IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 5, Issue 3, March 2016

Different Binarization Techniques for Improvisation of Degraded Documents

Shivagaje Pranoti Ashok[1], Prof. Mrs. A. S. Patil[2]

Student, E&TC Department, PVPIT, Bavdhan, Pune, India[1]

Asst. professor, E&TC Department, PVPIT, Pune, India[2]

Abstract: In Image binarization, segmentation is a major technique which is used for the separation of foreground and background. Image Binarization technique is used to segment the degraded images in order to gain images similar to that of the original images of the documents. One of the well-known techniques for binarization of documents is-Thresholding. This technique is further classified into Global and local thresholding techniques. There are also many techniques for Binary documentation process. Hence the work states the different techniques of binarization which are used for improving the quality of degraded documents.

Keywords: Image binarization, Segmentation, Binary documentation, Thresholding.

1. INTRODUCTION

The extraction of text stroke edge pixels from the gray-scale document images is termed as document image Binarization and it is usually performed in the document preprocessing stage.

The document Image processing tasks such as optical character recognition and document layout analyses are met using the Image binarization and hence it is an active area for research.

Thresholding is a well-known technique used for binarization of document images. Thresholding is further divided into global, local or hybrid.

In global approach threshold selection leads to a single threshold value for the entire image.Global thresholding has good performance in the case that there is a good separation between the foreground and the background. However often document images are exposed to degradations that weaken any guarantee for such a separation.

The **local thresholding technique** set different threshold for different target pixels depending on their neighborhood. Due to large variance in case of poor illuminated document or bleed through degradation we say these techniques are sensitive to background noises.

In **hybrid technique**, the approach combines global and local thresholding. In this technique the first step consists of carrying out global thresholding for classifying the part containing background and keep only the part containing foreground (text/graphics). In the second step, refinement of the image obtained in the first step is carried to obtain a sharper result.

In **dynamic Threshold Binarization** defines the threshold of a pixel with grey level values of its own and neighboring pixels and the coordinate of the pixel. This method is generally used for bad quality images viz the images with single-peak histogram.

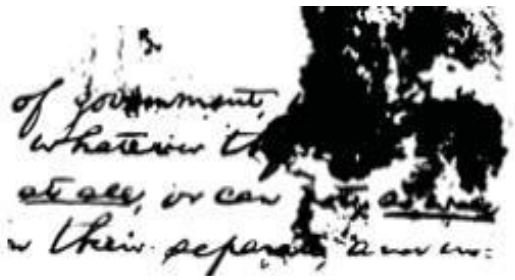


Fig 1: Input Image

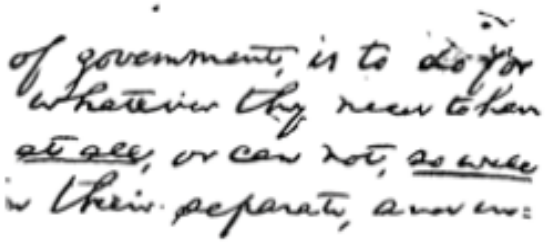


Fig 2: Binarized Image

Figure 1 has shown the input image for binarization and figure 2 has shown the binary image for the same Document image binarization is generally performed in the preprocessing phase of distinctive archive picture handling related requisitions, for example, optical characterdistinguishment (OCR) and report picture recovery.

2. DOCUMENT BINARIZATION TECHNIQUES

A. Image Binarization using Texture Features:

This technique is a versatile edge-based technique of binarization based on the texture features. The technique involves objective testing of DIBCO datasets, and subjective testing of the old corrupted documents given by the national library.

IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 5, Issue 3, March 2016

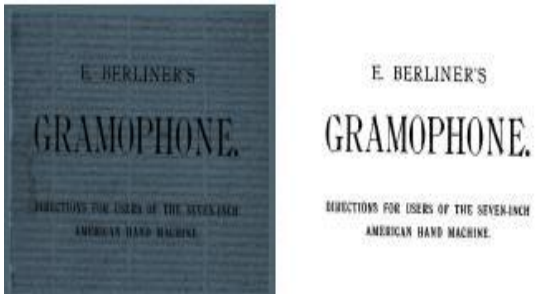
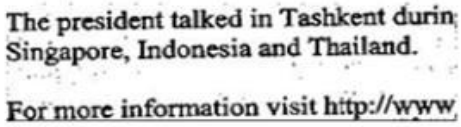


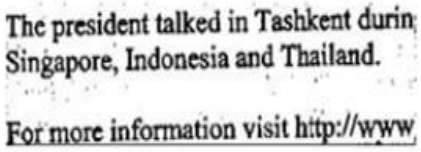
Fig 3: (a) Input (b) Output [Texture feature]

B. Adaptive Binarization for degraded images:

This method includes the utilization of dilation and erosion within the light black scale picture from which we get another picture in which shadow levels and light densities will be reduced. Binarization is planned in which we joined the system which enhanced Ni black and the neighboring threshold utilizing the little neighborhood which affected the mean values of the areas.



