To appear: in proceedings, IJCNN 2014, July 2014, Beijing China.
Final version to be available via IEEEXplore

Bayesian Network Scores Based Text Localization in Scene Images

Khalid Iqbal,Xu-Cheng Yin, *Member, IEEE,* Hong-Wei Hao, Sohail Asghar and Hazrat Ali

***Abstract*— Text localization in scene images is an essential and interesting task to analyze the image contents. In this work, a Bayesian network scores using K2 algorithm in conjunction with the geometric features based effective text localization method with the help of maximally stable extremal regions (MSERs). First, all MSER-based extracted candidate characters are directly compared with an existing text localization method to find text regions. Second, adjacent extracted MSER-based candidate characters are not encompassed into text regions due to strict edges constraint. Therefore, extracted candidate character regions are incorporated into text regions using selection rules. Third, K2 algorithm-based Bayesian networks scores are learned for the complimentary candidate character regions. Bayesian logistic regression classifier is built on the Bayesian network scores by computing the posterior probability of complimentary candidate character region corresponding to non-character candidates. The higher posterior probability of complimentary Candidate character regions are further grouped into words or sentences. Bayesian networks scores based text localization system, named as $BayesText$, is evaluated on ICDAR 2013 Robust Reading Competition (Challenge 2 Task 2.1: Text Localization) database. Experimental results have established significant competitive performance with the state-of-the-art text detection systems.**

I. Introduction

TEXTdetection in natural scene images is a challenging task in computer vision. The detected text has wide range of applications including license plate text localization [2], text extraction in video sequences [3], postal address box localization [4], automatic forms reading [5], globally document oriented systems [6], content-based image indexing [7], document image analysis [8] and to support visually impaired persons on sign board in streets [17]. In this context, text detection is a difficult task and attracts researchers to come up with a significant solution to overcome the poor performance of OCR engines [1].

Existing techniques for text detection in scene images can be classified into two classes: connected components (CC)-based and region-based techniques [9]. CC-based techniques extract components from scene images as candidate characters and group them if their geometric properties are similar by eliminating non-character regions with additional checks. Epshtem et al. [1] measured stroke width for each pixel. The approximately similar stroke width neighboring pixels are grouped into CCs. Consequently, text regions are found by filtering and aggregating CCs. Pan et al. [9] proposed hybrid method that exploits image regions to detect text candidates and extracts CCs as candidate characters by local binarization. False positive components are eliminated efficiently with the use of conditional random field (CRFs) model. Finally, character components are grouped into lines/words. Region-based techniques search an image for all possible text regions with use of sliding window to extract features of a region. These region features are classified using a classifier to identify text region. In order to identify text regions, Lee et al. [10] extracted six dissimilar classes of region features. These region features are classified by using Modest Adaboost to recognize text regions. Similarly, Yin et al. [11] used MSER (maximally stable extremal region) [18] to extract letter candidates by eliminating non-letter candidates by using geometric features. Similar geometric features letter candidates were grouped, using disjoint set, into candidate regions. Various features of candidate regions were extracted to build an Adaboost classifier for recognition of text regions. In this perspective, features of a letter candidate can help enhancing text localization accuracy in scene images rather than the use of a classifier. However, features of a letter candidate are restricted by their geometric shapes which are insufficient. Nevertheless, MSER-based extracted candidate character regions (letter candidates) can have comparable pixels that have a wide-ranging motivation for evaluation in conjunction with geometric features [11]. To evaluate the column-wise (features) pixels of MSER-based extracted candidate character region, Bayesian network scores are obtained by using K2 algorithm [12]. Bayesian network scores are classified with the Bayesian logistic regression classifier. In contrary to Yin et al. [34] low posterior probability based classified candidate character regions are eliminated rather than the higher posterior probability candidate character regions. Also, candidate selection rules [20] and perceptual text grouping [17] filters and repulsion scores [23] can be used to delineate text regions rather than using single-link clustering algorithm.

In this paper, we propose a Bayesian network scores in conjunction with geometric features of MSER-based text detection method for robustness and accuracy. First, all extracted MSER-based candidate character regions are matched with [11] for effectiveness that uses disjoint set to form

Khalid Iqbal is with Department of Computer Science and Technology, School of Computer and Communication Engineering, University of Science and Technology Beijing, Beijing 100083, P.R. China (email: kik.ustb@gmail.com).

Xu-Cheng Yin is with Department of Computer Science and Technology, School of Computer and Communication Engineering, University of Science and Technology Beijing, Beijing 100083, P.R. China (corresponding author, phone: 8610-82371191; fax: 8610-62332873; email:xuchengyin@ustb.edu.cn).

Hong-Wei Hao is with Institute of Automation, Chinese Academy of Sciences, Beijing 100190, P.R. China (email: hongwei.hao@ia.ac.cn).

Sohail Asghar is with University Institute of Information, PMAS-Arid Agriculture University Rawalpindi, Pakistan (email: sohail.asghar@uaar.edu.pk).

