Assiut Univ. J. of Mathematics and Computer Science *Printed ISSN 1687-4897*
47(2), pp58-73 (2018)

A SIMPLE NOVEL EFFECTIVE APPROACH TO DETECT CHARACTERS IN SCENE IMAGES

Abdel-Rahiem A. Hashem ***Abd El-Baset A. Ahmad***
Mathematics Dept., Faculty of Science, Assiut University
hashem_aer2@yahoo.com abdelbasetsalem@yahoo.com

Received: 19/4/2018 **Accepted:** 5/6/2018 **Available Online:** 20/12/2018

In this paper, we adopt a new local binarization method as an important stage in text detection problem in scene image. Then we compare this new binarization method with our previous proposed binarization method which depends on using the Naïve Bayes classifier to classify the image pixels into foreground and background sets. After this, we apply the connected components analysis to get the morphological properties of each connected component (CC). These properties make used in the next stage to distinguish between text region and non-text region. Our proposed binarization method outperforms the well-known adaptive threshold method with respect to the Precession metric. That is in this paper, we target to satisfy two benefits; the first benefit is to enhance our previous model which be used to extract the characters' regions in scene images. This enhancement will be shown from the maximizing the Recall metric and this enhancement is fulfilled by considering a new binarization method, Bradley method. The second benefit to show the difference between our previous binarization method and the new one, Bradley method with respect to Recall and Precession metrics.

Keywords: Binarization methods, Characters localization, Naïve Bayes classifier, Connected-components analysis.

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 [19].

Characters localization and extraction in scene images faced some obstacles resulted from; 1) the characters often conflict with other objects like the structural bars, company logo or some splash. 2) the colour0 of characters may differ slightly from the background. 3) Many font styles and size of characters may include in the same scene. 4) Also, the lighting condition in the scenes may vary [4].

There are two basic classes of text; the graphical text and scene text. The Graphical text is the text were added to the image or video after they are caught, for example, comments, inscriptions and marks and so on. The Scene text exists as an original parts of images when it is caught by a camera, for example, road names and traffic signs [5].

There are three major classes to solve this problem. The first class is edge-based techniques which give good detections but give more 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 colour and intensity. The last class is texture-based techniques, this class is unsuitable for small font and poor contrast [6].

The binarization step is considered a core step among all steps in image processing steps especially text localization. So as the binarization is best and obvious for an image, the detection will be better [7,8].

In this paper, we enhance our previous scheme [9] 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 binarization techniques instead of the used one in old work. Then we compare this enhanced method after we adopt the famous binarization method with our previous binarization technique [10] and explore the effects of the two binarization methods on extraction characters in scene images.

The rest of the paper arrange as follow: section 2 short overview of literature review, then in section 3, we discuss some related work. In section 4, we describe the proposed methodology. Experimental and results discussed in section 5. Finally, the conclusion and future work discussed in section 6.

2. LITERATURE REVIEW

In this section, we explore the binary methods that we used in this study, then we take an overview of Naïve Bayes classifier.

2.1 Binarization methods:

There are two types of the methods designed to convert the grayscale image into binary image. The first type, is global binarization methods which compute a single value as a threshold value for the entire image. The other type, is locally adaptive binarization. In the second type, we compute the threshold 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 binarization method [10] and the new one we adopt to enhance our algorithm designed to get characters in scene images [11,12].

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 neighbourhood 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 binarization method:

We adopt in this work our previous binarization scheme to get the binary image from the grayscale image. We summarize the steps of our method as the follow:

1. We read input RGB image.

2. We convert the RGB input image into the grayscale image.

3. We compute for each pixel in the grayscale image the mean value and the standard deviation by consideration the neighbourhood 3x3 window; where the considered pixel lies in the central of this window.

