Abstract

Binarization and Segmentation are considered to be the vital tasks in Optical Character Recognition (OCR) for document digitization. This paper discusses the applications of powerful Level set methodologies on these important tasks associated with OCR. Results acquired for these essential procedures in turn govern the accuracy of the OCR system. Conventionally Otsu method and Histogram profiling methods were used for binarization and segmentation purpose [1]. In this paper, we try to replace these different methods by a single procedure based on Level Set methodology, where segmentation and binarization for any document could be efficiently done in a single step. The main advantage of this method is that it does not require separate paragraph segmentation, line segmentation and character segmentation. We have also compared the working and performance of two algorithms based on Active Contour or
Combined Method of Level set with impact on Pre-processing for binarization of document images in Tamil Script


References


Index Terms

Computer Science
Pattern Recognition

Keywords
Level Set Model
Contour
Edge Based Model
Region Based Model
Image Segmentation
Parametric Active Contour
Non Parametric Active Contour