Hazrat Ali is with Department of Communication Engineering, School of Computer and Communication Engineering, University of Science and Technology Beijing, Beijing 100083, P. R. China. He is also with the Machine Learning Group, Department of Computing, City University London, United Kingdom. (email: engr.hazratali@yahoo.com).

text regions in scene images. However, direct comparison of extracted MSERs cannot effectively help in accuracy because of ANDing operation on the edges of candidate character regions and text regions of [11]. Second, we used candidate character regions selection rules in comparison with text regions in [11] to improve the performance. Despite this, several scene images text has not been localized by [11]. Third, Bayesian network scores of the complimentary candidate character region are computed using K2 algorithm [12] after resizing and thresholding. Bayesian network scores are used to train a Bayesian logistic regression classifier to filter out the non-candidate character region. Now, perceptual text grouping [17], selection rules [20] and repulsion scores [23] of the two candidate character regions are used to join them into words. Finally, by integrating the above steps, we build an accurate and robust text localization method in scene images. Our proposed method is evaluated on the recent benchmark ICDAR 2013 Robust Reading Competition (Challenge 2 Task 2.1: Text Localization) database and has achieved an $f - measure$ of 72.44% which is significant achievement.

The rest of the paper is organized as follows. Recent methods on MSER for scene text detection methods are reviewed in Section II. Section III describes our proposed text localization method. Experiments and results according to the performance evaluation criteria are presented in Section IV. Finally, the paper has been concluded.

II. Related Work

The integration of MSER has demonstrated significant performance in numerous real-world projects. However, existing MSER-based text localization methods detected repeating components and inadequate text regions constructing algorithms as key limitations. But, MSER algorithm has accomplished the fact to detect the most characters even in low quality images with low contrast and resolution; strong noise. Also, MSER allows detecting the most features of a candidate character region with significant response to define text regions in scene images. Therefore, repeating components detection by MSER-based methods is a minor problem that can be resolved while grouping candidate characters. An effective and novel MSER-based method is proposed by Yin et al. [11] to localize scene text that used geometric features and disjoint sets to eliminate non-characters and construct text regions respectively. Candidate text regions are classified by training an Adaboost classifier based on their features (horizontal and vertical variation, stroke width, color and geometry). Yin et al. [34] focused to prune MSERs by minimizing regularized variation. Single-link clustering algorithm is used to group candidate characters into text candidates by a new self-training distance learning algorithm. Text candidates with high posterior probabilities are eliminated to identify text regions. Fabrizio et al. [20] introduced novel approach that relies on the hypothesis generation/validation with the use of segmentation algorithm as well as use of SVM classifier and candidate character selection rules for grouping. Merino-Gracia et al. [17] also examined MSERs as candidate text region that relies on hierarchical relationship to quickly filter detected MSERs by using cascade of text classifying filters for grouping. Pan et al. [23] constructed a conditional random filed (CRF) to use visualize context for true detections with pairwise potentials including repulsion relationship.

The above mentioned MSER-based approaches still have insufficient text regions with limited number of features of a candidate character region. In contrary to [34] and in conjunction with [11], [17], [20], [24], Bayesian network scores for each candidate character regions are obtained through K2 algorithm [12] by discovering the causal link based on the intensity of pixels in a given column of MSER based extracted region. The key advantage of using Bayesian network scores in an incremental fashion is the discovery of scalability with an efficiency trade-off depending on the number of columns of candidate character region. However, we targeted to focus only detection quality and the unused candidate characters by Yin et al. [11] to enhance text detection accuracy with minimum effect on efficiency. In contrast to Yin et al. [34], low posterior probability candidate character regions are considered for grouping with the use of selection rules [20], few perceptual grouping filters [17] and repulsion scores from pairwise filters [23].

III. Text Localization in Scene Images

We incorporated Bayesian network scores in conjunction with Yins method [11] to localize text in scene images with remarkable improvements in comparison with the existing recent methods. The structure of our localization approach and sample results of each stage is shown in Figure 1. Text localization method follows the following stages.

$Stage1$: Candidate character regions are extracted using MSER algorithm and filtered by using constraint. Binarization is also performed on each candidate character region after resizing and thresholding to find candidate binary region. Additional details are presented in sub-section A.

$Stage2$: Text regions are constructed by pruning repeated components and locate text directly in images with the use of Yin et al. [11] that uses geometric features for text localization. However, several images text is still not recognized despite relaxing the constraints to include candidate character regions into text regions directly by using selection rules. The details about using selection rules are presented in subsection B.

$Stage3$: Each candidate binary character region has features according to the intensity of its pixel values. For each features of a candidate character region, Bayesian network scores are learnt for complimentary candidate binary characters using K2 algorithm [12] to train a Bayesian logistic regression classifier.

$Stage4$: A text classifier is used to classify the rest of true candidate character regions. Bayesian logistic regression classifier is trained to differentiate between candidate character and non-character regions corresponding to their labels [13], [14].