4. We use the Naïve Bayes classifier to classify the image pixels into two classes; foreground and background class. (the features we pass to the Naïve Bayes classifier are the mean and the standard deviation of each pixel in the grayscale image)

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

2.2 Naïve Bayes classifier (NB):

NB is considered one of the most efficient and effective inductive learning algorithms in the field of machine learning [12,13]. **NB** is a classification algorithm that is based on the application of the Bayes theorem with strong (naive) independence assumptions. Given a class variable **y**
and an independent feature vector $\mathbf{x_1}$ through $\mathbf{x_n}$; Bayes' theorem states the following relationship:

$$P(y|x_1, x_2, \dots, x_n) = \frac{P(y)P(x_1, x_2, \dots, x_n|y)}{P(x_1, x_2, \dots, x_n)}$$

In this paper, we use Naïve Bayes classifier to classify image pixels into two classes; background and foreground. We used this classifier because of it requires a small number of training data to estimate the parameters necessary for classification. Rather than achieve superiority in compare of other classification algorithms such as support vector machine (SVM), artificial neural network (ANN), decision tree (DT) [14].

As we mentioned that binarization step is a center stride in many strategies handle text extraction, Sue Wu and Adnan Amin [15] present their approach contingent upon thresholding to change the image into a binary image. Their proposed strategy depends on two phases. On the main stage, global thresholding is utilized to recognize the local/joint regions of the image (and to run connected component analysis over the image). The second stage is to perform thresholding on areas to separate parts of characters.

3. RELATED WORK

The recent research for many years ago focuses on designing methods mix two or three of the above categories. This is to make use the advantages of each method [16].

Lee and Kankanhalli [8] proposed their work to localize a character from a complex image background. In the first stage, potential character patterns are extracted by looking at contrasts in gray-levels between neighbouring pixel portions, then a local threshold is selected for each potential character, based on the gray-level of the pixel situated at the boundary of the character and the background. Then post-processing is running to dispose of noncharacters region like aspect ratio of bounding box contain the character and histogram in that bounding box [8].

Wang and Kangas [4], had proposed a robust, connected-component-based character locating 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 component-bounding box 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 works well in case when characters are square like Chinese or Korean letters but has some weakness when the letters are small and background is uneven lighting [4].

Kita and Toru [17] introduce their approach using clustering algorithm to get a binary image. This method contains 3 stages; the first stage, generation tentatively binary images via K-means clustering algorithm. The second stage, applying support vector machines to determine whether and to what degree each tentatively binary image represents a character or non-character. The third stage, selection single binary image with the maximum degree of "character-likeness" as an optimal binarization result.

Albeit numerous OCR frameworks work admirably on reported images under a controlled environment, they didn't give good results in scene text images [7]; this is a direct result of unacceptable binarization results of scene images. In their work [7], Shi C and at al plan framework to localize and recognize the characters, in localization stage, they utilize MSER to recognize candidate text areas, then apply tree-organized character models on these locales to eliminate false positives and find missing characters. Their method may fail in detecting the characters with the large deformation or distortion. Qiao Y and et al. [18] proposed a technique by determining an equation for calculating the desired threshold which used to separation image into object and background, this equation fundamentally to deal with small objects and relies on upon variance and intensity contrast between object and background.

4. METHODOLOGY

Preliminary:

Most researchers consider the text localization as a classification problem. This is because of we want to divide the image into two classes [19]; background class and the foreground class. In addition, we can consider the classification problem as a segmentation problem. The thresholding technique is one of the most important techniques used in the segmentation problem.

In this paper, we enhance our model described in Figure 1. This enhancement is carried out by replacement our previous method used to convert the grayscale image into the binary image by Bradley adapting thresholding method. Our binarization technique which state in [10] to divide an image into two classes; background and foreground depend on the Naïve Bayes classifier and use the mean and standard deviation as two features we pass to Naïve Bayes classifier. We define the text pixel when the maximum mean and minimum standard deviation occurs or the values closed to these two base values. On the other hand, we define the non-text pixels when the minimum mean with the maximum standard deviation values occurs. After this binarization step, we apply connected component analysis with shape properties and geometric features to remain text regions and discard non-text regions. We summarize the entire algorithm we used in Figure 1.

