Textual Characters Detection in Complex Scene Images based on Bradley Thresholding Technique in Compare with Statistical Thresholding Technique

Abdel-Rahiem A. Hashem
Associate professor
Faulty of Science
Assiut University

Ahmed I. Taloba
Assisstant professor
Faulty of Computers & Information
Assiut University

Majid A. Askar
Associate professor
Faulty of Computers & Information
Assiut University

ABSTRACT
Thresholding is an important process required to most tasks in computer vision especially characters' detection from scene images. So, in this paper, we focus to enhance our previous model which is used to detect the characters from scene images and has proven superior to most existing methods. This improvement is fulfilled by replacing the thresholding method in the entire model in the stage of converting the grayscale image into the binary one. The new thresholding method is known by Bradley method. Although our previous model got better when we use Bradley thresholding, the number of false detections still less than in the case of using our previous thresholding method in the entire designed model.

Keywords
Thresholding techniques, Characters detection, Naïve Bayes classifier, Connected-components analysis, Bradley thresholding

1. INTRODUCTION
The text plays a crucial role in image understanding, in despite of there are many elements may be contained in images like persons, animals, and other objects. Text still the most important to understand image for many applications; such that summarization, retrieval and indexing images and videos. To recognize text, we need firstly to localize it in images, so the text detection or localization is the basic step before carrying the recognition step [1,2,3]. Some authors consider the localization problem is more than the recognition problem because the first one is a binary classification problem but the recognition problem is a multi-classification problem [4].

Characters detection and extraction in scene images confronted a few snags came about because of; 1) the lighting condition in the scenes may change. 2) the similarity in color between characters and the background. 3) Likewise, the characters regularly clash with different items like the basic bars, organization logo or some sprinkle [5].

Two basic classes of text are founded; the artificial text and scene text. The artificial text is the text were putted in the image or video after they are gotten, for example, comments, inscriptions and marks and so on. The Scene text is founded as an original parts of images when it is gotten by a camera, for example, road names and traffic signs [6].

There are three major classes to solve this problem. The first class is edge-based techniques which give good detections but give large number of false detections in the case of complex background. The second class is connected component-based techniques which work well when text pixels in a specific region have the same properties like color and intensity. The last class is texture-based techniques, this class is inappropriate for small size font and poor contrast [7].

The thresholding stage is considered a core stage among all stages in image processing missions especially text localization. So as the thresholding is best and obvious for an image, the detection will be better [8,9].

In this paper, we enhance our previous scheme [10] to get the characters in scene images which depend on extracting the shape properties and geometric features for each region in the image. This enhancement is carried by adopting one of the famous thresholding techniques instead of the used one in old work. Then we compare this enhanced method after we adopt the famous thresholding method with our previous thresholding technique [11] and explore the effects of the two thresholding methods on extraction characters in scene images.

The remainder of the paper organize as follow: brief overview of literature review in section 2, and in section 3, we explore some related work. In section 4, the proposed methodology is described. Experimental results are discussed in section 5. Finally, the conclusion and future work are introduced in section 6.

2. LITERATURE REVIEW
In this section, we investigate the thresholding methods that we adopt in this study, then Naïve Bayes classifier is introduced.

2.1 Thresholding methods
There are two types of the methods placed to convert the grayscale image into binary image. The first type, is global thresholding methods which compute a single value as a threshold value for the whole image. The other type, is locally adaptive thresholding. In the second type, we compute the threshold value for each pixel in image depending on the intensity values of the pixels in a neighbourhood surround this pixel. In this overview, we recall our thresholding method [11] and the new one we adopt to enhance our algorithm designed to get characters in scene images [12,13].

2.1.1 Bradley method
This method is considered a simple extension of Wellner's method [13]. In Wellener's method, the moving average is computed for the last S pixels and compare to the current pixel to classify the current pixel whether it is black or white class. If the value of current pixel is t percent lower than the moving average then this pixel is set to black, otherwise it is set to white. Weller takes t = 15 and S = 1/8th of the image width. One of the weakness of this method is that the moving average

is not a good representation of the surrounding pixels because the neighborhood samples are not distributed evenly in all directions.