Fig. 1. Flowchart of Stage Wise Text Detection in Scene Images

A. Candidate Character Region Extraction and Filtration

The digital camera-based images may have distortions [15] with motivation by [16], [17] to localize text from scene images. Our method extracts MSER from scene image as candidate character regions. MSER was proposed by Matas et al. [18] to catch matching between different viewpoints of images. MSER is a well-known and best reported robust method against scale and light variation [19] that detects interest points by marking their locations. The location of an interest point in an image can have a range of points. Therefore, we extract the pixels of input image by taking minimum and maximum of located points of an interesting image region. Each extracted MSER is resized $(f \times f)$ to perform binarization using an average of pixels of the extracted region to define an adaptive threshold to perform binarization. Consider an image I, and a set of resized extracted MSERs, referred to as candidate character region, with f number of features. Binarization is performed to eliminate extra pixels of candidate characters using threshold as given by equation 1.

$$\tau = \frac{\Sigma_{c=1}^{k} rCCR_c}{k} \tag{1}$$

Where, CCR_c represent resized candidate character region pixels extracted from the input image, and $k = 25$ is the number of features. Binarization is performed on $rCCR_c$ by using τ as threshold to find candidate binary region using equation 2.

$$CBR = rCCR_c > \tau \tag{2}$$

Where, CBR represents candidate binary region. A constraint i.e. $(30 \leq$ Total number of Pixels in CBR $\leq 600)$ is applied on the number of pixels of a candidate binary character region to filter it as high or low density of pixels in CBR. The algorithmic steps of extraction and discarding of candidate character region are presented in a box as below.

Input $Image, P_{y\times x}, mserRegions, k = 25$
for $i = 1 : mserRegions$
$CCR_{ci} = Image(P_{minY}{:}P_{maxY}, P_{minX}{:}P_{maxX})$
$rCCR_{ci} = resize(CCR_c, k, k)$
$\tau = \frac{\Sigma_{c=1}^{k^2} rCCR_{ci}}{k^2}$
$CBR_i = rCCR_{ci} > \tau$
end for

B. Text Region Construction and Selection Rules

We directly carry on a step forward to consider groups of candidate character regions using effective text localization approach with the highest accuracy [11]. The effective text localization is proposed by Yin et al. [11] that extracts letter candidates and eliminates non-letter candidates using geometric information to construct text regions with the use of disjoint set by grouping similar letter candidates. Each candidate character region has four corner points as shown in Figure 2(a). Similarly, a text region can be defined as the group of candidate characters horizontally with the similar geometric features as shown by Figure 2(b). Therefore, text region identified by [11] are utilized to group extracted candidate character regions to localize text in scene images as shown in Figure 2.

However, a huge number of candidate characters are left by [11] as ungrouped. The complimentary candidate character regions are further examined by using adjacent candidate characters selection rules proposed by [20]. The selection rules, given by Figure 3, are used to group extracted candidate character regions with the detected text regions in 2(b). However, these constraints are closely defined to group candidate character regions with the localized text regions without affecting the existing results. Also, a candidate character region must be on the left side of the text region for improvement. From Figure 2 and Figure 3, slight improvement can be observed after grouping the adjacent candidate character regions with the detected text region. The slight improvement in text region is due to closely defined selection rules threshold values with an aim to reduce the effect to Figure 2 recognized text regions. Nevertheless, close-fitting constraints for grouping candidate character regions cannot

(a) Extracted Candidate Character Region Corners (*left*) and Text Regions (*right*)

(b) Text Regions Localized

Fig. 2. Grouping of Candidate Character Region using Yin et al. Method [11]

recognize maximum possible text regions with minimum error. Therefore, subsequent subsection learn Bayesian network score using K2 algorithm to further investigate remaining candidate character region for localization of text in scene images.

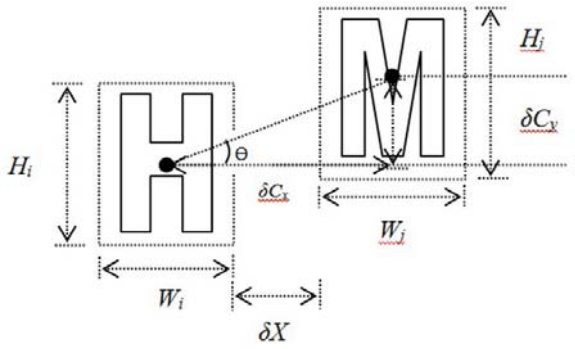


Fig. 3. Adjacent Candidate Characters Selection Rules

Edge angle ($i.e.\theta$=$\tan^{-1}\frac{\delta C_y}{\delta C_x} < 30°$), used by the OCR engine, is the angle between the edges of candidate character regions and the horizontal axis. This constraint limits the majority of candidate characters which are aligned horizontally or with slight slope. In addition, simple grouping procedure of the candidate character region to form words, lines or text region. According to Figure 3, H_i and H_j, are the heights of two candidate characters, vertical and horizontal centroid distances are δC_y and δC_x as well as distance between two candidate character regions along X-axis. Three conditions are tested before connecting two candidate character regions such as $Height\_Similarity = \frac{|H_i - H_j|}{min(H_i,H_j)} < 0.3$, $Alignment = \frac{\delta C_y}{min(H_i,H_j)} < 0.3$ and $Vicinity = \frac{\delta X}{min(H_i,H_j)} < 0.5$ which improved text detecting accuracy and can be observed in Figure 4.

