

## A Combination of Laplacian Energy, Global and Adaptive Techniques for Degraded Document Image Binarization

Md. Sujan Ali<sup>1</sup>, Mst. Jannatul Ferdous<sup>2</sup>

<sup>1,2</sup> Department of Computer Science and Engineering, Jatiya Kabi Kazi Nazrul Islam University, Trishal, Mymensingh, Bangladesh.

---

**Abstract:** Many document image binarization algorithms have previously been proposed for enhancing the performance of degraded document image binarization. This paper reviews algorithms for document image binarization. All of the algorithms have some advantages and disadvantages. To remove the drawbacks in this paper a combined approach is proposed that first combines different types of global and local thresholding techniques and then combines with the laplacian energy technique by logical OR operation. Experimentally the performance of the combined binarization technique is evaluated for both handwritten and printed degraded documents. The results showed that the performance of the proposed combined technique is better than some other techniques.

**Keywords:** image binarization, combination of binarization techniques, thresholding.

---

### I. Introduction

Binarization of document image is an important field of image processing. The purpose of the binarization is to extract text pixels from degraded gray-scale document image. It plays key role in the field of pattern recognition and document layout analysis. Thresholding is necessary technique for binarization of a gray-scale image. The gray levels of pixels belonging background are different from the gray levels of pixels belonging foreground (objects). The thresholding techniques are used to separate the objects from the background. Thresholding are usually two types: global thresholding and local or adaptive thresholding. The pixels of the image are classified into text or background according to a global threshold. Usually, such methods are simple and fast. On the other hand, they cannot be easily adapted in case the background noise is unevenly distributed in the entire image (e.g., smear or strains). The global methods are generally not perfect for degraded document images, because it selects only single threshold value for the entire image that's why in case of shadows and non-uniform illumination this method can not clearly separates foreground from background. In local thresholding the threshold value is selected considering the different region of an image. Usually averaging or median filter is used by preselected window size of the image. Local thresholding is called adaptive thresholding when a different threshold is used for different regions in the image. The disadvantage of the adaptive thresholding is that its performance totally depends on the window size and character stroke width.

Both global and local binarization techniques have drawbacks; combination of different existing image thresholding methods minimizes the drawbacks and produce better thresholding result. For a particular degraded document image, different document binarization techniques may generate different corresponding binarization results. Some techniques perform better on certain kinds of text image, while other techniques produce better binarized image for other kinds of degraded document image. Therefore, enhanced performance can be achieved by combining different kinds of binarization methods [6], [7], [8], [9]. Energy based segmentation is one of the better techniques for document image binarization [10]. This method integrates Laplacian operator to assess foreground and background labels, Canny edge detection technique to identify discontinuities and graph cut method to find minimum energy solution. The local information of an image is not considered carefully by this method during binarization.

At first this paper combines different existing thresholding techniques Otsu [1], Kittler et. al [2] and Niblack [3] and then combines the combined binarized image with the output binary image of Laplacian energy technique by logical OR operation. For experiment a set of handwritten and printed degraded text documents with corresponding ground truth images are taken from Document Image Binarization Contest (DIBCO-2009, DIBCO-2011). The experimental results showed that the proposed technique produce better binarized image than some other methods in F-measure and Peak Signal to Noise Ratio (PSNR) perspective.

### II. Survey of image binarization algorithms

Otsu [1] suggested the most popular global binarization algorithm using histogram analysis in 1979. To establish global threshold the algorithm minimizes the weighted sum of within-class variance of the foreground

and background pixels. The global threshold is selected automatically by a discriminant criterion. The pixels above the selected threshold values are set as background and lower than the threshold is set as foreground.

In Kittler et. al [2] method the foreground and background class conditional probability density functions are assumed to be Gaussian. The algorithm designed for a discriminant object from background in grayscale image. They considered the foreground and background probability distribution function are normal distributions. The algorithm minimizes the probability of classification error as well as solves a minimum error Gaussian density-fitting problem.