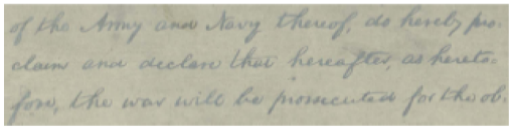
(a) Input Image



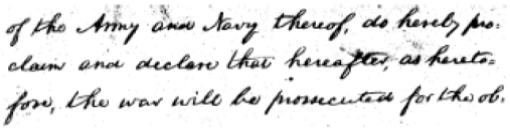
(b) Output Image

Fig 4: Adaptive thresholding

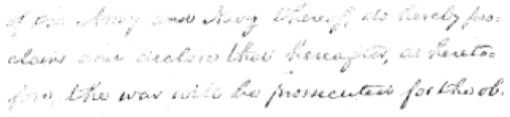
C. Combination of Document Image Binarization techniques:



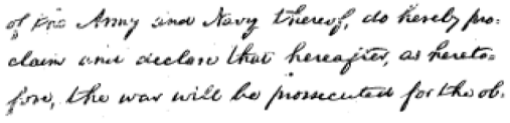
(a) One Handwritten Document Image Example



(b) Otsu's binarization result



(c) Sauvola's binarization result



(d) Combined result

Fig 5: Combination technique

This technique gives the categorization structure to join different thresholding methods and generate enhanced performance for the process of Image binarization. Considering the binarization outcome of reported methods, the document image pixels are divided into three sets: foreground pixels, background pixels, and uncertain pixel. Fig shows results of two methods separately and combinely.

D. Dynamic thresholding:

This is a Binarization technique similar to an iteration method which defines the threshold of a pixel with grey level values of its own and neighboring pixel and coordinate of the pixel. This technique is found effective on bad quality images, especially the ones with single peak histogram. Owing to the dynamic threshold calculation the method has high computation complexity and slow speed.

Consider an example for dynamic thresholding on an image and effect-

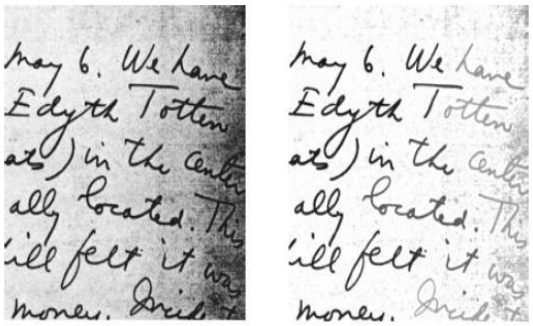


Fig 6: (a) Input Image (b) Output Image [Dynamic thresholding]

3. CONCLUSION

In this paper our main objective of evaluating the shortcomings of algorithms for degraded document images is overcomed by the application of vision processing. The key application of vision processing is Document Binarization. It is observed that each technique has its own benefits and limitations. No technique is best in every case. Enhancement can be done by finding a more reliable methodology

REFERENCES

[1] Su, Bolan, Shijian Lu, and Chew Lim Tan. "Combination of document image binarization techniques." Document Analysis and Recognition (ICDAR), 2011 International Conference on. IEEE, 2011.

[2] Su, Bolan, Shijian Lu, and Chew Lim Tan. "Robust document image binarization technique for degraded document images." Image Processing, IEEE Transactions on 22.4 (2013): 1408-1417.

[3] Pratikakis, Ioannis, Basilis Gatos, and Konstantinos Ntirogiannis."ICDAR 2013 Document Image Binarization Contest (DIBCO 2013)." Document Analysis and Recognition (ICDAR), 2013 12thInternational Conference on. IEEE, 2013.

[4] Jagroopkaur, Dr. Rajiv Mahajan, "A review of degraded document Image Binarization techniques ",Vol. 3, Issue 5, May 2014.