C. Bayesian Network Scores and K2 Algorithm

Bayesian network scores can be learnt in an incremental fashion. For large number of candidate binary region with k features, Bayesian network scores can have an efficiency problem. Therefore, aspect ratio is used to reject too narrow or too wide candidate character regions to increase the processing performance. The complimentary candidate character regions (candidate binary regions) are presented in Figure 5 $(Redcolor)$. For n^{th} candidate character region with height H_i and width W_i, aspect ratio can be defined by 3 with the use of constant threshold values $L_h = 0.3$,$L_w = 0.1$,$U_h = 1.15$,$U_w = 5$ throughout in our experiments.

$$L_h < \frac{H_i}{W_i} < U_h \quad and \quad L_w < \frac{W_i}{H_i} < U_w \tag{3}$$

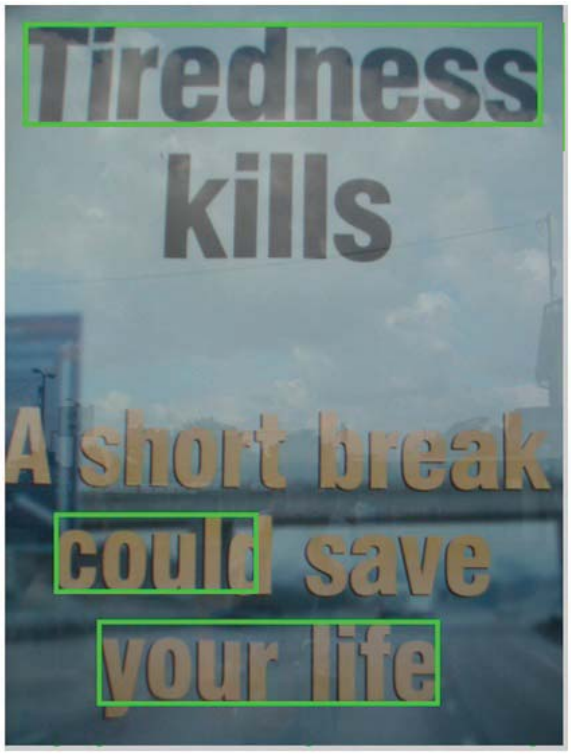


Fig. 4. Improvement in Text region using Selection Rules

Besides, Bayesian network is an extensively used model for data analysis. Bayesian networks [12] exploit a set of features of candidate binary region (as shown in Figure 6) of an image to measure probabilistic relationship graphically among features. The optimal determination of Bayesian network structure of a given data is NP-hard problem [21]. K2 algorithm [12] is one of the widely used efficient heuristic solutions to learn structural relationship dependencies among features according to a specified order. The initial feature in a given order has no parents. K2 then adds incrementally

parents for a current feature by increasing the score in resulting structure. When no predecessor features addition for a current feature as parent can increase the score, K2 stops adding parents to feature. As ordering of features are known beforehand, so search space is minimized make it efficient as well as parents of features can be chosen independently [22].

Fig. 5. Complimentary Candidate Character Regions

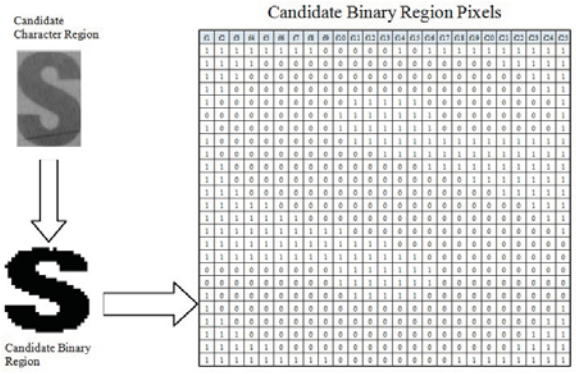


Fig. 6. Preprocessing of Extracted Candidate Character Region

Consider $F = \{f_1, f_2, f_3 \ldots f_k\}$ are the features of candidate binary region of scene image. All the candidate binary regions features are ordered from left to right for simplicity. To find scores of each feature, scoring function of K2 algorithm for i^{th} features is presented by equation 4. The final scores of all features in a network structure are obtained by multiplying the individual score of features of candidate binary character.

$$g(i, \pi_i) = \prod_{i=1}^{n} \prod_{j=1}^{T_i} \frac{(r_i - 1)!}{(N_{ij} + r_i - 1)!} \prod_{k=1}^{r_i} \alpha_{ijk}! \qquad (4)$$

Where, $g(i, \pi_i)$ represent the Bayesian network score using K2 algorithm for each feature of candidate binary character region of an image. However $g(i, \pi_i)$ (where, $i = 1$), has no parents and zero score. Therefore, we omit the first feature of candidate binary character region in all the test set ICDAR 2013 images. The purpose of neglecting initial score of candidates binary characters is to take the matrix product as a composite transformation of Bayesian network K2 scores. The algorithmic steps of CBR Bayesian network scores are presented as below.

