K2 Algorithm-based Text Detection with An Adaptive Classifier Threshold

Khalid Iqbal *kik.ustb@gmail.com*
Dept. of Computer Science and Technology,
School of Computer and Communication Engineering,
University of Science and Technology Beijing,
Beijing 100083, China.

Xu-Cheng Yin *xuchengyin@ustb.edu.cn*
Dept. of Computer Science and Technology,
School of Computer and Communication Engineering,
University of Science and Technology Beijing,
Beijing 100083, China.

Hong-Wei Hao *hongwei.hao@ia.ac.cn*
Institute of Automation
Chinese Academy of Sciences,
Beijing 100190, China.

Sohail Asghar *sohail.asghar@uaar.edu.pk*
University Institute of Information,
PMAS-Arid Agriculture University,
Rawalpindi, Pakistan

Hazrat Ali *engr.hazratali@yahoo.com*
Department of Computing,
City University London,
United Kingdom.

Abstract

In natural scene images, text detection is a challenging study area for dissimilar content-based image analysis tasks. In this paper, a Bayesian network scores are used to classify candidate character regions by computing posterior probabilities. The posterior probabilities are used to define an adaptive threshold to detect text in scene images with accuracy. Therefore, candidate character regions are extracted through maximally stable extremal region. K2 algorithm-based Bayesian network scores are learned by evaluating dependencies amongst features of a given candidate character region. Bayesian logistic regression classifier is trained to compute posterior probabilities to define an adaptive classifier threshold. The candidate character regions below from adaptive classifier threshold are discarded as non-character regions. Finally, text regions are detected with the use of effective text localization scheme based on geometric features. The entire system is evaluated on the ICDAR 2013 competition database. Experimental results produce competitive performance (precision, recall and harmonic mean) with the recently published algorithms.

Keywords: Bayesian Network, Adaptive Threshold, Bayesian Logistic Regression, Scene Image.

1. INTRODUCTION

With the advent of digital cameras at consumer-end, new challenges are opened to process and analyze the content of images. The image content captured with a camera may have several degradations such as distorted view, uneven illumination, curved or shadow effects. However,

image content can have useful information about vehicle license plates, landmarks recognition, gas/electricity meters and product recognition. The exploitation of image content constitutes a challenging research area to detect and recognize scene text. The extracted information from scene images has to be robustly located. The determination of locating text in scene images, with complex background, different illumination and variable fonts and size, and text orientation, refers to text localization.

Text localization techniques can be grouped into region-based, connected component (CC)-based [1] and hybrid methods [2]. Region-based techniques employ a sliding window to look for image text with the use of machine learning techniques for text identification. Sliding window based methods tend to be slow due to multi scale processing of images. A new text detection algorithm extracts six dissimilar classes of text features. Modest AdaBoost classifier is used to recognize text regions based on text features [3]. CC-based methods group extracted candidate characters into text regions with similar geometric features. CC-based methods are demanding to apply additional checks for eliminating false positives. To find CCs, stroke width for every pixel is computed to group neighboring pixels. These CCs were screened and grouped into text regions [4]. Pan et al. [2] 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. Recently, Yin et al. [5] extracted maximally stable extremal regions (MSERs) as letter candidates. Non-letter candidates are eliminated using geometric information. Candidate text regions are constructed by grouping similar letter candidates using disjoint set. For each candidate text region, vertical and horizontal variances, color, geometry and stroke width are extracted to identify text regions using Adaboost classifier. Besides, MSER based method is the winner of ICDAR 2011 Robust Reading Competition [6] with promising performance.

In this paper, we proposed a new robust and accurate MSER based approach to localize text in scene images. Firstly, MSER is extracted as candidate characters. Some of the non-character candidates are pruned based on total number of pixels. Second, Bayesian network scores for each candidate characters feature are obtained using K2 algorithm [7]. Third, Bayesian Logistic Regression Classifier (BLRC) [8, 9] is built from these score based features of candidate characters and used to identify text regions. Also, a posterior probability based adaptive threshold is used to filter out non-character candidates. In order to use ICDAR 2013 [10] competition dataset, candidate characters are grouped into words by using effective text localization approach [5]. The output of each step is shown in Figure 1. Our method is tested on the ICDAR 2013 test dataset. The experimental results show the remarkable performance in terms of accuracy.

The remainder structure of this paper is organized as follows. Section II explains the text localization method in perspective of candidate character extraction using MSER, Bayesian network score learning, Bayesian logistic regression classifier with an adaptive threshold in subsequent subsection and grouping of candidate character regions. Section III briefly describes the comparative performance of different text localization algorithm experimental results. In last section, the paper has been concluded.

