Journal of Umm Al-Qura University for Engineering and Architecture 11 (2020) 18-21

Journal of Umm Al-Qura University
for Engineering and Architecture

journal homepage: https://uqu.edu.sa/en/jea

Documents Binarization Based on Hierarchical Local Minima Thresholdx

Saleh A. Alshehri [a,*].

[a] *Department of Computer Science and Engineering, Jubail University College Royal Commission at Jubail, Jubail Industrial City, Saudi Arabia.*

ARTICLE INFO

Article History:
Submission date: 15/11/2019
Accepted date: 04/03/2020

Keywords:
Binarization, Local Minima, Document Thresholding.

ABSTRACT

Many applications, such as optical character recognition and documents archiving system, require document binarization as an initial step toward successful system. The document image to be binarized may contain impairments effects. This effect must be removed or mitigated in order to achieve good binarization performance. One important approach is to find a proper global threshold value. With much image quality degradation, this approach may not be feasible. Local thresholds can lead to better performance. The challenge is to determine the sub-images boundary where each subimage has it local threshold. This research suggests an automated method based on x and y projections to divide the original document image into several sub-images. Each sub-image would have its own impairment effect. The division is recursively and hierarchically performed until no further dividing take place or be user selection. The proposed method shows good results compared to methods presented in the literature.

1. Introduction:

Document binarization is an essential step in many applications. Optical character recognition application starts by pre-processing an image document that includes image binaraziation. An old document may not readable until it is binarized.

Document binarization is the process of converting color or grayscale images to black and white images. Color and gray image pixels values range between 0 and 255, while binary image pixels values are between 0 and 1. Converting color and gray images to binary image requires determining a threshold value such that any value greater than threshold is set to 1 and any value less than threshold value is reset to 0. Therefore, the problem becomes how to determine the threshold value [1-3]. Obviously, it may vary from one document to another. In addition, one threshold value may not be enough for one document [4].

There have been many research studies conducted for document binarization [1-9]. The focus of some researches was to build a system suitable for mobile devices since images in these devices suffer from illumination variations [1,6,8]. Removing the background pixels using means and mean variations was developed [2]. Another study used image training samples to detect text stroke width and generate the proper local threshold values [3]. Time consuming method was proposed where threshold is generated for each pixel based on image histogram [4]. Many other studies were based on Otsu's method with some modifications [5,7]. Each of these and other studies addresses the topic from different point of view. For example, enhancing the image quality as a first step is carried by Howe [10]. Another view is how to align the document to correct the skewness [11]. The main challenge in all of these studies is how to choose the window size to generate the local threshold.

Alshehri has developed a method to calculate the threshold values based on both local and global statistics [12]. The image is scaled to predefine size. Then, it is divided into four equal sub-images. The method showed acceptable results where it was ranked at the lower division of other 16 methods [12]. A problem may arise when deciding the sub-images sizes. Predefine number and sizes of these sub-images may not produce the best divisions for better threshold values calculation. The generated threshold is global. However it is based on local threshold for several windows. The choice of window size is not justifiable [12].

Projections have proven its suitability in document binarization systems. Histogram-based projection is developed. Usually, either one or multiple threshold values are calculated. This is done by dividing the document image into many sub-images [4,7]. However, it may not be closed to optimum. Therefore, the best way is to divide the document image into sub-images, each of which has its proper threshold value.

The focus in this research is to overcome the shortcomings in the previous research conducted by Alshehri [12]. First, proper procedure to divide the document image into sub-images is developed in contrast to fixed size subdivision. In addition, enhanced threshold values determination is suggested which is mainly based on projections. Even though the method can be used for handwrite text, the focus is on printed text images.

2. Threshold determination procedure

Written text in any document forms high pixel values. This is with assumption that the text is in black with high image pixels code values. Noise is represented by pixels values greater than background color "white" and less than the writing pixels values. The problem is that written text pixels values are not consistent and do not always have the highest values in all over the document.

Calculation of x and y projections show the x and y locations where high density of pixels values exist. Normally, valleys in y projection show the writing baseline locations. On the other hand, x projection shows the density of the high and low pixels. The intersection between the two projections lead to locate the areas with different noise values.

The procedure is developed as follows:

1. X and y projections are calculated:
   - Sum of all values over x axis and over y axis
2. Bothe projections were smoothed using moving average.
   - Three moving average: $X_{Proj(i)} = \frac{xp_{i-1}+xp_i+xp_{i+1}}{3}$
3. Local minima and local maxima were calculated.