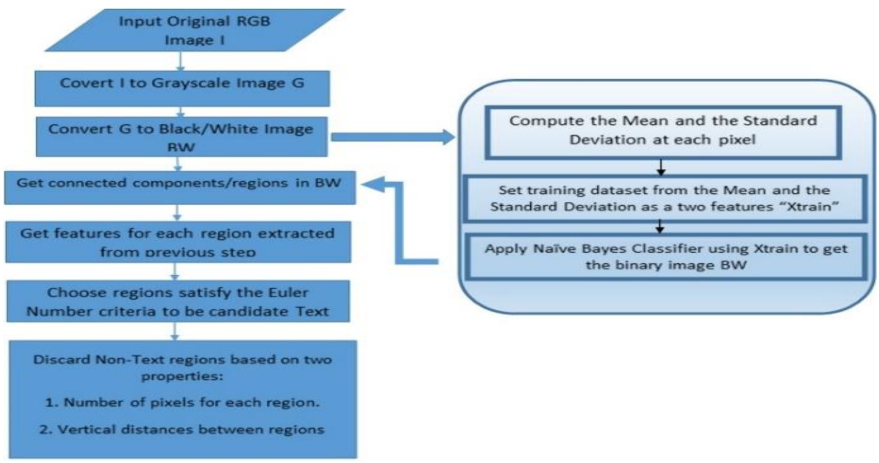

Fig.1: Flow Chart of proposed algorithm

Figure 1 show that the main two stages are: The first one is converting a grayscale image into binary image depending on the Naïve Bayes classifier as shown in the right side in Figure 1 with blurred blue colour.

The second stage with dark blue colour on the left side at Figure 1 show that we exploit the shape properties like EulerNumber and geometrical features like the center of region and vertical distance between regions.

The shape and geometric features we exploit and appear in the left side are explained in detail in [20].

EulerNumber is defined as the number of objects minus the number of holes in these objects, as shown on Figure 2. So, for some letters do not have any holes, the EulerNumber is 1, and so on. Also, we used the vertical distances to determine the character's regions existed in same horizontal line or almost in the same horizontal line by computing this vertical distance between the centers of regions. This feature with number of pixels in each region which we used to remain the text-regions and discard the non-text regions according to some threshold for each feature.

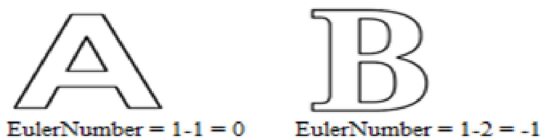

Fig. 2: Samples of Euler Numbers

When we trace most of the shape properties empirically such as ConvexHull, Eccentricity, Extent, Orientation, Solidity, number of pixels, and EulerNumber to explore which of these properties stronger than others in defining character regions. We found that EulerNumber property is the strongest to catch the maximum number of character regions among all properties.

The vertical distance is defined as follow, if we have two pixels with Cartesian coordinates; $\mathbf{P(x_1, y_1)}$ and $\mathbf{P(x_2, y_2)}$, the vertical distance computed from this formula:

Vertical Dist. $\mathbf{= | y_1 - y_2 |}$.

We can give some details for entire proposed in Figure 1 as follow:

1. Input RGB image.
2. Convert RGB image to Grayscale image, then convert the grayscale image to binary image according to Naïve Bayes classifier.
3. Apply the connected components analysis on the binary image to label the regions and use the features and the shape properties of each region extracted from connected components analysis.

4. Apply 3 features in this order; EulerNumber property to define text regions by suitable threshold, this property to select the candidate text regions. Then we apply two features to discard the non-text regions. The first features is the number of pixels in each extracted region, we discard the regions whose pixels less than **100** pixels as a threshold. This threshold of pixels need to adjust in the future work. Then we apply the second geometric features; vertical distance between each two regions, we discard the region has maximum distances between this region and other regions in the image.

