Abstract

In order to submit expense reports, multiple receipts are often scanned on a single page and the scanned images are submitted along with the expense report in order to get expenses reimbursed. These scanned images are manually verified to check the validity of the claimed expenses. In this paper, a method is presented to isolate receipt segments in an image and use Optical Character Recognition (OCR) to identify receipt amounts, reducing validation time and effort. Scanned images are processed to find the contours of all high-contrast objects in receipts, including letters. Minimum bounding rectangles (MBRs) are found for each of the contours. Spectral clustering is used to group these MBRs in order to find receipt clusters which correspond to individual receipts. These are then processed with OCR to aid the user with validation.

References

Identification of Receipts in a Multi-receipt Image using Spectral Clustering


Index Terms
Computer Science
Image Processing

Keywords
Clustering, Pattern Recognition, Receipt Recognition, and Spectral Clustering Algorithm.