Abstract

This paper deals with a syntactic approach for character recognition using approximate string matching and chain coding of characters. Here we deal only with the classification of characters and not on other phase of the character recognition process in a Optical character Recognition. The character image is first normalized to a specified size then by boundary detection process we detect the boundary of the character image. The character now converted to boundary curve representation of the characters. Then the curve is encoded to a sequence of numbers using Freeman chain coding. The coding scheme gives a sequence of numbers ranges from 0 to 7. Now the characters are in form of strings. For training set we will get a set of strings which is stored in the trie. The extracted unclassified character is also converted to string and searched in the trie. As we are dealing with the character which can be of different orientation so the searching is done with approximate string matching to support noisy character that of different orientation. For approximate string matching we use Look Ahead Branch and Bound scheme to prune path and make the approximation accurate and efficient. As we are using trie data structure, so it take uniform time and don’t dependent on the size of the input. When we performed our experimentation for noiseless character that is
printed character it successfully recognize all characters. But when we tested with the different variation of the character then it detect most of the character except some noisy character.

References


Index Terms

Computer Science

Pattern Recognition
Keywords
Syntactic Pattern Recognition  Freeman Chain Coding  trie  Character Recognition

Approximate string matching

Boundary detection.