2. K2 SCORES AND ADAPTIVE THRESHOLD BASED TEXT DETECTION METHOD

Text localization method is divided into four parts: candidate character extraction using MSER, Bayesian network scores for each candidate character, classification of candidate characters using Bayesian network scores with the use of adaptive threshold and group candidate characters by eliminating false positives.

2.1 Candidate Character Extraction with MSER

The digital camera-based images may have distortions [11] with motivation by [6, 12] to localize text from scene images. Our method extracts MSER from scene image as candidate characters region. MSER was proposed by Matas et al. [13] to catch matching between different viewpoints of images. MSER is a well-known and best reported robust method against scale and light variation [14]. Each extracted MSER is resized (*f X f*) to perform binarization using an adaptive threshold. Consider an image *I*, and a set of resized extracted MSERs E_{rc} referred as candidate character region, with *f* number of features. Binarization is performed to eliminate extra pixels of candidate characters using an adaptive threshold as given in equation 1. Additional check is applied on the number of pixels of all candidate character regions to filter out as non-character regions. The purpose of additional constraint is applied on candidate binary character region to minimize the processing time of learning Bayesian network scores using K2 algorithm [7].

$$\tau = \frac{1}{k}\sum_{r=1}^{k}\left(\sum_{c=1}^{k} CCR_{rc} \Big/ k\right) \rightarrow (1)$$

Where, CCR_{rc} represent candidate character region before application of binarization, and *k=25* is the number of features.

$$CBR = CCR > \tau \rightarrow (2)$$

Where, CBR represent the candidate binary character region. A candidate binary character region is discarded as non-character if it has less than or equal to 30 pixels or greater than or equal to 600 pixels.

2.2 K2 Algorithm-based Bayesian Network Scores

Bayesian network is an extensively used model for data analysis. Bayesian networks [7] exploit a set of features of candidate character binary region 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 [15]. K2 algorithm [7] 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 [16]. Consider $F = \{f_1, f_2, f_3 \ldots f_k\}$ are the features of candidate binary character region of scene image. All the candidate binary character 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 3. The final scores of all features in a network structure are obtained by multiplying the individual score of features of candidate binary character.

$$g(i,\pi_i) = \prod_{i=1}^{k}\prod_{j=1}^{q_i}\frac{(r_i-1)!}{(N_{ij}+r_i-1)!}\prod_{z=1}^{r_i}\alpha_{ijk}\,! \rightarrow (3)$$

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 omitting initial feature of candidates' binary characters is to produce point to point product of all Bayesian network K2 scores as given in equation 4.

$$K2F_{score} = [g(i-1, \pi_{i-1})] \circ [g(i-1, \pi_{i-1})]^T \rightarrow (4)$$

2.3 Classification of Bayesian Network Scores

Bayesian Logistic Regression is an approach that models a relationship between a Bayesian network scores and automatically 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 character regions *R*. The log-likelihood ratio $\ln\left(P(L=1/R,\omega)/P(L=0/R,\omega)\right)$ is assumed to be linear in R, such that conditional likelihood for *L=1* is given by the sigmoid function. $P(L=1/R,\omega) = 1/\left(1+e^{-\omega^T R}\right) = \delta(-\omega^T R)$. Similarly, $P(L=0/R,\omega) = 1-P(L=1/R,\omega) = 1/\left(1+e^{\omega^T R}\right)$ such that $P(L/R,\omega) = \delta(L\omega^T R)$.

Let training set τ composed of Bayesian network scores of an image candidate character regions *R* and their labels *L, i.e.* $\tau = \{\mathrm{L}, \mathrm{R}\}$ are input parameters to Bayesian Logistic Regression Classifier. Conditional probability $P(\omega/\tau)$ 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. The prior of this approximation and modeling of conjugate Gamma distribution can be represented as given by equations (5) and (6).

$$P(\omega/a) = \left(\frac{\alpha}{2\pi}\right)^{1/2} e^{\frac{\alpha}{2}\omega^T\omega} \rightarrow (5)$$

$$P(a) = \left(\frac{1}{\Gamma a_0}\right) b_0^{a_0} \alpha^{a_0-1} e^{-b_0\alpha} \rightarrow (6)$$

FIGURE 1: Text Detected in ICDAR 2013 Scene Images.