In Niblack [3] method the threshold value varies over the image, based on the local mean and local standard deviation. The threshold at pixel  $(x, y)$  is calculated as

$$TH(x,y) = m(x,y) + k \cdot s(x,y)$$

where  $m(x, y)$  and  $s(x, y)$  are the sample mean and standard deviation values, respectively, in a local neighborhood of  $(x, y)$ . To preserve local information in an image the size of the neighborhood should be small enough, but at the same time large enough to remove noise. The value of  $k$  is used to adjust how much of total print object boundary is taken as a part of the given object.

O'Gorman [4] proposed a global binarization method in which threshold value is calculated from a measure of local connectivity information. The thresholds are found at the intensity levels aiming to preserve the connectivity of regions.

Liu et al. [5] suggested a method for degraded text document image binarization that focuses on noisy and complex background problems. In this approach Gray-scale and run-length histogram analysis is used in object attribute thresholding. A set of Global thresholding value identifies which is used for final threshold selection utilizing local features.

B. Gatos et. al. [6], [7] proposed document image binarization approach based on combination of existing binarization techniques, incorporation of edge information and enhancement of image quality. Although edge property of an image is very important most of the image binarization methods do not consider the edge. The proposed method mainly concentrated on the edge of input document image.

Bolan Su et. al [8] presented a combination methods of document image binarization. From the document image the background, foreground and uncertain pixels are separated by the proposed framework efficiently. Depending on the pre-selected foreground and background sets the uncertain pixels are defined as either foreground or background pixels. In this paper a classifier is used iteratively to divide the three sets of pixels from the degraded document image.

E. Badekas et. al [9] proposed a binarization technique in which a Kohonen self-organizing neural network is used for the combination of results of different binarization algorithms. In this paper at first measures the best parameter values for the selected binarization techniques. The binarization results are produced using the estimated parameter values and the results are then send to the neural network to get the binarization information of the selected each techniques.

Nicholas R. Howe [10] proposed a new algorithm for degraded text document image binarization building upon energy based segmentation. First of all it uses Laplacian operator that separates concentration of brightness and darkness and hence the operator assess the local likelihood of background and foreground levels. In this algorithm a graph cut (maximum flow) method is used to efficiently compute the optimal binarization [11]. Finally, Canny edge detection technique is used so that discontinuities in the output binarization coincide with detected edges.

### **III. Combined approach (proposed)**

The combined method consists of i) combination of binarization results of some existing global and local techniques ii) document image binarization by Laplacian energy and iii) combination of the output results of i) and ii).

#### **3.1 Combination of different binarization results**

Suppose  $I_{IN(source)}(x, y)$  is input source image and  $R_1(x,y), R_2(x,y), R_3(x,y), \dots, R_N(x,y)$  represent the results of  $N$  different existing thresholding binarization methods of the source image  $I_{IN(source)}(x, y)$  where  $N=2m+1$  (odd number). After combining the all  $N$  different binarization results the binary image  $I_{B(combined)}(x, y)$  is generated according to the following formula [7]:

$$I_{B(combined)}(x, y) = \begin{cases} 1, & \text{if } \sum_{i=1}^{2m+1} R_i(x, y) > m \\ 0, & \text{otherwise} \end{cases} \quad (1)$$

**3.2 Binarization based upon Laplacian energy [10]**

Suppose  $I_B$  is a binary image of an input image  $I$  of size  $x \times y$ . If  $I_{B(DEAL)}$  is the ideal binarized image then energy function can be defined as

$$I_{B(DEAL)} = \operatorname{argmin}_{I_B} \epsilon (I_B) \tag{2}$$

A Markov random field (MRF), is a set of random variables having a Markov property described by an undirected graph. The term Markov property refers to the memoryless property of a stochastic process. A stochastic process has the Markov property if the conditional probability distribution of future states of the process depends only upon the present state, not on the sequence of events that preceded it. The Markov Random Field expresses the energy function  $\epsilon (I_B)$  as

$$\epsilon (I_B) = \sum_{i=0}^x \sum_{j=0}^y [L_{ij}^0 (1 - I_{Bij}) + L_{ij}^1 I_{Bij}] + \sum_{i=0}^{x-1} \sum_{j=0}^y C_{ij}^h (I_{Bij} \neq I_{B_{i+1,j}}) + \sum_{i=0}^x \sum_{j=0}^{y-1} C_{ij}^v (I_{Bij} \neq I_{B_{i,j+1}}) \tag{3}$$