In Bradley method, he uses the integral image to compute the average of neighbouring pixels on all sides. Instead of computing a running average of the last S pixels seen, he computes the average of an S x S window of pixels cantered around each pixel.

2.1.2 Our previous thresholding method

Recalling here our previous thresholding scheme to get the binary image from the grayscale image [11]. The steps of this method are summarized as follows:

Input: the RGB colour image.

Output: the binary (black/white) image.

Begin

Read input RGB image.

Convert the RGB input image into the grayscale image.

For *I* = 1 to *Number of pixels in the grayscale image*

Compute for each pixel in the grayscale image, the mean value and the standard deviation by consideration the neighborhood window with size 3x3; where the considered pixel lies in the central of this window.

Depending on the Naïve Bayes classifier, the image pixels are classified into two classes; background class and foreground with more likely to be text. (the features which the Naïve Bayes classifier depends on to fulfil the classification are the mean and the standard deviation of each pixel in the grayscale image).

End For

End

In the previous algorithm, the text pixel is defined when the value of mean statistic is maximum or closed to the maximum at this pixel and the value of standard deviation statistic is minimum or closed to the minimum at the same pixel.

In the next subsection we give short overview of the main theory in the Bayesian rule.

2.1.3 Naïve Bayes classifier (NB)

One of the most efficient and effective inductive learning algorithm in the machine learning field is Naïve Bayes (NB) [13,14]. NB is an important application of the well-known theorem; Bayesian rule in statistical induction. Given a class label, variable y

and independent features vector x1 through xn; Bayes' theorem is given from the following relationship:

$$P(y|x_1;x_2;\ldots;x_n) = \frac{P(y)p(x_1;x_2;\ldots;x_n|y)}{p(x_1;x_2;\ldots;x_n)}$$

Where the P(y│x_1;x_2;…;x_n) is the posterior probability value of the class variable y given an independent features vector, P(y) is the prior probability value of the class variable y, p(x_1;x_2;…;x_n│y) is the likelihood function of the independent features vector given the class variable, and p(x_1;x_2;…;x_n) is the probability model of the independent features vector.

When the features x_1; x_2;…; x_n are independent, the previous equation can be written as follows;

$$P(y|x_1;x_2;\ldots;x_n) = \frac{P(y)\prod_{i=1}^{n} p(x_i|y)}{\prod_{i=1}^{n} p(x_i)}$$

In this work, Naïve Bayes classifier is used to classify the pixels of an image into two sets; background pixels set and foreground pixels set. We used NB classifier because of it needs a small number of training data to compute the parameters necessary for classification. Moreover, NB classifier has acquired the superiority in compare of the existing classification techniques such as artificial neural network (ANN), support vector machine (SVM), decision tree (DT) [15].

As we mentioned that the thresholding step is a core step in many strategies handle text detection and extraction, So, Sue Wu and Adnan Amin introduce their approach depending on the thresholding. They apply thresholding for an image twice; the first one, is global thresholding to define the joint regions in the image. The second thresholding is applied for each region to separate the characters from background of the extracted region in previous step. [16]

3. RELATED WORK

The design techniques containing two or more of the main categories of text localization methods is the modern orientation for many years ago. This is to exploit the benefits and advantages of each category [17].

Lee and Kankanhalli [8] introduced their research to localize characters when the scene image has complex background. In the first stage, potential character patterns are extracted by testing contrasts in grayscales between neighbouring pixel segments, then for each potential character a local threshold is selected, based on the grayscale value of the pixel situated at the boundary of the character and the background. Then in post-processing stage, the patterns are analyse to discard of non-character regions [9].

