

Degraded Document Analysis and Extraction of Original Text Document: An Approach without Optical Character Recognition

Authors : L. Hamsaveni, Navya Prakash, Suresha

Abstract : Document Image Analysis recognizes text and graphics in documents acquired as images. An approach without Optical Character Recognition (OCR) for degraded document image analysis has been adopted in this paper. The technique involves document imaging methods such as Image Fusing and Speeded Up Robust Features (SURF) Detection to identify and extract the degraded regions from a set of document images to obtain an original document with complete information. In case, degraded document image captured is skewed, it has to be straightened (deskew) to perform further process. A special format of image storing known as YCbCr is used as a tool to convert the Grayscale image to RGB image format. The presented algorithm is tested on various types of degraded documents such as printed documents, handwritten documents, old script documents and handwritten image sketches in documents. The purpose of this research is to obtain an original document for a given set of degraded documents of the same source.

Keywords : grayscale image format, image fusing, RGB image format, SURF detection, YCbCr image format

Conference Title : ICVAIV 2017 : International Conference on Visual Analytics and Information Visualisation

Conference Location : London, United Kingdom

Conference Dates : January 19-20, 2017