* **Corresponding Author**
Department of Computer Science and Engineering, Jubail University College Royal Commission at Jubail, Jubail Industrial City, Saudi Arabia.
E-mail address: shehri@ucj.edu.sa (**Saleh A. Alshehri**).

- Calculate the moving difference on all x-projection and y projection then locates the locations where the sign change

4. The x and y threshold values are determined by the mean of the local maxima.
   - $X_{threshold} = mean(X_{projection})$, Ythreshold is calculated in similar way
5. If the difference between maximum and minimum values of the local minima is greater than 0.55 (empirically determine), possible crossings may exist.
6. If local minima fall below or above the main threshold for some specific number of points, the transition locations are determined at those locations.
7. The cross between these x and y transitions determined the sub-images boundary.

The procedure is applied recursively on each sub-image until either no more sub-image division occurs or the user selects to stop.

3. Method application example

The proposed method is applied on an the image was photographed with a mobile phone [12]. Figure 1 shows the original image and its x and y projections. As it can be noticed, y projections shows two possible locations where the local minima fall below the mean of the y local maxima of the projections. These locations are indicated by the red lines in Figure 1. Crosses in x-axis are located in similar manner. The resulting sub-images divisions are shown in Figure 1 lower right part. The sub-images are numbered from 1 to 9, staring from upper left to lower right direction.

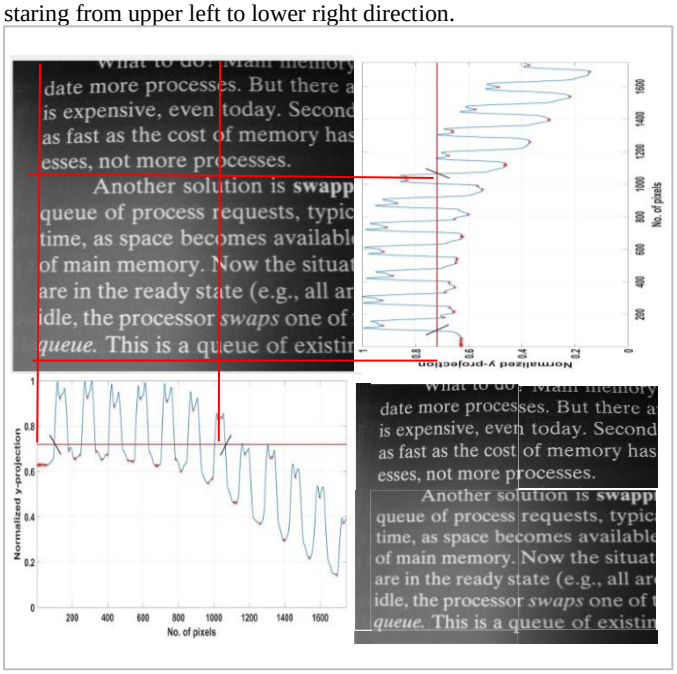
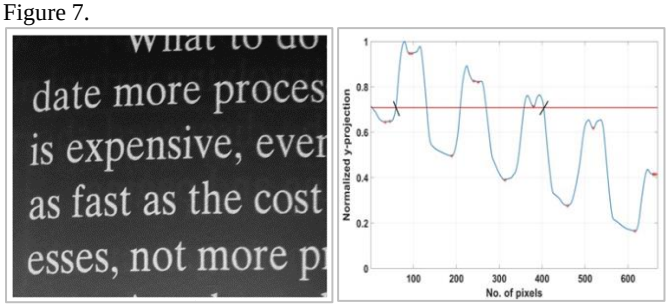
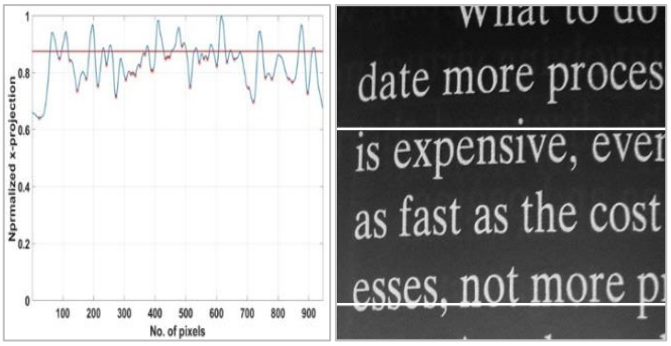


