ZText: Zone Based Text Localization in Natural Scene Images

Khalid Iqbal[1†] , Hazrat Ali[2†] , Qazi Mudassar Ilyas[3], Irfan Mehmood[4], Hikmat Ullah Khan[5], Zahoor-Ur-Rehman[1]

[1]COMSATS Institute of Information Technology Attock,
[2]COMSATS Institute of Information Technology Abbottabad,
[3]College of Computer Sciences and Information Technology, King Faisal University, Saudi Arabia
[4]Department of Computer Science and Engineering, Sejong University, Seoul, Korea
[5]COMSATS Institute of Information Technology Wah

Summary
The evaluation of natural scene images for text localization is an appealing task to examine the image contents. In this paper, MSER-based candidate character regions are initially compared with the geometric features based effective text localization method based to find text regions. In addition, zone-based features of MSER-based extracted complementary candidate characters are computed for respective zones including the regional features. Bayesian logistic regression classifier is trained on features complementary candidate characters. The complementary candidate character regions with higher posterior probability are considered as candidate characters or letters corresponding to non-candidate characters or letters. Adjacent complementary candidate characters with higher posterior probabilities are grouped into words and sentences. Consequently, zone-based text localization, named as ZText, is evaluated on ICDAR 2015 Robust Reading Competition benchmark dataset. The results of experiments have established amazing competitive performance with the recently published text localization algorithms.

Key words:
Zone-based Features, MSER, Classification, Bayesian Logistic Regression.

1. Introduction

A rapid increase of visually impaired persons causes by different diseases in eye and diabetes. These diseases may cause of traffic accidents on roads and streets. A technology based communication tool can provide a support to visually impaired people. In this perspective, Text localization in scene images is an appealing and complicated task for finding a technology based solution in the field of computer vision. Text extraction in video sequencing [2], automatic forms reading [3], postal address box [4], text in license plates [5], document image analysis [6], worldwide systems for document analysis [7] and image content image indexing [8] to help visually impaired people on streets sign boards [9]. With such dynamic background of natural scene images, text localization is a demanding task and drawing attention of researchers to find a practical solution to improve the poor performance of OCR engines [10] on color images with variation in color, font size, style and alignment orientation.

Text localization approaches can roughly be classified in four categories: edge based techniques, Morphological based techniques, Texture based techniques, connected component based techniques, and region based methods. First, edge based methods efficiently and effectively extracts text from documents and images. These methods considered edge strength, change in orientation and density. However, the accuracy of detected is not as good as required in the aforementioned computer vision fields. Second, geometrical structures extractions from images for shapes representation are known as morphological approaches. A morphology based feature can efficiently be extracted for character recognition and document image content. Third, texture-based techniques observed the distinct textual characteristics to identify text regions. Textual characteristics of text region can be detected by Gabor filter, FFT and Wavelet in a complex way. Fourth, CC-based methods extracted components from images as candidate characters and grouped according to similar geometric properties by eliminating non-character candidates with the help of constraints. Epshtem et al [10] calculated stroke width for every pixel. Neighboring pixels with similar stroke width are clustered as CCs. Pan et al. [11] exploited image candidate regions to localize text candidates by extracting CCs as candidate characters using local Binarization. Conditional random field (CRF) model is used to eliminate false positive components. To conclude, candidate components are grouped into lines and words. Lastly, Region-based approaches investigated images using sliding window to find features of a candidate region. Candidate region features are then classified to locate text regions. In order to recognize text regions, Lee et al. [12] extracted 6 different classes of regions as features. Modest Adaboost method is used to detect text regions. Similarly, Yin et al. [1] used MSER [13] to extract candidate letters. Non-candidate letters are eliminated with the help of geometric features. Candidate letters with similar geometry

Manuscript received April 5, 2017
Manuscript revised April 20, 2017

are set into text candidate regions by disjoint set. A limited number of features were extracted to train Adaboost for detecting image text. However, independent features of letters can help to improve text detection in natural scene images. Thus, features of candidate letters are confined to their geometric shapes. Nonetheless, candidate character regions (letter candidates) extracted through MSER can have enough number of pixels for textual content in combination with geometric features [1]. To find a sufficient amount features, zoning can be performed on candidate characters as well as to train Bayesian logistic regression for eliminating complementary non-candidate characters. The higher posterior probabilities of a complementary candidate character of [1] are considered as region of interest in contrary to Yin et al. [14]. Candidate selection rules [15], perceptual text grouping filters [9] and repulsion scores [16] are used to detect text regions rather than using single-link clustering algorithm. In this context, robust feature search is still a challenge of MSER-based extracted candidate letters.