Initialize $Order = 1{:}k, u = 4$
for $j = 1 : numberOfImages$
  $mserRegions = mserFeatures(j))$
  $Score = k2Score = []$
  for $i = 1 : mserRegions$
    Input CBR_i
    $Score = K2(CBR_i, Order, u)$
    $k2Score = [k2Score; Score]$
  end for
  $CBR_{score} = k2Score(:, 2{:}end) * k2Score(:, 2{:}end)^T$
end for

D. Classification of Bayesian Networks Scores

Bayesian Logistic Regression is used to avoid over-fitting, classification effectiveness and efficient in fitting and at prediction time [13], [14]. Bayesian Logistic Regression allows modeling a relationship between a Bayesian network scores corresponding to labeled candidate character regions, in which a statistical analysis is taken under Bayes rule. The application of Bayesian Logistic Regression can be applied in pattern recognition to classify Bayesian network scores of candidate character regions for two or more classes. The most important advantage of Bayesian Logistic Regression is its dominating classification accuracy while estimating a probabilistic relationship between Bayesian network scores and labeled candidate character regions.

Suppose, $L = \{0, 1\}$ be the labels of Bayesian network scores of candidate binary character regions C. The log-likelihood ratio $\ln(P(L = 1/C, \omega)/P(L = 0/C, \omega))$ is assumed to be linear in R, such that conditional likelihood for $L = 1$ is given by the sigmoid function $P(L = 1/C, \omega) = \frac{1}{1+e^{-\omega^T C}} = \sigma(-\omega^T C)$. Similarly, $P(L = 0/C, \omega) = 1 - P(L = 1/C, \omega) = \frac{1}{1+e^{\omega^T C}}$, such that $P(L/C, \omega) = \sigma\omega^T C)$.

Let training set T composed of Bayesian network scores of an image candidate character regions R and their labels L,i.e. $\mathrm{T} = \{\mathrm{L}, \mathrm{C}\}$ are input parameters to Bayesian Logistic Regression Classifier. Conditional probability $P(\omega/\mathrm{T})$ and priori probability $P(\omega)$ are used to compute posterior probability. As sigmoid likelihood data does not permit a conjugate-exponential prior to find posterior analytic expression. The quadratic approximation ω is used exponentially, such that conjugate-Gaussian prior with parameterization of α as hyper-parameter, a_0 and b_0 as hyper prior parameters. The prior of this approximation and modeling of conjugate

Gamma distribution can be represented as given by equations 5 and 6.

$$P(\omega/\alpha) = (\frac{\alpha}{2\pi})^{1/2} e^{\frac{\alpha}{2}\omega^t\omega} \quad (5)$$

$$P(\alpha) = (\frac{1}{\Gamma a_0}) b_0^{a_0} \alpha^{a_0-1} e^{-b_0\alpha} \quad (6)$$

Candidate character regions are eliminated if its posterior probability is less than 0.5. The output of the classified candidate character regions is shown in Figure 7 (red color).

Fig. 7. Complimentary Classified Candidate Character Regions ($RedColor$) and Text Regions ($GreenColor$)

E. Grouping of Scores based Classified Complimentary Candidate Characters

From equation 3, height similarity and alignment measures are adapted for the remaining set of candidate character regions. The experimental values of height similarity and alignment are exclusively fluctuating from 70% to 90% and 5% to 15% respectively. In addition, range of horizontal centroid difference i.e. $|C_{xi}, C_{xj}|$ is used to group similar candidate character regions [20] as given by equation 7.

$$0.1max(\frac{H_i}{2}, H_j 2) < |C_{xi}, C_{xj}| < 2max(\frac{H_i}{2}, \frac{H_j}{2}) \quad (7)$$

Similarly, vertical centroid difference i.e. i.e. $|C_{yi}, C_{yj}|$ is used to group two candidate character regions to find text region as given by equation 8.

$$|C_{yi}, C_{yj}| < 0.1max(\frac{W_i}{2}, \frac{W_j}{2}) \quad (8)$$

Lastly, Repulsion Score [23] measures the degree of repulsion between two candidate character regions. However, two candidate character regions are considered as likely to group if repulsion score is greater than 0.7 and computed by equation 9.

$$RepulsionScore = 1 - min(1, \frac{\|C_i, C_j\|}{max(D_i, D_j)}) \quad (9)$$

Where, C_i,C_j are the centroid coordinates and D_i,D_j are half diagonals of the two candidate character regions.