Figure 1. Original image, y-projection, x-projection and the resulting sub-images

Figure 2, Figure 3 and Figure 4 show second level sub-images divisions for sub-images 2, 3 and 6, respectively. The boundaries of each sub-imaging are based on the x and y projections sown in every figure. It can be noticed that x-projection did not produce any possible crossings. Therefore, the sub-images were further divided into sub-images vertically based on y projections. Also, both x and y projections did not produce any crossings points for sub-image 5 as shown in Figure 5. Figure 6 shows the original sub-images and its binarized sub-images. The final binarized document is shown in Figure 7.

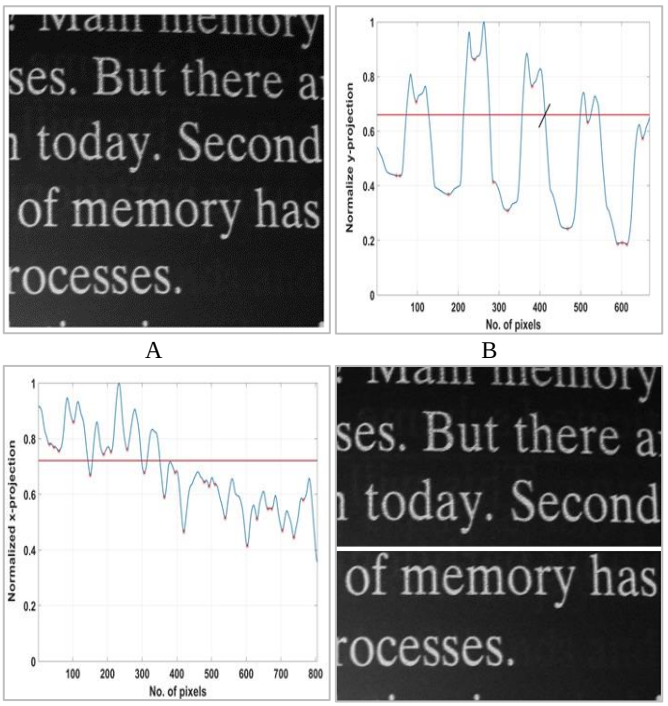


A B



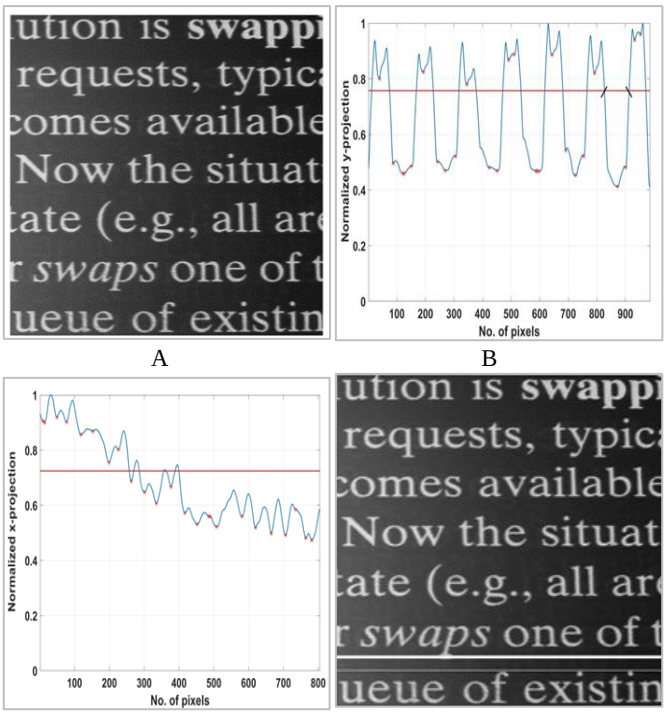
C D

Figure 2. The sub-image number 2 sub-images. (a) Original image, (b) y-projection, (c) x-projection, and (d) resulting sub-images



A B C D

Figure 3. The sub-image number 3 sub-images. (a) Original image, (b) y-projection, (c) x-projection, and (d) resulting sub-images



A B C D

Figure 4. The sub-image number 6 sub-images. (a) Original image, (b) y-projection, (c) x-projection, and (d) resulting sub-images