In this paper, we propose a hybrid system based on zones features and geometric features of MSER-based text detection method for robustness and accuracy. First, MSER-based extracted candidate characters are compared with [1] for text regions based on disjoint set. However, a direct comparison of extracted MSERs ignored various candidate characters that can be grouped into text region. Second, selection rules in comparison with text regions in [1] are used to enhance accuracy. Despite this, other text regions can be found from complementary candidate regions ignored by Yin et al. [1]. Therefore, Bayesian logistic regression classifier is trained on the zones features of complementary set character regions. Now, perceptual text grouping, selection rules and repulsion score of adjacent complementary candidate character region are used to group them into words. Consequently, we build an accurate text localization system in natural scene images. Our proposed method is evaluated on ICDAR 2015 Robust Reading Competition (Challenge 2: “Focused Scene Text”, Task 2.1: Text Localization) benchmark database and has achieved an f-measure of 72.60% which is a significant achievement.

The remaining paper is organized as follows. Latest MSER-based methods for natural scene text localization are reviewed in Section 2. Section 3 explains the proposed method, named as ZText. Experimental results are presented in section 4. Lastly, the paper has been ended up with a conclusion.

2. Related Works

MSER has been integrated in a number of real-world projects for significant performance. Still, MSER-based text detecting techniques consist of repeating components and inadequate text regions. This fundamental limitation is a major challenge while incorporating MSER for text detection in street scene images. Instead of this, MSER can detect maximum number of candidate characters in quality compromised images having strong noise, low contrast and resolution. Likewise, MSER permits to detect features of candidate regions. Therefore, repeating components detection is minor problem by MSER-based methods and can be adjusted while aligning candidates. An new MSER-based approach is suggested by Yin et al. [1] to retrieve natural scene text. Candidate text regions are constructed and then classified by training Adaboost on features like horizontal and vertical variation, stroke width, color and geometry. Yin et al. [14] pruned MSERs by minimizing regularized variation. Candidate characters are grouped into text candidates by using single link clustering by learning the self-training distance. Text candidates with higher posterior probabilities are removed to recognize text regions. Fabrizio et al. [15] validated a hypothesis by segmentation in conjunction with SVM classifier and candidate characters are grouped using selection rules. Meri-Gracia et al. [9] focused on MSERs as candidate text region in a hierarchical relationship and are grouped with the use of cascaded text classification filters. Pan et al. [16] developed conditional random field (CRF) to exploit visualize context for valid candidate’s pair wise detection and with repulsion link.

The aforementioned MSER-based techniques have not enough text regions based on a subset of features of candidates. In contrary to [14] and in conjunction with [1, 9, 15, 16], zoning can be performed to obtain features in large quantity for candidate character region. The key advantage of zoning is to find an adequate amount of features of different parts of a given candidate character region. In addition, zoning can perform better than the use of Bayesian network scores for text localization. Consequently, zoning can improve the detection quality of text in scene images in conjunction with Yin et al. [1]. In contrast to Yin et al. [14], high posterior probability candidate characters regions are considered for grouping with the help of some perceptual grouping filters [9], repulsion score [16] and selection rules [15].

3. ZText: Proposed Method

In this section, we explain the proposed ZText (i.e. Zone features of candidate characters based Text detection method).

3.1 Extracting and Filtering Candidate Character Regions

The images taken by digital camera may have deformity [17] that motivated us to detect text in natural scene images. ZText uses MSER [13] to extract candidate character regions from scene images. MSER catches matching in images for retrieval. MSER can detect region of interest even in scale and light variation [18] by marking their locations. Therefore, minimum and maximum of located points are used to extract interesting regions. Each extracted MSER is resized into (w x w) window to perform binarization for zoning process using an adaptive threshold ƛ which is average of k resized candidate character regions having window size w=25. Candidate Binary regions are obtained by performing binarization on each window. Candidate binary regions are retained if total number of pixels ≥30 and ≤600. The purpose of applying this constraint is to filter a CBR with too high or too low density of pixels.

3.2 Zone-based Features of Candidate Characters

First of all, MSER-based candidate regions (as shown in Figure 1) are resized into window with a w x w size (i.e. w=25). To perform zoning process, a resized candidate character region is divided into 9 sub regions and known as zones given by Figure 2. Features are extracted of each zone rather than the intact candidate character region. The benefit of using zoning is to obtain fine details as well as positions of line segments of a given candidate character region. To extract line segments of a zone, the whole skeleton of candidate character region can be traversed according to each zone. Therefore, pixels defined of candidate character skeleton in a given zone can be defined as starters, minor starters and intersections.

