encouraging results.

Reference

- M. C. Padma and P. Nagbhushan, Identification and separation of text words of Kannada, Hindi, and English languages through discriminating features, Proceedings of NCDAR 2003, 252-260
Index Terms

Computer Science

Pattern Recognition

Key words

Multi-script documents

handwritten script

Discrete Cosine Transform

Wavelets

K-NN classifier