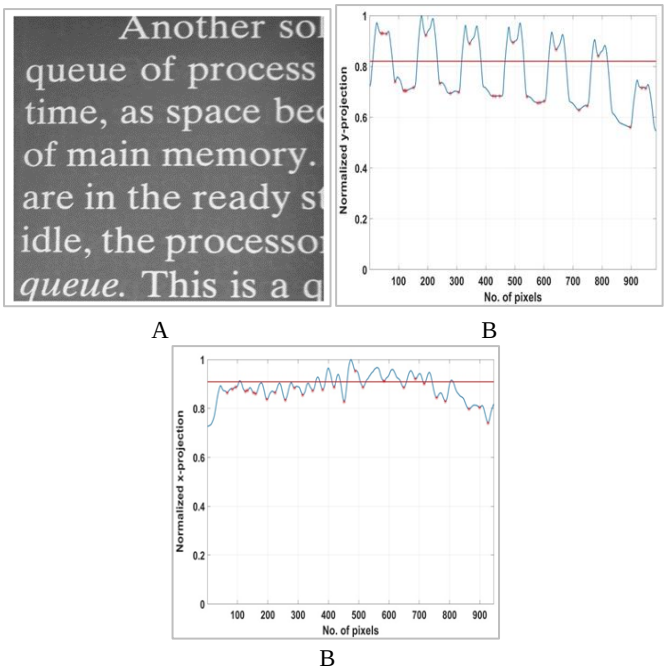


Figure 5. Subimage number 5. (a) Original image. (b) y-projection, (c) x-projection

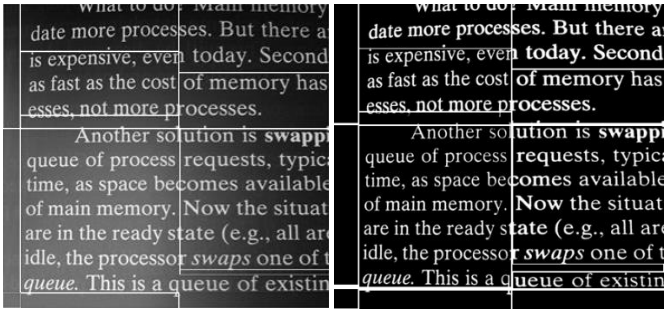


Figure 6. Original sub-images and its binarized sub-images.

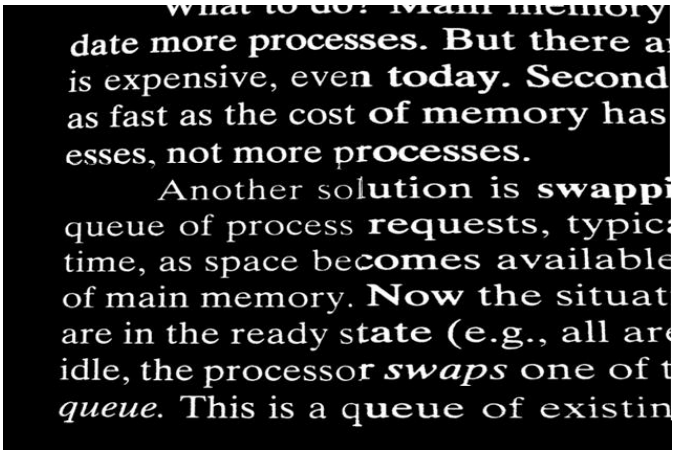


Figure 7. Final binarized document.

The proposed method is applied also on various documents. Figure 8 shows the result when applied on a historical document [9].

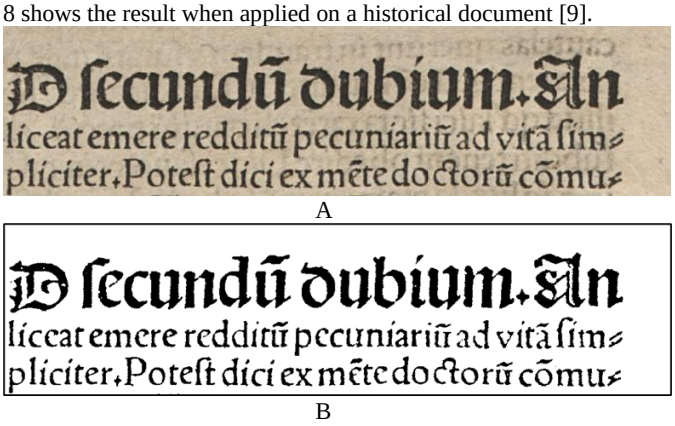


Figure 8. The result of applying the proposed method on a document from [9], (a) original document and (b) is the binarized document.