Considering the resized binary image as shown in Figure 1(b), character traversal starts to perform zoning on the candidate binary character region. To traverse candidate character, line segments are extracted for each zone to populate a list. Starters and intersections are found in a given zone of a candidate character. Minor starters are created along the traversal path. Minor starters are processed after the processing of all starters. All the line segments and their pixel positions are stored. After visiting all the pixels of a given candidate character, feature extraction algorithm stops.

Now, line segments are identified and classified into horizontal and vertical line, and left and right diagonal lines by extracting a direction vector [19] for determining its type. After determining the type of line segment, a feature vector is formed corresponding to each zone of a given candidate character region. Each zone feature vector consisted of horizontal line, vertical line, left and right diagonal lines, normalized length of horizontal, vertical, left and right diagonal lines and normalized area of skeleton. The length and number of any particular line type are normalized using Equation (1) and (2) respectively.

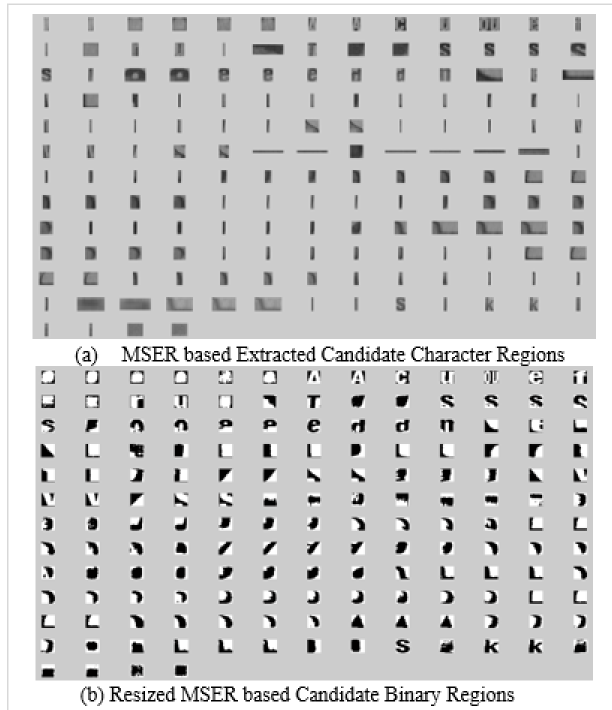

Fig. 1 MSER based Candidate Binary Regions

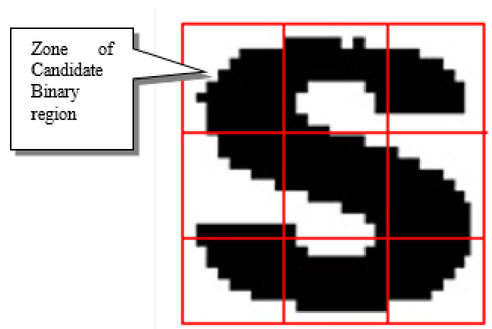

Fig. 2 9 Zones of Candidate Binary Region

$$l = T_{p/l} / T_{zp} \quad (1)$$

$$v = 1 - 2N/10 \quad (2)$$

Where, l is the length of any particular line, $T_{p/l}$ are the total pixels in a given line type and T_{zp} are the total zone pixels. Similarly, v is the number of a given particular line; N is the number of line.

The candidate binary character region is splitted into 9 sub regions with an aim to extract features of each zone. After extraction of zone features, features of the entire candidate binary character region are extracted based on regional properties such as *Euler value*, *Region area* and *Eccentricity*. First, *Euler value* is the dissimilarity between the number of objects and the number of holes in the candidate binary character region. Second, *Region Area* can be defined as the fraction of the number of pixels in the skeleton and the total number of pixels in the candidate binary character region. Lastly, *Eccentricity* is defined as the odd behavior of the smallest ellipse that meets the required skeleton of the candidate binary character region.

3.3 Text Regions Construction with Selection Rules

For enhancing accuracy, MSER-based extracted candidate character regions are compared with [1]. Each candidate character region has four corner points. Text regions are formed after comparing these 4 points with [1] and presented in Figure 3. Besides Figure 3 result, a large number of candidate characters are missed by [1] while constructing text regions. The missed candidate characters are inspected for their adjacency according to selection rules [15]. The selection rules, formulated through

Fig. 3 Text Regions formed using Yin et al. method [1]

Figure 4, can be utilized to group missed candidate characters to enhance accuracy without affecting the existing results. Slight improvement in text regions can be observed. Despite this, close-fitting constraints cannot recognize maximum possible text regions. Therefore, zoning is performed to learn features for the missed candidate characters for better results.