Where,  $L_{ij}^0$  is the cost of assigning label 0 and  $L_{ij}^1$  is the cost of assigning label 1 to the pixel at (i,j).  $C_{ij}^h$  and  $C_{ij}^v$  represent the cost of a label mismatch between  $I_{Bij}$  and its neighbor to the south or east respectively. The expression of  $C_{ij}^h$  and  $C_{ij}^v$  are as follows:

$$C_{ij}^h = \begin{cases} 0 & \text{if } E_{ij} \wedge (I_{ij} < I_{i+1,j}) \\ 0 & \text{if } E_{i+1,j} \wedge (I_{ij} \geq I_{i+1,j}) \\ c & \text{otherwise} \end{cases} \tag{4}$$

$$C_{ij}^v = \begin{cases} 0 & \text{if } E_{ij} \wedge (I_{ij} < I_{i,j+1}) \\ 0 & \text{if } E_{i,j+1} \wedge (I_{ij} \geq I_{i,j+1}) \\ c & \text{otherwise} \end{cases} \tag{5}$$

Where the presence or absence of canny detected edge at pixel (i,j) is represented by  $E_{ij}$ . The costs of label identify areas of converging and diverging gradients:

$$L_{ij}^0 = \nabla^2 I_{ij} \tag{6}$$

$$L_{ij}^1 = -\nabla^2 I_{ij} \tag{7}$$

For certain bright outlier pixels,

$$L_{ij}^1 = \begin{cases} -\nabla^2 I_{ij} & H_{ij} \leq 2 S_{ij} \\ \tau & H_{ij} > 2 S_{ij} \end{cases} \tag{8}$$

Where,

$\tau$  is a constant for certain bright outlier pixel.

$$H_{ij} = I_{ij} - G_r ( I_{ij} ) \tag{9}$$

$$S_{ij} = \sqrt{G_r ( H_{ij}^2 )} \tag{10}$$

Here,  $G_r ( I_{ij} )$  is the convoluted image of  $I$  with Gaussian kernel of radius  $r$ .  $H_{ij}$  is the resemble of image  $I$  and  $S$  is the standard deviation of  $H$ . After setting best parameter values and completing post processing step the binarized image  $I_B(x, y)$  is produced.

**3.3 Combination of the output images obtained in 3.1 and 3.2**

The pixel values in a binary image represents in the form of two intensity levels that may be 0 and 1. But in the real case the binary image low intensity level is normally 0 and the high intensity level is 255. In this case, the 255 can be taken as logical 1 when applying the logical operators on the binary images. Using this convention we can carry out logical operations on images simply by applying truth-table combination rules to the pixel values from a pair of input images. Normally, corresponding pixels from each of two identically sized binary input images are compared to produce the output image, which is another binary image of the same size. In this combined approach I combine the two binary images obtained by the combination of different binarization techniques  $I_{B(combined)}(x, y)$  and by the Laplacian energy  $I_B(x, y)$  using the equation (11) that will produce better binarized output image  $I_{B(output)}(x, y)$ .

$$I_{B(output)}(x, y) = \begin{cases} 0, & \text{if } I_{B(combined)}(x, y) = I_B(x, y) = 0 \\ 1, & \text{otherwise} \end{cases} \quad (11)$$

#### IV. Experiment Results

In this paper for experiment a set of text documents both handwritten and printed has been taken from DIBCO-2009 and DIBCO-2011 for training and corresponding ground truth images for objective evaluations. The binarized images from input text documents (Fig. 1) are obtained according to Otsu [1], Kittler et. al [2] and Niblack [3] algorithms. The combination of output images from these algorithms produced an image by equation (1) shown in Fig. 2.

Fig. 3 shows the binarization output by Laplacian energy. The output images by the combination of global thresholding (Otsu [1], Kittler et. al [2]), adaptive thresholding (Niblack [3]) and Laplacian energy technique (Nicholas R. Howe [10]) are shown in Fig. 4.