Wang and Kangas [5], had proposed their robust technique to locate the characters depending on the connected-component method, this based on colour clustering which is utilized to partitioned the colour image into homogeneous colour layers. After this they used black adjacency graph (BAG) to analysed every connected component in colour layer, and the bounding box of component is computed. Then, they used an aligning-and-merging-analysis (AMA) scheme to locate all the potential characters depending on the bounding boxes information of connected components in all colour layers. The last step is discarding the false detections using identification of characters. This method gives good results in case when characters are square like Chinese or Korean letters whereas gives bad results when the letters are small and background is uneven lighting [5].

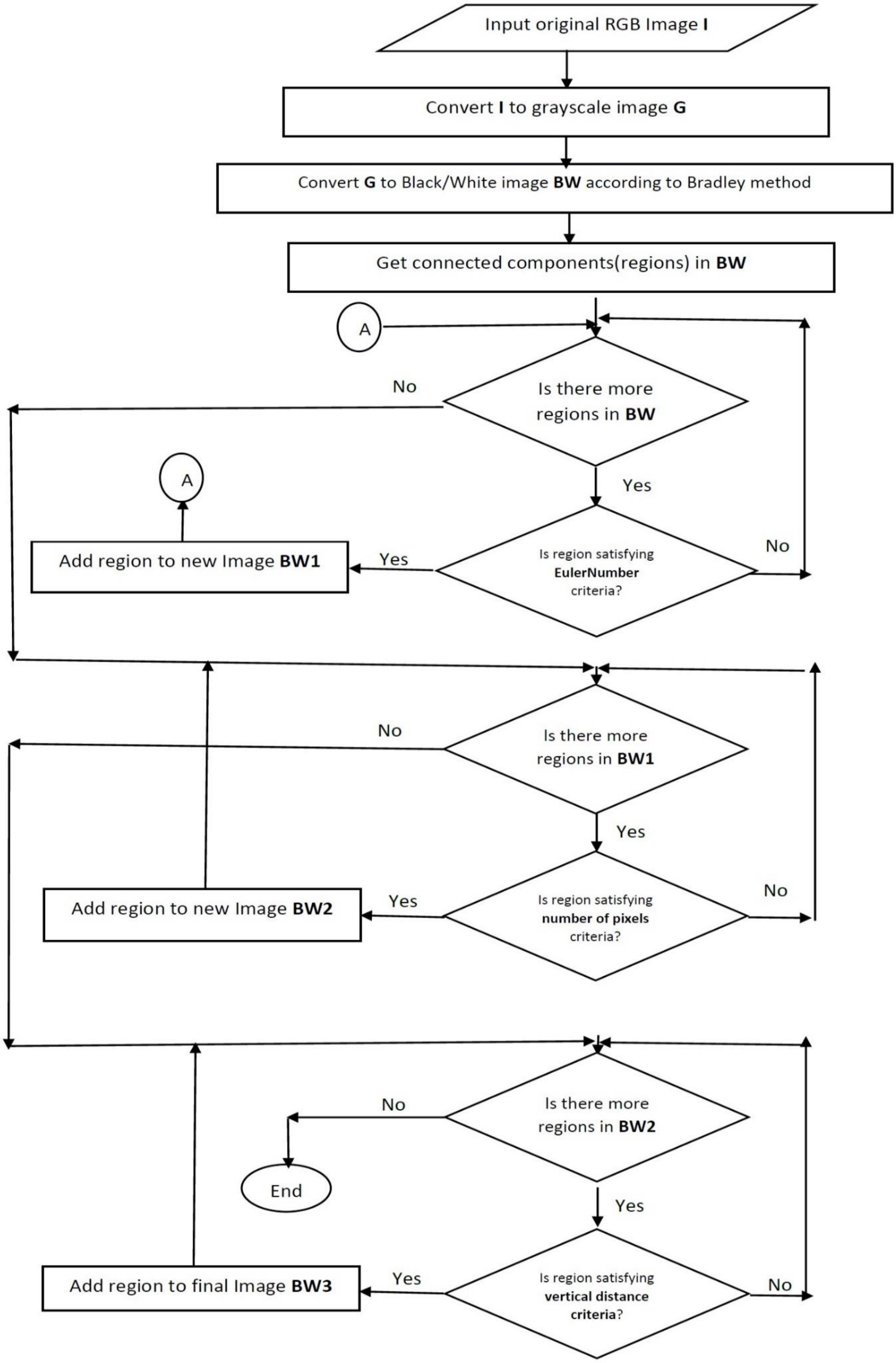

Fig 1: Flow Chart of proposed algorithm

Kita and Toru [18] introduce their technique depending on clustering algorithm to get a binary image. They arrange three stages in this method; the first stage, generation temporarily binary images using the K-means clustering algorithm. In the second stage, we need to determine whether and to what degree each temporarily binary image represents a character or non-character, this degree of character or non-character is called "character-likeness" and this stage is fulfilled using the support vector machines. The third stage, to select a single binary image amongst the temporarily binary images with the maximum degree of "character-likeness" as an optimal thresholding result.

Though many OCR techniques work well on documented images under constrained environment, they did not give perfect results in scene text images [8]; this is a direct result of inadmissible thresholding results of scene images. In their work [8], Shi C and at al., introduce an algorithm to detect and recognize the textual characters. In detection stage, they make use MSER to detect the probable text regions, then the tree-organized character model is applied on these regions to remove the false detections and find missing characters. One of the weaknesses of their method that it may fail when the characters have a large distortion. Qiao Y and et al., [19] proposed a technique by derived a formula for computing the required threshold which used to separate image into object and background, this formula basically to deal with the

4. METHODOLOGY

Preliminary:

Text detection/localization has been addressed as a classification problem in most of recent research, where the input image is divided into a foreground and a background. Other researchers have addressed the same problem from a segmentation perspective [4]

In this paper, we utilize a Bradly adapting thresholding module as a replacement to the classical Naive Bayes dependent thresholding module used in [20].

