ISSN: 2277-9655
Scientific Journal Impact Factor: 3.449
(ISRA), Impact Factor: 1.852

IJESRT

INTERNATIONAL JOURNAL OF ENGINEERING SCIENCES & RESEARCH TECHNOLOGY

A Review on Adaptive Approach for Degraded Document Binarization

Shivani Goyal[*1], Naresh Kumar Garg[2]
[*1] Student, [2]Assistant Professor ., CSE dept., GZS_PTU Campus, Bathinda, India
er.shivanigoyal14@gmail.com

Abstract

This paper deals with an adaptive approach for binarization and enhancement of degraded documents. considerate the text from badly degraded document image is very difficult job due to high variations between the background and the foreground text of different document images. Alot of algorithms have previously been proposed for improving the thresholding of degraded document images. Not any of the algorithm can solve all types of problems, but a few of the algorithms are better than others for particular situation. There are various methodologies to improve the clarity of degraded documents such as:- Sauvola, Gato, Ni-black, Otsu etc. To remove degradations from documents like :- spoilt background, smudge, holes, gaps, irregular illumination and so on.

Keywords: binarization, degraded documents, thresholding, illumination and digital image processing.

Introduction

Binarization is the initial step of most document image analysis system and refers to the change of the gray-scale image to a binary image. Binarization is an important step in digital image processing modules because a good binarization sets the base for advanced document image analysis. Binarization generally distinguishes text areas from background.

Historical Documents Historical documents collections are a precious resource for human. There is a great collection of historical documents that have invaluable knowledge about the history and religion. Some of the people are allowed access to such collections because the maintenance of the material is of great concern.

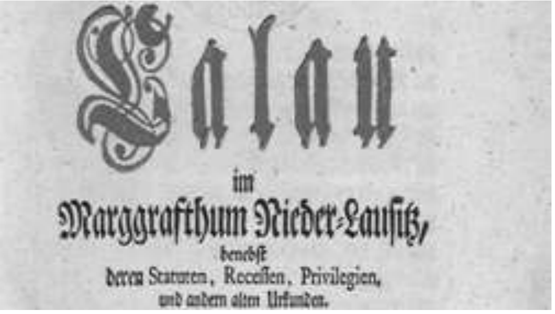

Fig. 1: Historical Document

Degraded Documents Degradations show regularly and may arise due to some reasons. Adaptive degraded document image binarization gives a consistent result. It is mostly performed by clearing up any unnecessary objects appeared in the document, hiding background, remove noise and fill gaps or holes in the foreground and ultimately recover the quality of the character.

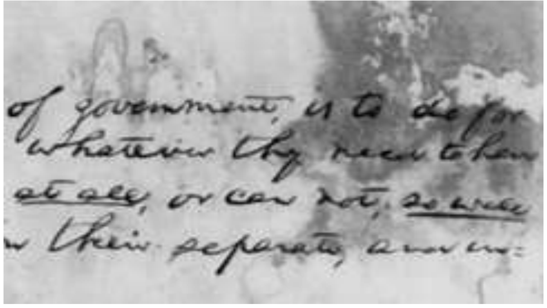

Fig. 2 : Degraded Document

Historical & Degraded document images collected from (DIBCO) dataset of images.

Existing techniques

Image binarization converts an image up to 256 gray levels to a black and white figure. Generally, binarization is used as a pre-processing step to divide the text pixels from the background pixels. The easiest way to get an image binarized is to select a threshold value, and arrange all pixels with values above this threshold as white, and all other pixels as black. Generally, approaches that deal with document image binarization are either global or local thresholding.

Global Thresholding Global binarization algorithms used a single global threshold value for entire document image. The global threshold is used to divide image pixels and background image pixels of objects.

Local Thresholding Introduce an algorithm that calculates a pixelwise threshold by moving a rectangular window across the image. Local binarization algorithms used a different local threshold value for each pixel of document image.

Historical Manuscripts by **Chun Che Fung 2010 [4]** Survey is performed on different binarization techniques such as Otsu, Sauvola, Ni-Black algorithm and different evaluation measurements have been taken This technique result in noise reduction and in selection of automatic optimal binarization algorithm.