2.4 Posterior Probability-based Adaptive Threshold

To filter non-character regions, an adaptive threshold is computed using posterior probabilities computed by Bayesian Logistic regression in previous subsection. For simplicity, we use geometric mean and standard deviation of $P(\omega/a)$ to filter non-character candidate regions. An adaptive threshold is computed by equation 7.

$$T = \sqrt[n]{\prod_{i=1}^{n} P_i(\omega/a)} - 3\left(\frac{n\sum_{i=1}^{n} P_i(\omega/a)^2 - (\sum_{i=1}^{n} P_i(\omega/a))^2}{n^2}\right) \rightarrow (7)$$

The entire candidate character region of an image classified through Bayesian network scores are eliminated if their posterior probability is less than or equal to adaptive threshold value $\mathbb{T}$.

2.5 Candidate Character Regions Filtering and Grouping

Aspect ratio [12] is a cascade of filter that can be used to eliminate too wide or too narrow candidate regions. For example, W_i is a width and H_i is the height of candidate region. In this way, candidate character regions are further reduced after adaptive threshold.

$$l_{min} < \frac{W_i}{H_i} < l_{max} \rightarrow (8)$$

In all our experiments, adaptive thresholds are adapted empirically. However, aspect ratio threshold (l_{min}=*0.1* and l_{max}=*10)* are fixed empirically throughout the experiment. Now, candidate character regions are grouped using effective text localization method proposed by Yin et al. [5]. This technique constructs candidate regions with the use of geometric information and group similar candidate character region using disjoint sets. Also, vertical and horizontal variances, color, geometry and stroke width features of each candidate character regions are extracted to build Adaboost classifier for identification of text regions. In contrary to [5], we used Bayesian network based scores and Bayesian logistic regression classifier with additional adaptable and fixed constraints. The output text regions after grouping are presented in Figure 1.

3. EXPERIMENTAL RESULTS

The Robust Reading Competition was organized by ICDAR 2013 [6, 10, 17] with an objective to deliver performances of different algorithms as a benchmark. The evaluation system used by ICDAR 2013 demonstrated object level precision and recall of detecting algorithms according to their superiority of detection. Our text localization method used test dataset of ICDAR 2013 Robust Reading Competition Challenge 2 Task 1 [10]. The whole dataset consisted of 233 images containing text of different fonts and colors with different backgrounds.

3.1 Performance Evaluation of Text Localization System

ICDAR 2013 dataset was used in most recent text detection competition as a benchmark. To provide a baseline comparison, we tested our method on the publicly available dataset in [10]. The output of an algorithm is a set of rectangles designating bounding boxes for detected words in scene images and referred as *estimate* as given by Figure 1. A set of ground truth boxes is provided in the dataset which referred to *targets*. The match of an estimate and target by a text detecting algorithm can be defined as the area of intersection between two rectangles divided by the area of minimum bounding box containing both rectangles. The identical rectangles have value one and zero for rectangles with no intersection. Therefore, a closest match was found between the estimated and targeted rectangles to measure the performance. For this purpose, precision, recall and harmonic mean are used to recognize the matching of intersecting rectangles. Precision is the ratio of the number of matched detected rectangles to the total number of matched and unmatched detected rectangles. For example, M represents the matched detected rectangles and U represents the unmatched detected rectangles, Precision can be defined as given by equation 9.

$$Precision = \frac{M}{M+U} \rightarrow (9)$$

Similarly, recall can be defined as the matched detected rectangles to the total number targeted rectangles. For example, M are the matched and T are the targeted rectangles, recall can be defined as given by equation 10.

$$Recall = \frac{M}{M+T} \rightarrow (10)$$

Finally, F-measure can be defined as the harmonic mean of precision and recall and can represented as given by equation 11.

$$F-Measure = \frac{2}{\frac{1}{Precision}+\frac{1}{Recall}} \rightarrow (11)$$

Consequently, performances of our localization method with some of the published algorithms in ICDAR 2013 competition are presented in Table 1. Our text localization method shows a competitive recall *62.37*, precision *84.97* and a *71.94* harmonic mean, which is competitive with the leading methods reported by [10]. However, our text localization method performs better than the ICDAR 2011 Robust Reading Competition methods reported by [6].

Method	Recall	Precision	HMean
USTB_TexStar [27]	66.45	88.47	75.89
Text Spotter [20], [21], [22]	64.84	87.51	74.49
CASIA_NLPR [23], [24]	68.24	78.89	73.18
Text_Detector_CASIA [25], [26]	62.85	84.70	72.16
Our Method (LocaTeXT)	**62.37**	**84.97**	**71.94**
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 [18], [19]	53.42	74.15	62.10
Baseline	34.74	60.76	44.21
Inkam	35.27	31.20	33.11

TABLE 1: Performance (%) Comparison of Text Detection Methods on ICDAR 2013 Dataset.