The thresholding approach in [11] uses Naive Bayes to divides the image into foreground and background by exploiting the mean and the standard deviation of different images regions. The method associates text pixels with high mean and low standard deviation values. Non-text pixels, on the other hand, are associated with minimum mean and maximum standard deviation.

The final vote on text and non-text regions of the input image is done based on the derived geometric properties of each region (mainly obtained through connected component analysis) [11,20]. The Figure 1 show the proposed algorithm.

The main steps of the proposed algorithms in Figure1 are:

1. Read RGB image.

2. Convert RGB image to Grayscale image, then convert the grayscale image to binary image using Bradley thresholding.

3. Apply the connected components analysis (CCs) for the binary image to identify all regions in the binary image, then we use the geometric features and the shape properties of each region.

4- Choosing text regions: based on the previous step, the maximum number of text regions is selected by thresholding the Euler Number property. Non-text regions are then discarded based on multiple discarding features including (1) Region area: smaller regions tend to be non-text areas, (2) Vertical distance: regions that lies far apart from other potential text regions tend to be non-text areas.

The preceding algorithm as the same one in our past work in [20] except the step No. 2; which target to convert the grayscale into the binary image.

The shape properties and geometric features are explained in detail in [21].

EN property of any region in an image is defined as the number of objects minus the number of holes in these objects, as shown in Figure 2. For example, the characters do not have any holes, whose EN is 1, and etc. Also, the vertical distances are used to define the regions represent characters existed in the same horizontal line or nearly in the same horizontal line by computing the vertical distance between the centers of regions, Figure 3 shows the modality of discarding the non-text regions according to the vertical distances between the candidates' regions. By the two features; the vertical distances and the number of pixels in each region we can remain the text-regions and discard the non-text regions according to some threshold for each feature.

Fig 2: Samples of Euler Numbers (EN)

Tracking most of the shape properties empirically such as Solidity, ConvexHull, Extent, Orientation, Eccentricity, number of pixels, and EulerNumber to examine which of these properties more precise than others in determining character regions. We found that EulerNumber property is the best among all properties to detect the maximum number of character regions.

