Bank Form Classification using Document Layout Analysis and Image Processing Techniques

Shashikala H K

Department of Computer Science and Engineering
School of Engineering and Technology, Jain University
Bangalore Rural, India
shashi.hk85@gmail.com

Abstract- **Every day thousands of forms are filled out and submitted across the world, in banks, post offices, government organizations, educational institutions etc. These include electronic forms as well as physical forms. All of these forms irrespective of their origin are at some stage made digital and stored electronically to address issues of physical storage, form degradation and data accessibility. Document layout analysis is a basic step in converting document images into electronic form. This conversion is laborious and can be made more efficient (in terms of throughput and human resource) by automating most of the conversion process using document layout analysis techniques. Document classification is an important step in Office Automation, Digital Libraries, and other document image analysis applications. Physical forms require human supervision for any operations done on the form. Digitization of these forms reduces human resources, also reduces any human redundancy involved with the operation on the physical forms. This paper addresses the initial stage of this automation, namely, bank form classification and decipherment of fields. The former recognizes the type of the bank form and the latter extracts regions of useful data from the classified bank form. The proposed work aims to provide accurate bank form classification along with noise removal, skew detection and correction, finally layout analysis is carried out to extract fields like name, address, signature from the classified forms.**

Keywords—Document processing, Document classification, Classification algorithms. Document layout Analysis, Skew detection and correction, Noise removal, Bank forms, optical character recognition.