Then we apply the same previous algorithm but we use in the stage of converting the grayscale image into binary image instead of our binarization method, we used the Bradley adaptive thresholding method.

5. EXPERIMENTAL RESULTS

We applied our scheme to the dataset ICDAR2013, which contain of 233 images. In each image in this dataset we compute the number of exact letters in this image, and then we compute manually the number of true letters (**TP**) extracted by proposed method, after we get the True Positive, we used 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**)

After we got the probable candidates to be regions of text according to EulerNumber, this shape feature will provide false regions (non-text regions), so in the next step we use the two elimination features; number of pixels in each region and the vertical distance between each two regions for all candidate’s regions. The first feature is used as a filter to discard non-text regions is the number of pixels for each region. We use **100** pixels as a threshold to remain region as a text region. Secondly, we used the vertical distance to decrease the false positives,

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 3 the difference between the effects of our binarization method which be based on Naïve Bayes and the Bradley thresholding method.

Table1: Comparison show the effects of our binarization 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

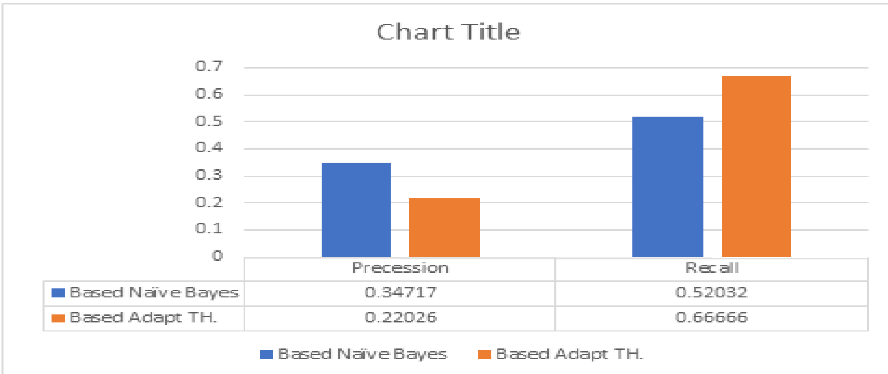

Fig. 3: Comparison show the effects of our binarization method and Bradley method in detecting characters in scene images

We present in the next Figure 4, samples images that reveal the superiority of our binarization model of the Bradley thresholoding model in specially in the complex images.