Now we define the vertical distance we used, if we have two pixels in Cartesian coordinates; **P** (x1, y1) and **P** (x2, y2), the vertical distance is given from next formula, and shown in the next Figure 3:

Vertical Dist. = | y1 – y2 |.

The next Figure 3 shows that the importance of vertical distances in discarding the outlier regions that do not represent any character. The shown image contains 12 different objects, of them 6 represent characters; **ob5, ob6, ob7, ob8, ob9 and ob10.** The other objects in the image are not characters. As example, the **ob1** in the shown figure is far from all objects. So, it will be discarded because it does not represent a character.

5. EXPERIMENTAL RESULTS

We consider the well-known dataset ICDAR2013 to apply our new approach, which contains of 233 images. For each image in this dataset we count the number of exact letters in this image, and then we count manually the number of true letters (TP) extracted by proposed approach. After we get the True Positive using the proposed approach, we use the following equations to compute the False Positives (FP), False negative (FN) and True Negative (TN).

False Positives (FP) = Final detected letters by method - TP

False Negatives (FN) = exact letters - TP

True Negatives (TN) = all candidate objects – (exact letters + FP)

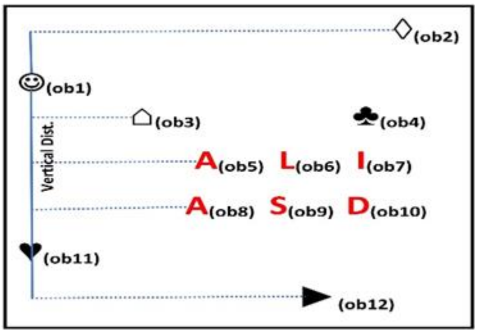

Fig 3: Samples of vertical distances between the regions

After we got the possible candidates to be text regions according to EulerNumber property, this shape property may give false regions (non-text regions). In the next step we use the two features to eliminate the non-text regions. The first feature is the number of pixels in each region and the second one is the vertical distance between each two regions for all candidate regions. For the first feature is used as a filter to discard non-text regions, we use 100 pixels as a threshold to define the text region. Secondly, we used the vertical distance between any two candidate regions and discard the region that has largest number of maximum distances to decrease the false regions.

We compute the basic metrics as follow:

Precession = TP / (TP + FP)

Accuracy = (TP + TN) / (TP + TN + FP + FN)

Recall = TP / (TP + FN)

Now, we present in the next Table 1 and Figure 4 the difference between our text localization approach with Bradley thresholding and the same text localization approach with our thresholding method which depend on the Naïve Bayes [20].

Table 1. Comparison show the effects of our thresholding method and Bradley method in detecting characters in scene images

	Precession	Recall
Based Naïve Bayes	0.34717	0.52032
Based Adapt TH.	0.22026	0.66666

From the previous Table 1 and Figure 4, we show that the proposed characters localization method after adopting the adaptive threshold (Bradley method) in the stage of converting the grayscale image into binary image outperforms the same system in case of using our thresholding method with respect to metric Recall. However, the Precession metric when we used our thresholding scheme is better than the corresponding metric in case of using Bradley method as shown in the previous Table 1 and previous Figure 4. In the next Figure 5, we present samples images which show that the superiority of using Bradley especially the images have the blurring and strong brightness.

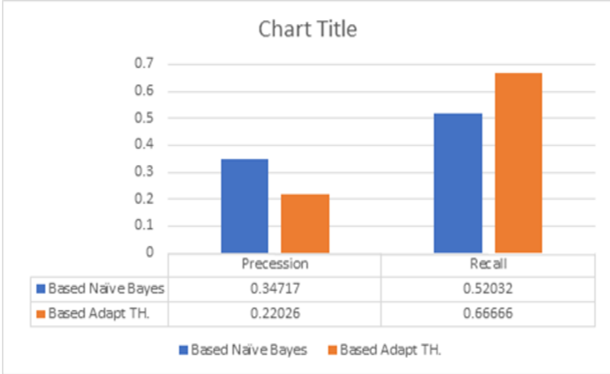

Fig 4: Comparison show the effects of our thresholding method and Bradley method in detecting characters in scene images