F-measure, measure of binarization quality is evaluated for the output results of this paper. The F-measure is defined as

$$F\text{-measure} = \frac{2 \times \text{recall} \times \text{precision}}{\text{recall} + \text{precision}} \quad (12)$$

Where  $\text{recall} = \frac{\text{True Positive (TP)}}{\text{True positive (TP)} + \text{False Negative (FN)}}$

and

$$\text{precision} = \frac{\text{True Positive (TP)}}{\text{True positive (TP)} + \text{False Positive (FP)}}$$

Peak Signal to Noise Ratio (PSNR) is a measure of how close is an image to another. The higher the value of PSNR indicates the higher the similarity of the two images. The PSNR is defined as

$$PSNR = 10 \log \left( \frac{1}{MSE} \right) \quad (13)$$

In this experiment the F-measure and PSNR are evaluated for both handwritten and printed documents. Table 1 summarize the results of the method described in this paper. According to comparison performed, the proposed combined method has the best overall performance with F-measure and PSNR.

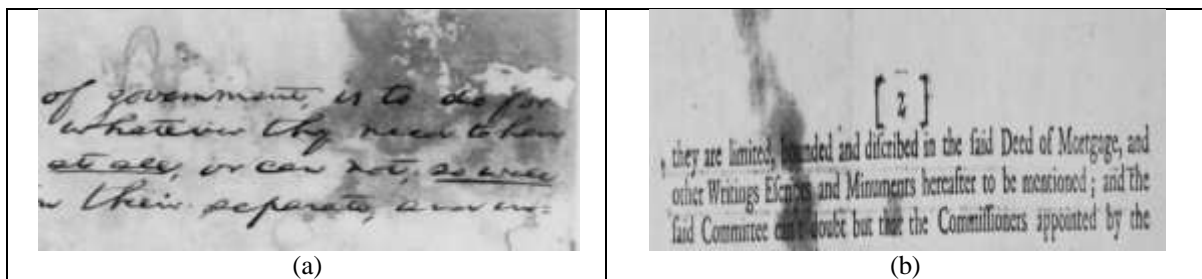


Fig. 1: Input text document images taken from DIBCO-2009: (a) handwritten document, (b) printed document.

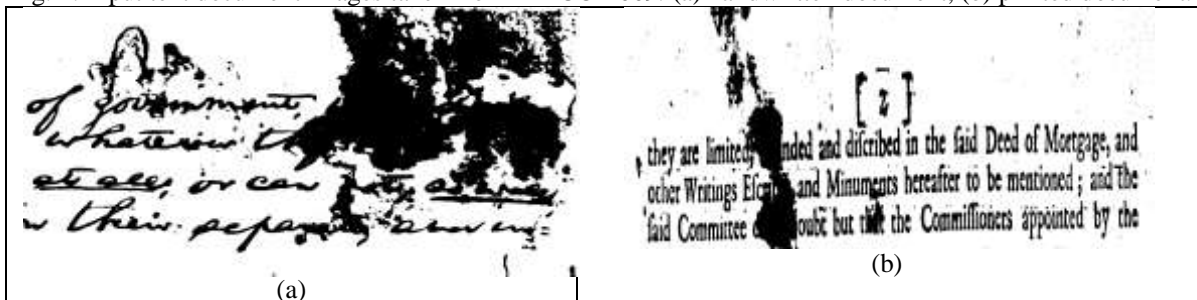


Fig. 2: Combination of Otsu [1], Kittler et. al [2] and Niblack [3] algorithms: (a) handwritten document, (b) printed document.

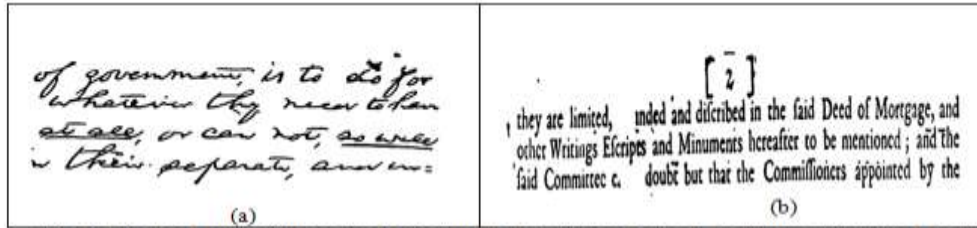


Fig. 3: Binarized output images by Laplacian energy: (a) handwritten document, (b) printed document.