Original Image **Result based Bradley** **Result based our binarization scheme**

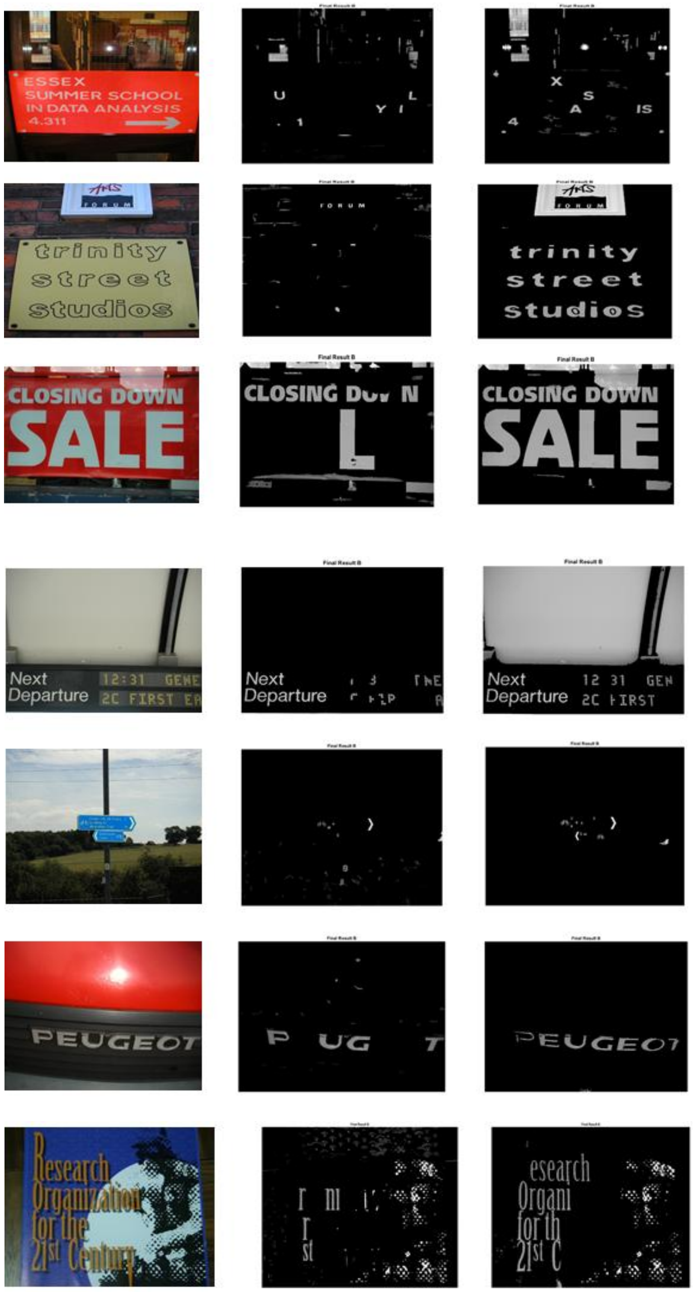

Fig. 4: Samples where our binarization method is better than Bradley thresholding method.

The next images in Figure 5 reveal that Bradley thresholding is better when the images are having blurring or strong illumination:

