

Roll No.:	13104033
Title:	Image Binarization of Historical Degraded Document Images
Author(s):	Chirag
Supervisor(s):	Gupta, Sumana Venkatesh, K S
Keyword(s):	Degraded Documents Pixels F-measure MPM
Subject(s):	Image Processing

Abstract: Document image binarization in the presence of different types of degradations is a challenging task. Binarization is the preprocessing step required to separate out text region from non-text region in a document. It is an important step in document analysis as it affects the performance of Optical Character Recognition (OCR). In this thesis we present a novel method of binarization of historical document images affected by typical degradations such as stain, ink-bleeding and nonuniform background. The method can handle document images affected by multiple defects. The method is based on choice of appropriate thresholding techniques. The proposed method is tested on different degraded data sets and its performance is compared with existing methods in terms of F-Measure, Misclassification Penalty Metric (MPM) and Peak Signal to Noise Ratio (PSNR).

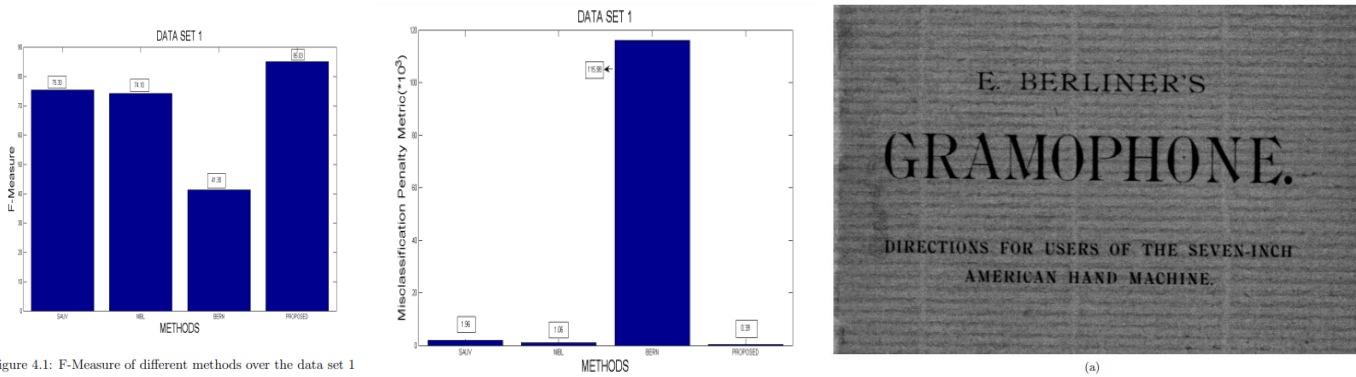


Figure 4.1: F-Measure of different methods over the data set 1



Figure 1.3: Degraded Documents affected by non-uniform background

(b) Resultant image after applying Niblack's method

(d) Resultant image after applying Sauvola's method

(e) Resultant image after applying Proposed method