4. REFERENCES

[1] 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, 20(3):800 –813, March 2011.

[2] Y.-F. Pan,, X. Hou, and C.-L. Liu. "A hybrid approach to detect and localize texts in natural scene images." Image Processing, IEEE Transactions on 20, no. 3 (2011): 800-813.

[3] J.-J. Lee, P.-H. Lee, S.-W. Lee, A. Yuille, and C. Koch. Adaboost for text detection in natural scene. In ICDAR 2011, pages 429 –434, September 2011.

[4] B. Epshtein, E. Ofek, and Y. Wexler. Detecting text in natural scenes with stroke width transform. In CVPR 2010, pages 2963 –2970, June 2010.

[5] 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." In Pattern Recognition (ICPR), 2012 21st International Conference on, pp. 725-728. IEEE, 2012.

[6] A. Shahab, F. Shafait, and A. Dengel. ICDAR 2011 robust reading competition challenge 2: Reading text in scene images. In ICDAR 2011, pages 1491 –1496, September 2011.

[7] 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.

[8] K. Iqbal, X.-C. Yin, H.-W. Hao, X. Yin, H. Ali, (2013), Classifier Comparison for MSER-Based Text Classification in Scene Images, In Neural Networks (IJCNN), The 2013 International Joint Conference on (pp. 1-6). IEEE.

[9] Hosmer, David W.; Lemeshow, Stanley (2000). “Applied Logistic Regression” (2nd ed.). Wiley. ISBN 0-471-35632-8.

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

[11] X.-C. Yin, H.-W. Hao, J. Sun, and S. Naoi. Robustvanishing point detection for MobileCam-based documents. In ICDAR 2011, pages 136 – 140, September 2011.

[12] C. Merino-Gracia, K. Lenc, M. Mirmehdi, A head-mounted device for recognizing text in natural scenes, Camera-Based Document Analysis and Recognition, pages 29--41, 2012.

[13] J. Matas, O. Chum, M. Urban, and T. Pajdla. Robust wide baseline stereo from maximally stable extremal regions. In British Machine Vision Conference 2002, volume 1, pages 384–393, 2002.

[14] 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, 65(1):43–72, November 2005.

[15] 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.

[16] N. Friedman, D. Koller, Being Bayesian About Network Structure: A Bayesian Approach to Structure Discovery in Bayesian Networks, Machine Learning 50 (1-2) (2003) 95-125.

[17] D. Karatzas, S. Robles Mestre, J. Mas, F. Nourbakhsh, P. Pratim Roy , "ICDAR 2011 Robust Reading Competition - Challenge 1: Reading Text in Born-Digital Images (Web and Email)", In Proc. 11th *International Conference of Document Analysis and Recognition*, 2011, IEEE CPS, pp. 1485-1490.

[18] J. Fabrizio, B. Marcotegui, and M. Cord, “Text segmentation in natural scenes using toggle-mapping,” in Proc. Int. Conf. on Image Processing,2009.

[19] J. Fabrizio, B. Marcotegui, M. Cord, “Text detection in street level images,” Pattern Analysis and Applications, 2013, Volume 16, Issue 4, pp 519-533.

[20] L. Neumann and J. Matas, “A method for text localization and recognition in real-world images,” in Proc. Asian Conf. on Computer Vision, 2010, pp. 2067–2078.

[21] L. Neumann, and J. Matas. Real-time scene text localization and recognition. in Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on. 2012., pp. 3538–3545.

[22] L Neumann, J Matas, “On combining multiple segmentations in scene text recognition,” in Proc. Int. Conf. on Document Analysis and Recognition, 2013. International Conference on Document Analysis and Recognition (ICDAR 2013), Washington D.C., 2013.

[23] Y.-M. Zhang, K.-Z. Huang, and C.-L. Liu, “Fast and robust graph-based transductive learning via minimum tree cut,” in Proc. Int. Conf. on Data Mining, 2011.

[24] C.-L. L. B. Bai, F. Yin, “Scene text localization using gradient local correlation,” in Proc. Int. Conf. on Document Analysis and Recognition, 2013.

[25] C. Shi, C. Wang, B. Xiao, Y. Zhang, S. Gao, and Z. Zhang, “Scene text recognition using part-based tree-structured character detections,” in Proc. Int. Conf. on Computer Vision and Pattern Recognition, 2013.

[26] 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.

[27] 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.