Original Image **Result based Bradley** **Result based our binarization scheme**

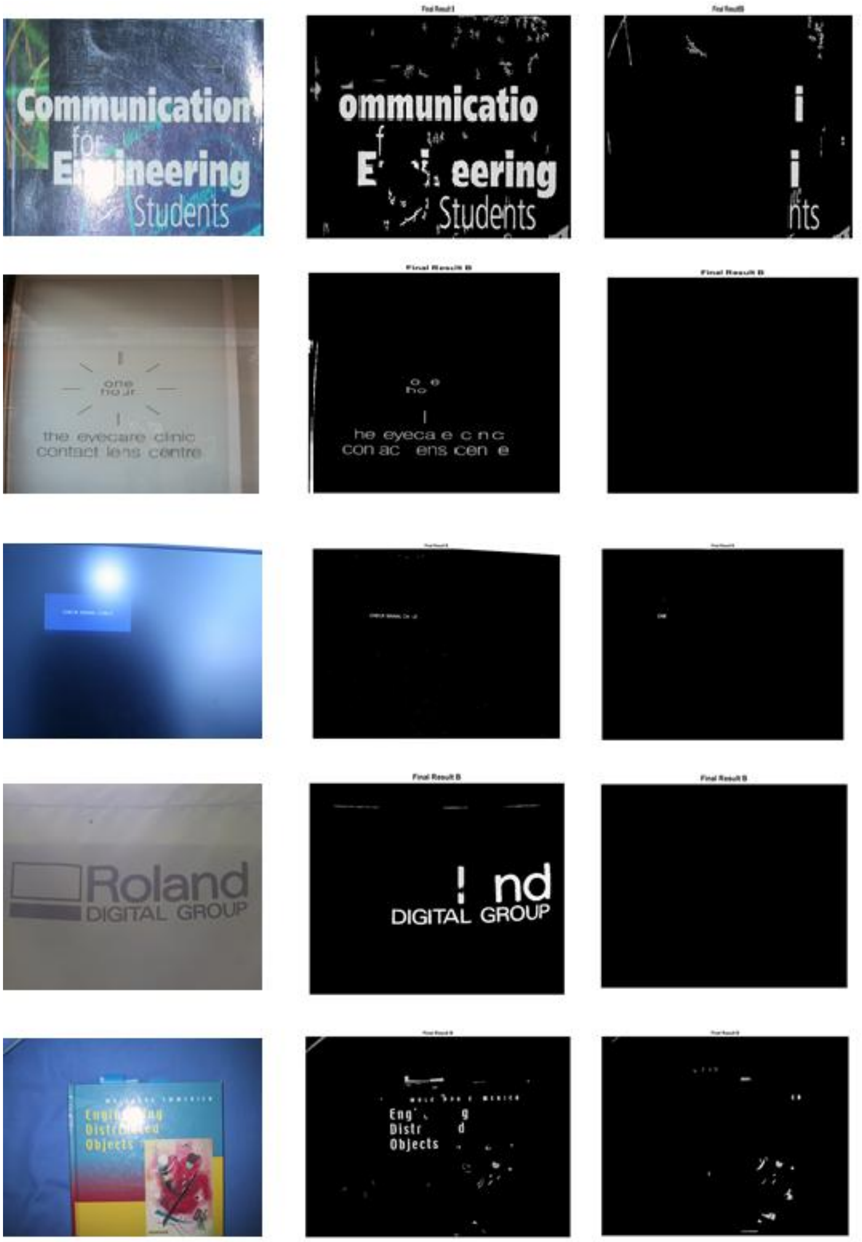

Fig. 5: Samples where adaptive thresholding is better than our binarization method

From the previous Table 1 and Figure 3, The proposed 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 binarization method with respect to metric **Recall.** However**,** the **Precession** metric when we used our binarization scheme is better than the corresponding metric in case of using Bradley method as shown in the previous table and previous Figure.

In the next Table 2 and Figure 6, we show comparison of our binarization technique with some existing thresholding techniques through their effects in detection characters in scene images. These methods are considered because they are the most famous effective methods in converting grayscale image into binary images.

Table 2: Comparison show the effects of six binarization methods in detecting characters in scene images

	Precession	Recall
EulerDist(im2bw)	0.350829795	0.477834239
EulerDist(Otsu)	0.368068713	0.510690501
EulerDist(Bayes)	**0.347170253**	**0.520329478**
EulerDist(Niblack)	0.045333936	0.140729057
EulerDist(Bernsen)	0.097683934	0.234314756
Based Adapt TH.	**0.22026**	**0.66666**

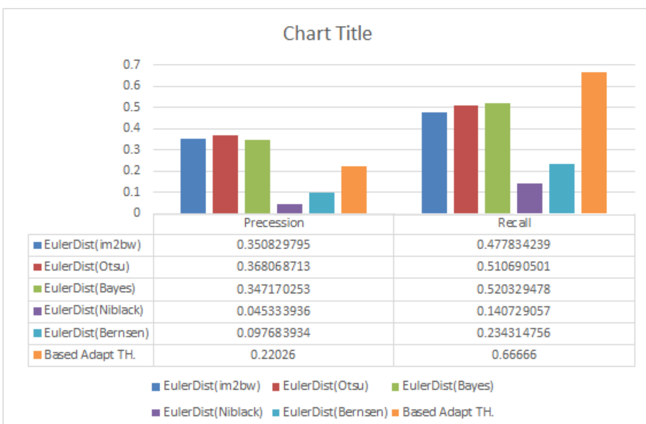

Fig. 6: Comparison between five binarization schemes and their effects in finding text characters.

In the previous Table 2 and Figure 6, we show that the highest **Recall** belongs to the last scheme whose binarization method depend on the

Bradley method. The following this our binarization technique which depend on Naïve Bayes classifier. This indicates that the maximum number of characters got from this binarization; Bradley method in spite of the superiority of our binarization in some images. The number of images that Naïve Bayes exceeds in extracted characters is 83 images, but the number of images that Otsu exceeds in extracted letters is 60 images. Although the **Recall** is highest in case of Bradley binarization, the **Precession** is the highest in our binarization method and the Otsu binarization method, this means the least false detections occur in our binarization and the Otsu binarization techniques.