Original Image	Result based Bradley	Result based our thresholding scheme

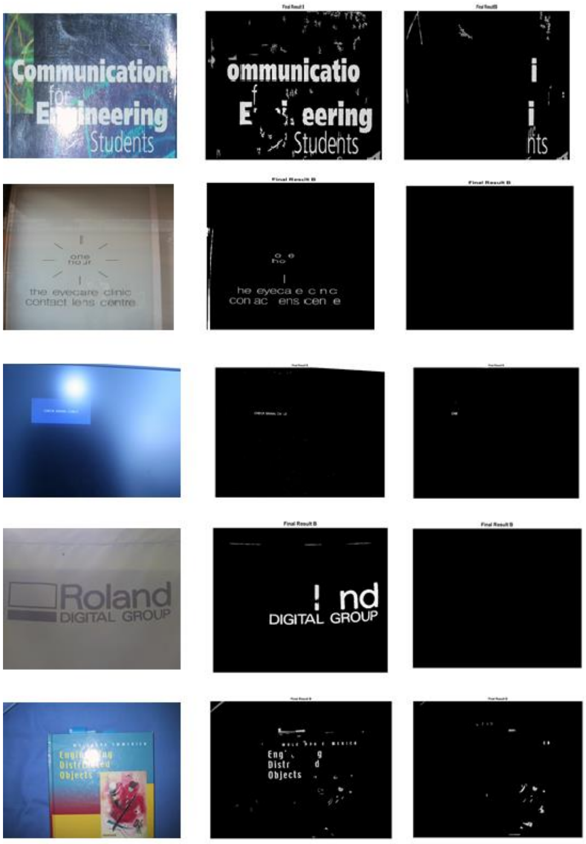

Fig 5: Samples images show that the using of adaptive thresholding is better than our thresholding method

We present in the next Figure6, samples images that reveal the superiority of our thresholding model to the Bradley thresholding model in especially in the complex images.

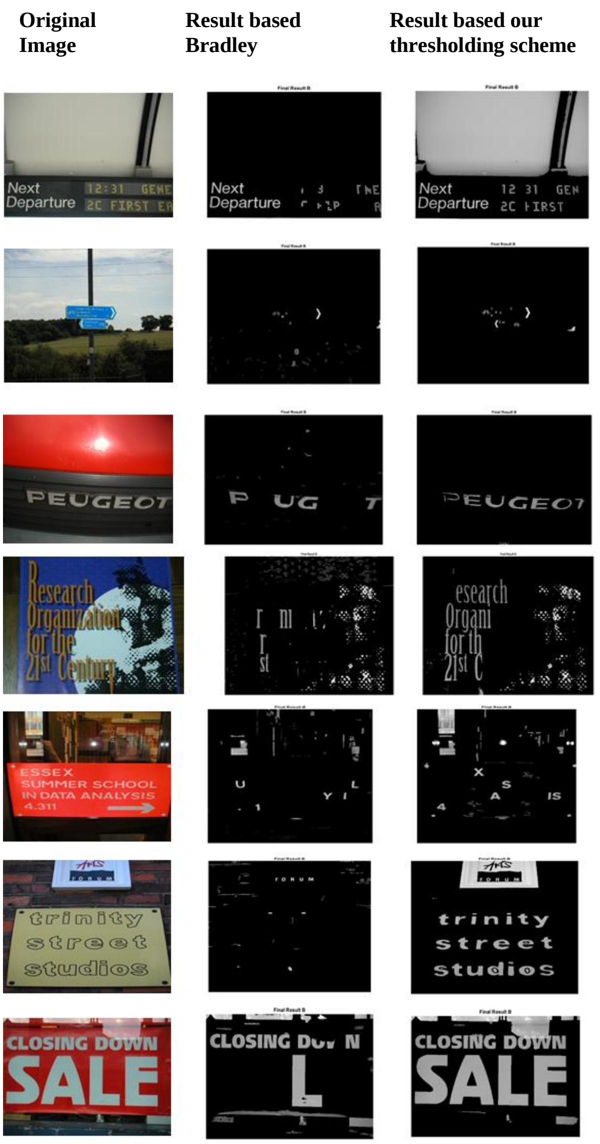

Fig 6: Samples images show that the using of our thresholding method is better than Bradley method