4. Results and discussion

The proposed method illustrated the possible auto-dividing selection of document threshold for Binarization application. The dividing process continued until not further need based on the x and y projection or by user selection of the subdividing levels. The proposed method was compared with the methods developed by Alshehri [12]. Alshehri's method was ranked 13 among 16 different methods. The presented method in this research could reach the fourth rank among the other methods. This used metrics were as presented in [9].

a. $$FM = \frac{2\ x\ Recall\ x\ Precision}{Recall + Precision}, \quad (1)$$

where $Recall = \frac{TP}{TP+FN}$, TP, FP and FN denote the True positive, False position and False negative values respectively.

b. $$p-FM = \frac{2\ x\ p{-}Recall\ x\ Precision}{p{-}REcall+Precision}, \quad (2)$$

where $p-Recall = \frac{\sum_{x=1,y=1}^{x=M,y=N} SG(x,y).B(x,y)}{\sum_{x=1,y=1}^{x=M,y=N} SG(x,y)}$

$SG(x,y) = \begin{cases} 0, & background \\ 1, & text \end{cases}$ and B is the MxN binary image

c. $$PSNR = 10\log\left(\frac{C^2}{MSE}\right), \quad (3)$$

where $MSE = \frac{\sum_{x=1}^{M}\sum_{y=1}^{N}(I(x,y)-I'(x,y))^2}{MN}$, c is the maximum pixel value, I and I' are the pixel intensity of both the original and processed images.

d. $$NRM = \frac{NR_{FN}+NR_{FP}}{2}, \quad (4)$$

where $NR_{FN} = \frac{N_{FN}}{N_{FN}+N_{TP}}, NR_{FP} = \frac{N_{FP}}{N_{FP}+N_{TN}}$, N_{TP} denotes true positives, N_{FP} denotes false positive, N_{TN} denotes true negative and N_{FN} denotes false negative.

e. $$MPM = \frac{MP_{FN}+MP_{FP}}{2}, \quad (5)$$

where $MP_{FN} = \frac{\sum_{i=1}^{N_{FN}} d_{FN}^{i}}{D}, MP_{FP} = \frac{\sum_{j=1}^{N_{FP}} d_{FP}^{j}}{D}$, d_{FN}^{i} and d_{FP}^{j} denote the distance of the i[th] false negative and the j[th] false positive pixel from the contour of the test in the ground truth image.

The five metric measures were obtained to be as the 4[th] entry in Table 1.

Table 1. Evaluation results including the proposed method [9]

Rank	Method	FM	p-FM	PSNR	NRM	MPM
1	1	91.50	93.58	19.78	5.98	0.49
	2	89.70	95.15	19.15	8.18	0.29
2	3	91.78	94.43	19.67	4.77	1.33
3	Proposed method	92.37	95.07	15.59	0.07	0.40
4	14	89.73	90.11	18.90	5.77	0.41
5	10	87.98	90.83	18.26	7.67	0.37
6	13	86.85	92.43	18.19	9.98	0.23
7	8	86.13	88.80	17.62	8.68	0.37
8	17	85.71	91.68	17.63	10.42	1.18
9	16	83.22	91.24	17.19	13.15	0.50
10	12	85.06	89.63	17.56	10.48	3.80
11	9	83.51	86.88	17.24	13.02	0.94
	11	82.99	87.55	17.02	12.83	0.69
12	15	81.39	81.91	15.60	5.53	1.66
13	6	84.95	86.89	16.82	11.47	48.63
14	Alshehri's Meth. [12]	77.11	78.14	15.01	0.20	3.22
15	7	82.29	89.56	16.61	13.19	2.84
16	5	73.51	78.96	15.95	19.95	1.04
16	4	57.73	66.42	14.29	28.41	1.10

The proposed method divides the document into segments based on x and y projections. The projections generate mountains and valleys as shown in Figure 1. When there is abrupt change in these mountains and valleys trend besides the number of local minima and maxima, a new document subdivision is assumed. The local threshold is then generated based on the mean of the projections values.

