Abstract

The desire to access, search and explore large amount of documents had paved the way for digitizing and storing the document in computer for easy access. But doing a word search (word spotting) in a scanned document is difficult, since the entire document is saved as an image. Different methods are already proposed which recognizes the characters from the scanned document and converts it into a text document in which the word spotting is done. In proposed method, a set of features are extracted from each character and the feature vector is converted to a floating point value. This floating point is a combination of quadrant densities obtained from the character and its aspect ratio with their respective importance. Now using this floating point recognizing the character with reference to a trained set of floating point values can be done. Now when the user searches for a certain word, spatial adjacency algorithm is used to spot the searched keyword directly in the image. Character recognition is one of the most dynamic part of today's artificial intelligence systems. Here the proposing system analysis the similarity between various characters in a language and trains itself to identify and understand similar characters using the previously learned data.
- Stefan Klink, German Research centre for the Artificial Intelligence (DFKI, GmbH), Germany
- Andreas Dengel, Thomas Kieninger German Research center for the Artificial Intelligence (DFKI, GmbH), Germany
- Sami Lais is a freelance writer in Takoma Park, Md.
- T. M. Rath, R. Manmatha et al. [trath,manmatha] @ cs. umass. edu

Index Terms

Computer Science
Artificial Intelligence

Keywords