In the next Table 2 and Figure 7, we show comparison of our text localization approach which adopt the Bradley thresholding technique with some existing text localization techniques. We take these results from [22].

Table 2. Comparison between some recent text localization methods with our text localization technique in scene images

Method	Recall
USTB_TexStar_7_4_2013	0.6645
TextSpotter_5_4_2013	0.6484
UMD_IntegratedDisrimination_29_8_2013	0.6226
Ztext_22_7_2015	0.6351
BayesText_23_12_2013	0.6351
Text_detector_CASIA_8_4_2013	0.6285
P-SSD.v1_13_8_2017	0.6552
DetectText_18_8_2014	0.6605
VGGMaxNet_10_23_3_2015	0.5487
I2R_NUS_8_4_2013	0.6617
TH-TextLoc_8_4_2013	0.6519
CNN_Text_2_4_2017	0.6568
TEST17_17_12_2015	0.5246
Text Detection_6_4_2013	0.5342
CNN_24_3_2017	0.6491
MSER with LocalSWT_18_8_2015	0.4537
MSERs_16_9_2014	0.5178
MSER_18_8_2015	0.4462
Baseline_25_4_2013	0.3474
Inkam_10_4_2013	0.3527
The Proposed Method	**0.6666**

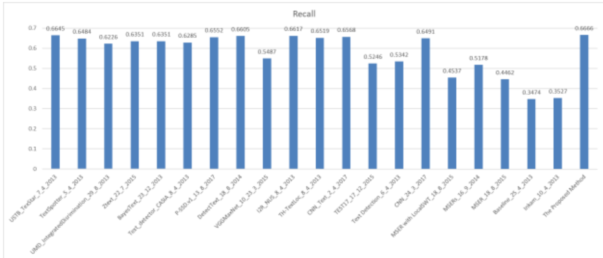

Fig 7: Comparison between some recent text localization methods with our text localization technique in scene images

In the previous Table 2 and Figure 7, we show that the highest Recall among all approaches belongs to the proposed technique in this work whose thresholding method depend on the Bradley method. This means that the maximum number of characters got from our approach.

6. DISCUSSION AND CONCLUSION

In this work, we compare our preceding thresholding method with new technique; Bradley thresholding method through study their effects in detecting characters from scene images. This is not the only target of this work. The second target to enhance our previous technique to detect characters in scene images by using this Bradley adaptive thresholding instead of our thresholding method. We built the entire system from two main stages. The first one, is converting the grayscale image into binary image, this stage is fulfilled by two types of thresholding; our thresholding technique and Bradley method. The second stage, is using the shape properties and some simple geometric features to catch the characger regions and discarding the non-character regions.

Although the simplicity of the two elementary statistics; the mean and the standard deviation, the study reveals the significance of these elementary statistics; the mean and the standard deviation which we make use as attributes that the Naïve Bayes depend on to classify the pixels of the image according to these statistics attributes. Also, it outperforms the Bradley method with respect to Precession metric, this means that the number of false detections in the case of using Bradley thresholding are more than the number of false detections in case of using the thresholding which depending on the Naïve Bayes.

As a future work, we try to design the metric to measure the regularity of the shapes we extract from the connected components analysis to reduce the false detected regions. Also, we don't use the gradient of the image in our classifier in this work, so we hope to make use this important feature in the next work. Also, we try to make use the horizontal beside of the vertical distances between candidate regions, this will be more effective in defining text regions and discarding non-text regions. Also, the need to place any heuristic model to adapt the used threshold of number of pixels is required in the future work.