The F-measure is an important measure. It is a compromise between recall and precision. When both recall and precision are high, FM is high. It is clear from Table 1 that the proposed method produces the highest FM value among all other methods. For skeleton based measure, p-FM is the suggested measure. The proposed method shows the second height among other methods. For both NRM and MPM measure, the lower the better. The proposed method has the lowest NRM among all other methods. However, it did not overcome other

methods in terms of MPM and PSNR yet it was ranked 12 positions better than Alshehri's method. The various measures for Figure 8 are (FM = 86.51, p-FM = 93.05, PSNR = 12.91, NRM = 0.12 and MPM = 1.37). This result shows good general performance of the proposed method.

It is worth mentioning that Sauvola's method is based on average histogram of local window [4]. Even though it showed good results, it suffers from two problems. First, it is very time consuming method since the threshold is calculated for each pixel. Second, choosing the window size is not justifiable. The method developed by Geraud and Lazzara is based on Sauvola's method [9]. However it was ranked the sixteenth while the proposed method was ranked the third as shown in Table 1.

5. Conclusions

Image hierarchical subdividing based on x and y projection was presented to obtain candidates of image division's boundary. The local minima and maxima were necessary to be calculated. The local minima were compared with local maxima mean. A series of consecutive local minima above the mean of the local maxima show possible locations of noise. The boundary points of x-axis and y-axis form the sub-images. In each subimage, the threshold is calculated locally, and binarization is done of that subimage. The process is repeated until no more sub-images can be obtained or by the used selection. The result was very promising compared to the published results. The proposed method is undergoing investigation by the author to become a generalized method in order to be applied for handwritten.

References

[1] Oliveira, Daniel Marques and Lins , Rafael Dueire, "A New Method for Shading Removal and Binarization of Documents Acquired with Portable Digital Cameras", Proceedings of the 3rd International Workshop on Camera-Based Document Analysis and Recognition (CBDAR), 2009.

[2] Singh, T., Roy, S., Sing, O. et al. "A new local adaptive thresholding technique in binarization", International Journal of Computer Science Issues, 8(2), 271–277, 2011.

[3] Shi, J., Ray, N., Zhang, H. "Shape based local thresholding for binarization of document images", Pattern Recognition Letters, 33, 24–32, 2012.

[4] Sauvola, J., PietikaKinen, M. "Adaptive document image Binarization", Pattern Recognition, Vol. 33, pp. 225-236, 2000.

[5] Almeida, Marcos, Lins, Rafael Dueire, Bernardino , Rodrigo, Jesus, Darlisson and Bruno Lima, "A New Binarization Algorithm for Historical Documents", J. Imaging, 4 (27), pp. 1-12, 2018.

[6] Jawas, N., Wihandika, R. C., Arifin, A. Z. "Local binarization for document images captured by cameras with decision tree", Jurnal Ilmu Komputer dan Informasi, 5(1), 24–30, 2012.

[7] Messaoud, I., Amiri, H., El Abed, H., Margner, V. "New binarization approach based on text block extraction", 2011 International Conference on Document Analysis and Recognition (ICDAR), 1205–1209, Beijing, 2011.

[8] Chou, C., Lin, W., Chang, F. "A binarization method with learning-built rules for document images produced by cameras", Pattern Recognition, 43, 1518–1530, 2010.

[9] Pratikakis, I., Gatos, B., Ntirogiannis, K. "H-DIBCO 2010-Handwritten document image binarization competition", 12th International Conference on Frontiers in Handwriting Recognition (ICFHR), 727–732, Kolkata, 2010.

[10] Nicholas R. Howe "Document Binarization with Automatic Parameter Tuning", International Journal on Document Analysis and Recognition (IJDAR), September 2013, 16 (3), pp 247–258.

[11] Lins, Rafael Dueire and Avila, Bruno "A New Algorithm for Skew Detection in Images of Documents", Image Analysis and Recognition: International Conference, ICIAR 2004, Porto, Portugal, September 29-October 1, 2004, Proceedings, Part II.

[12] Alshehri, Saleh "Document Image Binarization Method That Compromises Between Global and Local Thresholding Techniques and Automates the Free Parameter Selection", 2nd Global Conference on Computer Science, Software, Networks and Engineering, (COMENG-2014), 06 - 08 November 2014, Efes Sürmeli Hotel & Convention Center – İzmir, Kuşadası, Turkey.