F. Grouping of Complimentary Candidate Characters using Connected Components

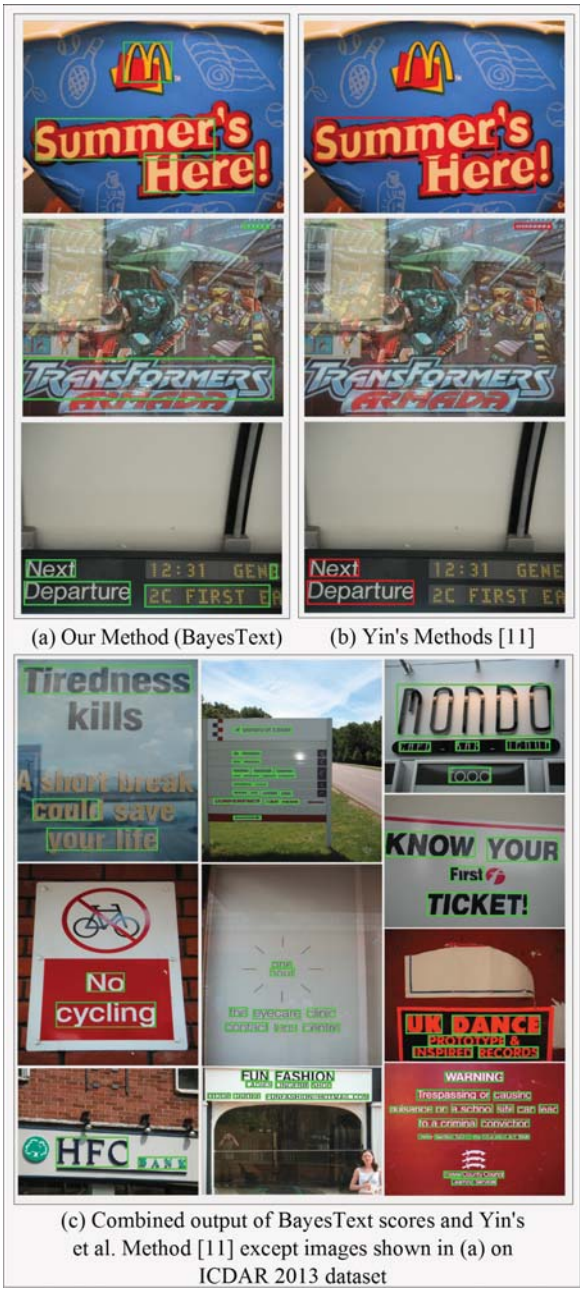
The above filtered measures are used to group candidate character regions. The closed complimentary classified candidate character regions are grouped into text regions using connected components and the extracted grouped text regions are shown in Figure 8. Therefore, connected components with 8-neighbour metric, having at least a sum of 8 pixels, has been used to localize text based on the complimentary classified adjacent candidate character regions. Finally, the bounding box of complimentary classified text regions in Figure 8 and text region bounding box in $green$ color in Figure 7 are combined into Figure 9 which reflects slight improvement in scene image text detection.

Fig. 8. Complimentary Classified Candidate Characters Grouping

IV. EXPERIMENTS

The ICDAR 2013 Robust Reading Competition (Challenge 2 Task 2.1: Text Localization) database [25] is widely used benchmarking database for text localization algorithms. The database contains 233 testing images. The Bayesian network using K2 algorithm scores based method is trained on this database. In order to measure the performance of our method, text regions are constructed with the use of Yin et al. method [11] as well as to enlarge the text regions by adding closed candidate character region with the help of constraints presented by Figure 3. Now, the complimentary set of candidate character regions are taken into account to learn Bayesian network scores by K2 algorithm with an objective to train a Bayesian logistic regression classifier [13], [14]. After classification, the horizontally aligned candidate characters are grouped according to filters described in sub-section E of the previous section. The complimented

classified candidate characters are grouped using connected components for localization of text regions in the form of words and sentences. The extracted localized text regions are shown in Figure 9.



(a) Our Method (BayesText) (b) Yin's Methods [11]

(c) Combined output of BayesText scores and Yin's et al. Method [11] except images shown in (a) on ICDAR 2013 dataset

Fig. 9. Text Regions Localized in Scene Images

A. Performance Evaluation of Text Localization Systems

To evaluate the performance of our method with the state-of-the-art text localization methods, the ICDAR 2013 text localization competition scene image database and evaluation procedure are used. This database is provided with the ground truth patterns for a given scene image text as words. The detection quality is evaluated through precision and recall, defined as $p = \frac{\sum_{r_e \varepsilon E_b} m(r_e, T_g)}{|E_b|}$ and $r = \frac{\sum_{r_t \varepsilon T_g} m(r_t, E_b)}{|T_g|}$, where $m(r, R_b)$ defines the best match for a rectangular bounding box r in a set of rectangles R_b. The set of ground-truths and estimated rectangular bounding boxes refers to T_g and E_b respectively. All the algorithms are evaluated and compared with respect to $f - measure$. $F - measure$ is the harmonic mean of recall and precision defined as $f = \frac{2}{\frac{1}{precision} + \frac{1}{recall}}$. However, there are certain difficulties in order to get the relevant evaluation. The difficulties in fair evaluation are lack of consistency in ground truths, granularity of annotation and uneven weight for text regions by averaging precision and recall over all images after computing them for each image [20].

