Abstract

The field of digital-image processing has experienced dramatic growth and increasingly widespread applicability in recent years. Fortunately, advances in computer technology have kept pace with the rapid growth in volume of image data in these and other applications. Digital-image processing has become economical in many fields of research and in industrial and military applications. While each application has requirements unique from the others, all are concerned with faster, cheaper, more accurate, and more extensive computation.

Analysis of document images for information extraction has become very prominent in recent
past. Wide variety of information, which has been conventionally stored on paper, is now being converted into electronic form for better storage and intelligent processing. This needs processing of documents using image analysis, processing methods. This article provides an overview of various methods used for digital image processing using three main components: Pre-processing, Feature extraction and the Classification. Pre-processing includes Image acquisition, Binarization, identification, Layout analysis, feature extraction and classification. Classification is an important step in Office Automation, Digital Libraries, and other document image analysis applications. This article examines the various methods used for document image processing in order to achieve a processed document having high quality, accuracy and fast retrieval.

**Reference**

- RANGACHAR KASTURI1, LAWRENCE, and O’GORMAN2, Document image analysis: A primer, S”adhan”a Vol. 27, Part 1, February 2002, pp. 3–22.

Index Terms

Computer Science  
Signal Processing

Key words

Document  
Analysis  
Processing

Classification