7. ACKNOWLEDGMENTS

Many thanks for Dr. Ahmed Gawish, Information Tecnology Dept at Assiut University for his support and help.

8. REFERENCES

[1] Chunna T and et al. 2017 Natural scene text detection with MC–MR candidate extraction and coarse-to-fine filtering, Neurocomputing, 260 (112 – 122).

[2] Zhang H. and er al. 2013 Text extraction from natural scene image: a survey, Neurocomputing 122 (310 - 323).

[3] Yin X.C. and et al. 2016 Text detection, tracking and recognition in video: a comprehensive survey, IEEE Trans. Image Process., 25(6).

[4] Matias V. Toro and et al. 2016 Histogram of Stroke Widths for Multi-script Text Detection and Verification in Road Scenes, IFAC (International Federation of Automatic Control) Hosting by Elsevier Ltd.

[5] Kongqiao Wang and Jari A. Kangas. 2003 Character location in scene images from digital camera, Pattern Recognition, 36 (2287 – 2299).

[6] Qixiang Ye and David Doermann. 2015 Text Detection and Recognition in Imagery: A Survey, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 37, no. 7, pp. 1480-1500.

[7] Shivakumara P. and et al. 2008 Efficient Video Text Detection using Edge Features, IEEE 19th International Conference on Pattern Recognition.

[8] Shi C. and et al. 2014 End-to-end scene text recognition using tree-structured models, Pattern Recognition, vol. 47.

[9] Lee C. and Kankanhalli A. 1994 Automatic Extraction of Characters in Complex Scene Images, International Joiurnal of Pattern Recognition and Artifial Intelligence, vol. 9, No. 1.

[10] Abdel-Rahiem A. Hashem and et al. 2017 Extraction Characters from Scene Image based on Shape Properties and Geometric Features, International Journal of Computer Applications, Vol. 169, No. 3.

[11] Abdel-Rahiem A. Hashem and et al. 2017 A Comparison study on text detection in scene images based on connected component analysis, IJCSIS, vol. 15, no. 2, pp. 127-139.

[12] Trier D. and Jain K. 1995 Goal-Directed Evaluation of Thresholding Methods, IEEE Transaction on PAMI.

[13] Sahoo P. and et al. 1988 A Survey of Thresholding Techniques, COMPUTER VISION, GRAPHICS, AND IMAGE PROCESSING, 41(233-260).

[14] Bradley D. and Roth G. 2007 Adaptive thresholding using the integral image, Journal of Graphics Tools, Vol. 12, No. 2, PP. 13-21.

[15] Yi-Feng Pan and et al. 2009 Text Localization in Natural Scene Images based on Conditional Random Field, 10th International Conference on Document Analysis and Recognition.

[16] Shivakumara P. and et al. 2008 Efficient Video Text Detection using Edge Features, IEEE 19th International Conference on Pattern Recognition.

[17] Chunmei Liu and et al. 2005 Text Detection in Images Based on Unsupervised Classification of Edge-based Features, in Proceedings of Eight International Conference on Document Analysis and Recognition, pp. 610-614.

[18] Kita K. and Toru W. 2010 Thresholding of Color Characters in scene images Using k-means Clustering and Support Vector Machines, International Conference on Pattern Recognition.

[19] Yu Qiao and et al. 2007 Thresholding based on variance and intensity contrast, Pattern Recognition, 40, pp. 596-608.

[20] Abdel-Rahiem A. Hashem, Mohd Yamani and Abd El-Baset Ahmad. 2018 "Comparative study of different thresholding methods through their effects in characters localization in scene images", Data & Knowledge Engineering, Vol. 117.

[21] https://www.mathworks.com/help/images/ref/regionpr, ops.html

[22] http://rrc.cvc.uab.es/?ch=2&com=evaluation&task=1&f=1&e=1