B. Results

We evaluate our method on ICDAR 2013 benchmark database [25] and our reported results are presented in Table I. A direct comparison of MSER-based candidate character region with Yin et al . [11] gives $f - meausre$ as 72.38. However, our method (BayesText) provides competitive performance result 72.44 as $f - measure$ and can be ranked as 4^{th} out of 10 on the most recently published results. Figure 9 illustrates several successful results on ICDAR 2013 database scene images. Figure 9(a) gives slight enhancement over Yin et al. [11] in comparison with Figure 9(b) and Figure 9(c) reflects equivalent results by combining Bayesian network scores with Yin et al [11].

TABLE I
PERFORMANCE (%) COMPARISON OF TEXT LOCALIZATION METHODS ON ICDAR 2013 DATABASE

Method	Recall	Precision	f-Measure
USTB_TexStar [34]	66.45	88.47	75.89
Text Spotter [26], [27], [28]	64.84	87.51	74.49
CASIA_NLPR [29], [30]	68.24	78.89	73.18
Our Method (BayesText)	**63.51**	**84.30**	**72.44**
Text_detector_CASIA [31], [32]	62.85	84.70	72.16
I2R NUS FAR	69.00	75.08	71.91
I2R NUS	66.17	72.54	69.21
TH-TextLoc	65.19	69.96	67.49
Text Detection [20], [33]	53.42	74.15	62.10
Baseline	34.74	60.76	44.21
Inkam	35.27	31.20	33.11

V. CONCLUSION AND FUTURE WORK

In this paper, MSER-based scene text localization method is presented with the partial use of Bayesian network scores in conjunction with an effective text localization method based on geometric features. Bayesian network scores are computed for each candidate character region. Bayesian logistic regression classifier is trained to estimate the posterior probability of complimentary candidate characters to eliminate non-character candidates which assists in building a more powerful character classifier. Consequently, by integrating Bayesian network scores, we build a robust scene text localization method that demonstrates a remarkable and

rational comparative performance with the recent state-of-the-art methods. However, our method (BayesText) is dependent on Yin's et al. [11] method for direct comparison of MSER-based extracted candidate character regions. Besides, automatic labeling of MSER-based extracted candidates requires preprocessing. In this context, removal of dependency on Yin's et al. [11] and preprocessing step for automatic labeling of candidates could be our possible future work for further generalization and improvements.

ACKNOWLEDGMENT

The research was supported by National Natural Science Foundation of China (61105018, 61175020).

REFERENCES

[1] B. Epshtein, E. Ofek, and Y. Wexler, *Detecting text in natural scenes with stroke width transform.* In Computer Vision and Pattern Recognition, pp. 2963-2970, 2010.

[2] C. Arth, F. Limberger, and H. Bischof, *Real-time license plate recognition on an embedded DSP-platform.* IEEE International Conference on Computer Vision and Pattern Recognition, pp. 1–8, 2007.

[3] C. Wolf, J. M. Jolion, and F. Chassaing, *Text localization, enhancement and binarization in multimedia documents.* In Proceedings of the International Conference on Pattern Recognition, pp. 1037–1040, 2002.

[4] P.W. Palumbo, S.N. Srihari,J. Soh, R. Sridhar, and V. Demjanenko, *Postal address block location in real time.* Computer, Vol. 25, Issue 7, pp. 34–42, 1992.

[5] E. Kavallieratou, D. Balcan,M. Popa, and N. Fakotakis, *Handwritten text localization in skewed documents.* In International Conference on Image Processing, pp. I:1102–1105, 2001.

[6] F. Wahl, K. Wong,and R. Casey, *Block segmentation and text extraction in mixed text/image documents.* Computer Graphics and Image Processing, Vol. 20, Issue 4, pp. 375–390, 1982.

[7] R. Lienhart, and W. Eelsberg, *Automatic text segmentation and text recognition for video indexing.* Multimedia Systems, Vol. 8, Issue 1, pp. 69–81, 2000.

[8] Y. Xiao, and H. Yan, *Text region extraction in a document image based on the delaunay tessellation.* Pattern Recognition, Vol. 36, Issue 3, pp. 799–809, 2003.

[9] Y.-F. Pan, X. Hou, and C.-L. Liu, *A hybrid approach to detect and localize texts in natural scene images.* IEEE Transactions on Image Processing, Vol. 20, No. 3, pp. 800–813, 2011.

[10] J.-J. Lee, P.-H. Lee, S.-W. Lee, A. Yuille, and C. Koch, *Adaboost for text detection in natural scene.* In International COnference on Document Analysis and Recognition, pp. 429-434, 2011.

[11] X. Yin, X.-C. Yin, H.-W. Hao, and K. Iqbal, *Effective text localization in natural scene images with MSER, geometry-based grouping and AdaBoost.* 21st IEEE International Conference on Pattern Recognition, pp. 725–728, 2012.

[12] G. F. Cooper and E. Herskovits, *A Bayesian method for the induction of probabilistic networks from data.* Machine Learning, Vol. 9, No.4, pp. 309–347, 1992.

