Abstract

Achieving high accuracies in recognition of handwritten text is a challenging research problem and never exhausting. The factors that instill challenges in handwritten character recognition include high degree of variability in writing, script type and the type of documents etc. In this paper, we focus on recognition of handwritten Telugu text commonly found in document images. The character set include all the vowels, consonants and single level vowel consonant clusters chosen in accordance with the commonly used terminology employed in composition pre-printed documents such as admission forms. In this paper, an algorithm is devised that performs zone-based feature extraction of the segmented character images. In the proposed work, the Gabor features are extracted from the character image at zone level and its efficiency is evaluated individually on two different zone representations and entire image at various scales and orientations. The classification and recognition performance is analyzed using nearest neighbor classifier, Naïve Bayesian, multi-class SVM and probabilistic neural networks classifiers. The efficiency of the classifiers are also tested with statistical, Histogram of Gradients and Hu moments’ feature extraction methods and the best accuracy of the system is
found to be 84.8% for Gabor features with zone representation 2 and with multi-class SVM classifier.

References


32. Tan, Hin-Leong., “Hybrid feature-based and template matching optical character
analysis in Indian bilingual documents”, International Conference on Intelligent Sensing and
Recognition Using Alignment to Prototypes”, International Conference on Document Analysis
and Recognition (ICDAR), 2005.
35. Chih-Wei Hsu, Chih-Jen Lin,” A Comparison of Methods for Multi-class Support Vector
36. Sabine Barrat, Salvatore Tabbone, Patrick Nourissier ,”A Bayesian classifier for symbol
character recognition using probabilistic neural networks”, Pattern Recognition, Vol. 30, Issue 8,
38. Vincent Labatut, Hocine Cherifi, “Accuracy measures for comparison of classifiers”,
descriptors for robust character recognition”, In 2011 International Conference on Document

Index Terms

Computer Science  Pattern Recognition

Keywords

Handwritten Telugu text, Character recognition, Gabor feature, Multi-class SVM, Naïve
Bayesian classifier, nearest neighbor.