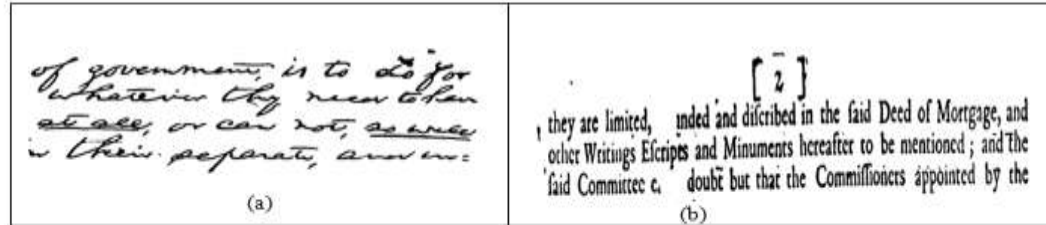


Fig. 4: Combination of different thresholding and Laplacian energy techniques: (a) handwritten document, (b) printed document.

Table 1

Document Types	Method	F-measure (%)	PSNR
Handwritten	<b>Proposed</b>	<b>92.19851</b>	<b>19.20</b>
	Nicholas R. Howe [10]	91.83638	18.97
	Niblack [3]	41.01443	6.79
	Otsu [1]	40.55702	6.73
	Kittler et. al [2]	30.00139	4.66
Printed	<b>Proposed</b>	<b>92.70648</b>	<b>18.15</b>
	Nicholas R. Howe [10]	92.11111	17.77
	Otsu [1]	82.59100	13.75
	Kittler et. al [2]	65.02197	9.49
	Niblack [3]	53.55494	7.46

## V. Conclusion

This paper presents a combined algorithm for text document image binarization after survey of some binarization techniques. The proposed combination technique mainly combines the combined result of some local and global thresholding methods with the output binarized image by Laplacian energy using logical OR operation. Experiments over the dataset of DIBCO-2009 and DIBCO-2011 demonstrate the better performance of the proposed method compare to other state-of-the-art techniques. Experimental results show that the proposed method can enhance the performance of binarization methods significantly.

## References

- [1] Otsu, N. , A threshold selection method from gray-level histograms, IEEE transaction on system, man and cybernetics SMC-8, pages 62-66.
- [2] Kittler, J., Illingworth, J., Minimum error thresholding, Pattern recognition 19(1), pages 41-47.
- [3] Niblack, W., An Introduction to Digital Image Processing, Prentice Hall, Englewood Cliffs, NJ, pages 115-116.
- [4] L. O'Gorman, Binarization and multithresholding of document images using connectivity, CVGIP: Graph. Models Image Processing 56 (6) (1994) 496-506.
- [5] Y. Liu, R. Fenrich, S.N. Srihari, An object attribute thresholding algorithm for document image binarization, International Conference on Document Analysis and Recognition, ICDAR '93, Japan, 1993, pages 278-281.
- [7] B. Gatos, I.Pratikakis, and S. J. Perantonis, Efficient banarization of historical and degraded document image, In Int. Workshop on Document Analysis and Systems, pages 447-454, 2008.
- [8] B. Gatos, I.Pratikakis, and S. J. Perantonis, Improved document image binarization by using a combination of multiple binarization techniques and adapted edge information, Int. Conference on Pattern Recognition, pages 1-4, 2008.
- [9] Bolan Su, Shijian Lu, and Chew Lim Tan, Combination of Document Image Binarization Techniques, 2011 International Conference on Document Analysis and Recognition (ICDAR), E. Badeks, N. Papamarkos, Optimal Combination of Document Binarization Techniques Using a Self-organizing Map Neural Network, Engineering Application of Artificial Intelligence, vol. 20, pages 11-24, 2007.
- [10] Nicholas R. Howe, A laplacian energy for document binarization, 2011 International Conference on Document Analysis and Recognition.
- [11] Y.Boykov and V. Kolmogorov, An Experimental comparison of Min-cut/Max-flow algorithms for energy minimization in vision, IEEE trans. On pattern analysis and machine intelligence, pages 1124-1137.