A SURVEY ON PRACTICES OF DOCUMENT IMAGE SPLITTING

Mrs. Rasika G. Khade
Department of Computer Engineering
Sardar Vallabhbhai National Institute of Technology, Surat, India

Dr. Krupa N. Jariwala
Department of Computer Engineering
Sardar Vallabhbhai National Institute of Technology, Surat, India

Traditionally, papery documents were used for communication and data storage. Moreover, a large number of old documents and books around the world are kept in the files and which are threatened vanishing. Thus, it is essential to preserve this heritage and make it available for everyone in easily understandable form. So, the management of paper documents must be done in well-organized and integrated way. The ultimate key would be a computer, which deals with a paper document as effectively as it able to do with other digital media. Thus to involve computers for processing of paper documents it becomes necessary to convert hardcopy of documents into softcopy. This sort of conversation can be done through the scanners or cameras, which stores the paper documents in the form of document image. The accessibility is one of the major concerns of user. The OCR (Optical Character Recognition) recognizes characters in a document, but OCR process only textual part of document image and non-text components are skipped. There are several applications (such as, scanned copy of technical book or journal may contain diagrams or tables, official documents or application forms may contain some special symbol or organization logo) where each part of document is equally important. Document layout analysis (DLA) can be used for such task. DLA is nothing but the task of splitting up document images into various different sections such as scripts, pictures, charts, logos, symbols and tables. Though, it is complex problem because of the diversity of document structure. The major focus of this paper is on the detailed study of common approaches, features used for splitting of document image into separate parts and to give future research direction for researchers.