Binarizing Degraded Document Images by **Y.H. CHIU ET. AL. 2012 [11]** This paper proposes a two-stage parameter-free window-based method to binarize the degraded document images. Empirical results demonstrate that the proposed method is competitive when compared to the existing adaptive binarization methods and achieves better performance in accuracy, precision and F-measure.

Historical Document Images by **K. NTIROGIANNIS, B.GTOS, 2009 [8]** This paper improves the adaptive logical level techniques by making the window variable to extract the essential features as charter stroke width (SW) as some character have different SW so there will be different SW values. Adaptive Logical Level Techniques and the proposed method were compared and show the f-measurement for the global parameter for handwritten and printed image has best performance at value equal to 0.2 and 0.4 for global parameter.

Conclusion

This review has concentrated on the techniques of image binarization. In this paper Basically, Study of Traditional Binarization Techniques. No single technique could be claimed as the best method. Several techniques and algorithms depend on foreground and background of an image. Our main focus is to effectively binarize the document images suffering from strain & smudge, spoilt backround, holes, spot and various illumination effect by applying Adaptive Binarization Techniques.

References

[1] B.Gatos, I. Pratikakis and S.J. Perantonis, “Adaptive Degraded Document Image Binarization”, Pattern Recognition, Vol. 39(3), PP: 317 – 327, 2006.

[2] B. Gatos, I. Pratikakis and S.J. Perantonis, “Efficient Binarization Of Historical And Degraded Document Images ’’, IEEE Transactions on Image Processing, Vol. 7, PP: 447 - 454, 2008.

[3] B. Gatos, I. Pratikakis and S.J. Perantonis, “Improve Document Image Binarization by Using a Combination of Multiple Binarization Techniques and Adapted Edge Information”, Proceedings of the 19th International Conference on Pattern Recognition, PP: 1 - 4, 2008.

[4] Fung, C.C. and Chamchong, R., “A Review of Evaluation of Optimal Binarization Technique for Character Segmentation in Historical Manuscripts”, 3rd International Conference on Knowledge Discovery and Data Mining, Phuket, IEEE, PP: 236 – 240, 2010.

[5] J. He, Q.D.M. Do, A.C. Downton, and J.H. Kim., “A Comparison of Binarization Methods for Historical Archive Documents”, In International Conference on Document Analysis and Recognition, PP: 538 – 542, 2005

[6] J.J. Sauvola, T. Seppanen, S. Haapakoski, and M. Pietikäinen, “Adaptive Document Binarization", International Conference on Document Analysis and Recognition (ICDAR), Vol. 1, PP: 147 – 152, 1997.

[7] João Marcelo Monte da Silva, Rafael Dueire Lins, “A New And Efficient Algorithm To Binarize Document Images Removing Back-To-Front Interface”, JUCS, Vol. 14(2), PP: 229 - 313, 2008.

[8] K. Ntirogiannis, B. Gatos and I. Pratikakis, “A Modified Adaptive Logical Level Binarization Techniques For Historical Document Images”, 10th International Conference on Document analysis and Recognition (ICDAR), Barcelona, Spain, IEEE, PP: 1171 – 1175, 2009.

[9] Ntogas Nikolaos, Ventzas Dimitrios, A Binarization method for Historical Manuscripts, 12th WSEAS International Conference on Comunications, Heraklion, Greece, PP: 23 – 25, 2008.

[10] Yahia S. Halabi, Zaid SA, Faris Hamdan, Khaled Haj Yousef, “Modeling Adaptive Degraded Document Image Binarization and Optical Character System”, Euro Journals Publishing, Inc., Vol.28 No.1, PP: 14 - 32, 2009.

[11] *Y.H. Chiu* et al. *"Parameter-free based two-stage method for binarizing degraded document images" in Pattern Recognition 45, Elsevier, PP: 4250–4262, 2012.*

[12] *You Yang, "OCR Oriented Binarization Method of Document Image," Image and Signal Processing, IEEE, Vol. 4, PP: 622 - 625, 2008.*