Abstract

Recognition techniques for printed and handwritten text in scanned documents are significantly different. In this paper, we propose method to automatically identify the signature in the scanned document images. This helps to retrieve the document images based on the signature. A simple region growing algorithm is used to segment the document into a number of patches. A patch is composed of many closely located components. A component is a one piece of connected foreground pixels (say 8 connectivity). We extracted the state features of all the patches to identify the signature in the document images. A label for each such segmented
 Discrimination between Printed and Handwritten Text in Documents

patch is inferred using neural network model (NN) and support vector machine (SVM). These models are flexible enough to include signature as a type of handwriting and isolate it from machine-print. From experimental results we found that classification rate for SVM is superior over NN.

Reference

- Christopher M. Bishop Pattern Recognition and Machine Learning

Index Terms

Computer Science

Pattern Recognition

Key words

Document analysis

text identification

machine vision

signature detection
retrieval