[13] K. Iqbal, X.-C. Yin, H.-W. Hao, X. Yin, and H. Ali, *Classifier Comparison for MSER-Based Text Classification in Scene Images.* The 2013 International joint Conference on Neural Networks, pp. 1–6, 2013.

[14] D.W. Hosmer,and S. Lemeshow, *Applied Logistic Regression.* (2nd edition). Wiley. ISBN 0-471-35632-8.

[15] X.-C. Yin, H.-W. Hao, J. Sun, and S. Naoi. *Robust Vanishing point detection for MobileCam-based documents.* IEEE International Conference on Document Analysis and Recognition, pp. 136-140, 2011.

[16] A. Shahab, F. Shafait, and A. Dengel, *ICDAR 2011 robust reading competition challenge 2: Reading text in scene images.* International Conference on Document Analysis and Recognition, pp. 1491-1496, 2011.

[17] C. Merino-Gracia, K. Lenc, and M. Mirmehdi, *A head-mounted device for recognizing text in natural scenes.*, Camera-Based Document Analysis and Recognition, Springer Berlin Heidelberg, pp. 29–41, 2012.

[18] J. Matas, O. Chum, M. Urban, and T. Pajdla, *Robust wide baseline stereo from maximally stable extremal regions.* In British Machine Vision Conference, Vol. 1, pp. 384-393, 2002.

[19] K. Mikolajczyk, T. Tuytelaars, C. Schmid, A. Zisserman, J. Matas, F. Schaffalitzky, T. Kadir, and L. V. Gool, *A Comparison of Affine Region Detectors.* International Journal of Computer Vision, Vol. 65, Issue 1, pp. 43-72, 2005.

[20] J. Fabrizio, B. Marcotegui, and M. Cord, *Text detection in street level images.* Pattern Analysis and Applications, pp. 1–15, 2013.

[21] D. M. Chickering, *Learning Bayesian networks is NP-complete.* In D. Fisher and H.J. Lenz, editors, Learning from Data: Articial Intelligence and Statistics V, Springer-Verlag, pp. 121–130, 1996.

[22] N. Friedman, and D. Koller, *Being Bayesian About Network Structure: A Bayesian Approach to Structure Discovery in Bayesian Networks.* Machine Learning, Vol. 50, Issue (1-2), pp. 95–125, 2003.

[23] J. Pan, Y. Chen, B. Anderson, P. Berkhin, and T. Kanade, *Effectively leveraging visual context to detect texts in natural scenes.* In Proceedings of Asian Conference Computer Vision 2012.

[24] R.M. Haralick, and L.G. Shapiro. *Computer and Robot Vision.* Volume I. Addison-Wesley, 1992. pp. 28-48.

[25] D. Karatzas, F. Shafait, S. Uchida, M. Iwamura, S.R. Mestre, J. Mas, D. F. Mota, J. A. Almazan, L.P. de las Heras, *ICDAR 2013 Robust Reading Competition.* IEEE 2013 12th International Conference on Document Analysis and Recognition (ICDAR), pp. 1484–1493, 2013.

[26] L. Neumann, and J. Matas, *A method for text localization and recognition in real-world images.* In Proceedings of Asian Conference on Computer Vision, pp. 2067-2078, 2010.

[27] Neumann, Lukas, and Jiri Matas, *Real-time scene text localization and recognition.* In Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on, pp. 3538–3545. IEEE, 2012.

[28] Neumann, Lukas, and Jiri Matas, *On combining multiple segmentations in scene text recognition.* In Document Analysis and Recognition (ICDAR), 2013 12th International Conference on, IEEE, pp. 523–527, 2013.

[29] Y.-M. Zhang, K.-Z. Huang, and C.-L. Liu, *Fast and robust graph-based transductive learning via minimum tree cut.* In Data Mining (ICDM), 2011 IEEE 11th International Conference on, IEEE, pp. 952–961, 2011.

[30] B. Bai, F. Yin, and C.L. Liu. *Scene text localization using gradient local correlation.* 2013 12th IEEE International Conference on Document Analysis and Recognition, pp. 1380–1384, 2013.

[31] C. Shi, C. Wang, B. Xiao, Y. Zhang, S. Gao, and Z. Zhang, *Scene text recognition using part-based tree-structured character detections.* International Conference on Computer Vision and Pattern Recognition, pp. 2961–2968, 2013.

[32] C. Shi, C. Wang, B. Xiao, Y. Zhang, and S. Gao, *Scene text detection using graph model built upon maximally stable extremal regions.* Pattern Recognition Letters, Vol. 34, No. 2, pp. 107-116, 2013.

[33] J. Fabrizio, B. Marcotegui, and M. Cord, *Text segmentation in natural scenes using toggle-mapping.* IEEE Inernational Conference on Image Processing, pp. 2373–2376, 2009.

[34] X.-C Yin, X. Yin, K. Huang, and H.-W. Hao *Robust Text Detection in Natural Scene Images.*, IEEE Transactions on Pattern Analysis and Machine Intelligence, preprint, 2013.