Abstract

Nowadays, it is observed that English script has interspersed within the Indian languages. So there is a need for an optical character recognition (OCR) system which can recognize these bilingual documents and store it for future use. Hence, in this paper an OCR system is proposed that can read documents containing Gujarati and English scripts (Only digits). These scripts have many features in common and hence a single system can be modelled to recognize them. Here, we have used template matching classifier. The normalized feature vector is used as a feature to classify English and Gujarati digits. The system shows a good performance for multi-font, size independent printed bilingual English- Gujarati digits. An average classification rate 98.30% is obtained for Gujarati digits and 98.88% is obtained for English digits at character
level.

References


Index Terms

Computer Science  
Computational Intelligence

Keywords

Segmentation  Normalization  Vector  Template  Correlation