6. DISCUSSION AND CONCLUSION

In this paper, we compare our previous binarization method with new technique; Bradley thresholding method through study their effects in extracting characters from scene images. This is not the only target of this work. The second target to enhance our previous technique to extract characters in scene images by using this Bradley adaptive thresholding instead of our binarization 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.

The study shows the importance of the two basic statistics; the mean and the standard deviation which we depend on as a features the Naïve Bayes have been used to classify the image pixels according to these statistics features. Also, this study shows that our binarization scheme outperforms the most well-known scheme among the binarization methods which called Otsu method with respect to Recall metric. Also, it outperforms the Bradley method with respect to Precession metric.

As a future work, more related features will be used to our classifier, including the gradient of the image and the entropy to reduce the false positives regions. Also, we try to make use the horizontal beside of the vertical distances between candidate regions, this will be more effective in defining and discarding text regions.

REFERENCES

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

[20] https://www.mathworks.com/help/images/ref/regionprops.html

طريقة بسيطة جديدة وفعالة لإستخراج الرموز والحروف النصية من الصورالطبيعية

د. عبد الرحيم أحمد هاشم و أ.د عبد الباسط عبد الله أحمد
قسم الرياضيات - كلية العلوم - جامعة أسيوط

فى هذا البحث يتم معالجة مشكلة استخراج النصوص من الصور الطبيعية لأهمية ذلك فى كثير من التطبيقات العملية مثل مساعدة فاقدى البصر أثناء السير فى التعرف على أسماء الشوارع والميادين والأماكن أثناء سيرهم وذلك بتوصيل مجسات تستطيع قراءة النصوص وتحويلها إلى صوت يصل لأذن الشخص. كذلك لأهميتها لدى إدارات المرور لمعرفة أسماء وأرقام السيارات المارة بسرعات عالية، إلى غير ذلك من التطبيقات العديدة.

يعتمد هذا البحث على تحويل الصورة الرمادية اللون إلى صورة ثنائية (أبيض وأسود) يسهل معها اكتشاف محتويات الصورة objects ومن هذا المحتويات ما هو رموز نصية ومنها ما هو غير ذلك، ثم عن طريق استخدام بعض الخصائص الهندسية والسمات الظاهرية لهذه الكائنات يمكن استبعاد المناطق غير النصية Non-character regions واستبقاء المناطق النصية Character regions فقط.

تم الاعتماد على طريقتين مختلفتين فى تحويل الصورة من رمادية إلى ثنائية أولاهما وهى المقدمة من قبل الباحث وفيها يتم استخدام مصنف Classifier ليتم تقسيم بكسالات الصورة إلى بكسالات خاصة بالكائنات وأخرى تنتمى للخلفية، وهذا المصنف هو Naïve Bayes Classifier والخصائص التى اعتمد عليها المصنف هى المتوسط والإنحراف المعيارى لكل بكسل فى الصورة الرمادية اللون. ويتم حساب هذا المتوسط والإنحراف المعيارى عند كل بكسل فى الصورة عن طريق أخذ جوار 3x3 حول هذه البكسل.

أمّا الطريقة الثانية فهى طريقة برادلى Bradley لتحويل الصورة إلى ثنائية ، وقمنا بمقارنة النتائج مع بعض الطرق المشهورة والواعدة فى تحويل الصورة الرمادية إلى ثنائية ووجدنا تفوق لطريقتنا المعتمدة على الـ Naïve Bayes Classifier على الطرق الواردة بالبحث، وقمنا بتعظيم هذه النتائج عند استخدام طريقة برادلى Bradley بدلا من طريقة Naïve Bayes فى التحويل إلى ثنائية.