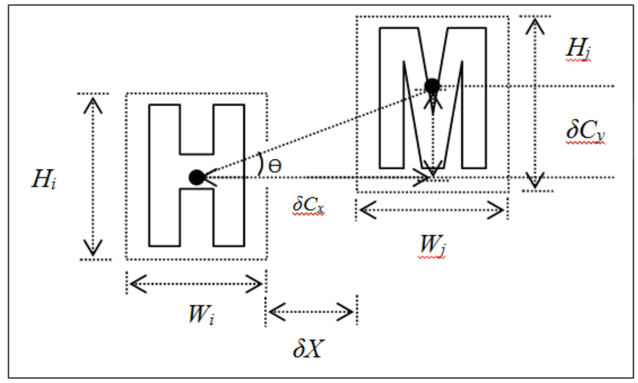

Fig. 4 Adjacent Candidate Characters Regions Geometry for Selection Rules

Candidate character regions can be grouped according to their angle, vicinity, horizontal alignment and height similarity constraints. Edge angle can be used to align characters horizontally (i.e. $a=\tan^{-1}(\delta Cy/\delta Cx)<30°$). With the use of angle, alignment can further be ensured by applying a constraint, derived from Figure 4, (i.e. $Align=\delta Cy/\min(H_i,H_j)<0.3$). Similarly from Figure 4, height similarity and vicinity constraints are applied for grouping candidates characters (i.e. $h_s=|H_i - H_j|/\min(H_i, H_j)<0.3$, $V=\delta X/\min(H_i, H_j)<0.5$).

3.4 Zoning based Features of Complementary Candidate Characters

To the best of our knowledge, zoning has not yet been considered by the academia for text detection. Therefore, zoning based features of complementary candidate character regions are learnt. The reason behind learning zones complementary candidate character features is to overcome the limitation of poor efficiency. Besides, rejection of too wide or too narrow regions is eliminated by using aspect ratio and computed by Equation (3) with the use of threshold values ($t_h=0.3$, $t_w=0.1$, $T_h=1.15$, $T_w=5$) in our zoning experiments.

$$t_h < H_i/W_i < T_k \text{ and } t_w < W_i/H_i < T_w \tag{3}$$

Besides, zoning based extracted features as well as whole candidate character region features can be used for data analysis. First, a candidate character region is divided into 9 zones to compute 81 features. Each zone height and width is computed by dividing the total number of rows and total number of columns by 3 of a given candidate character region. A feature vector of a given zone is formed on a number and type of line segments as explained in Section 3. Second, the entire candidate character region features are extracted to measure the regional properties including Euler number, eccentricity, orientation and extent. Consequently, feature vector formed is based on 81 candidate based zones with 4 regional features. The following are the steps to find a feature vector for a given candidate binary region.

3.5 Classification of Complementary Zone-based Features

Bayesian logistic regression classifier can be used to keep away from over-fitting problem with great accuracy [20, 21]. The objective of using this classifier is to model a relationship between complementary candidate characters zone and regional features corresponding to their class according to Bayes rule. The essential use of Bayesian logistic regression is its great accuracy [22] while figuring out a probabilistic relationship between feature vector and class labels of complementary candidate character regions. Let $T=\{z_1,z_2,\ldots,z_{81},r_1,r_2,r_3,r_4,L\}$ be the test set composed of *81* zones and *4* regional features of complementary candidate binary regions (C) with class labels $L=\{0,1\}$. The log-likelihood ratio $ln(L=1/T,\omega)/P(L=0/T,\omega))$ is assumed to be linear in C. The conditional likelihood for $L=1$ can be computed by the use of sigmoid function $P(L=1/T,\omega) = 1-1/\left(1+e^{-\omega^t T}\right) = \sigma\left(-\omega^t C\right)$. In the same way, $P(L=0/T,\omega) = 1-P(L=1/T,\omega) = 1/\left(1+e^{\omega^t T}\right)$, such that $P(L/T,\omega)=\sigma\omega^t T$. Posterior probability complementary candidate characters can be computed by conditional probability and prior probability with the help of conjugate-Gaussian prior including hyper-parameters. Equations (4) and (5) can be used to approximate and model the conjugate Gamma distribution.

$$P(\omega/\alpha) = (\alpha/2\pi)^{\frac{1}{2}} e^{\frac{\alpha}{2}\omega^t\omega} \tag{4}$$

$$\mathrm{P}(\alpha) = \left(1/\Gamma a_0\right) b_0^{a_0} \alpha^{a_0-1} e^{-b_0\alpha} \tag{5}$$

Where ω: conjugate-Gaussian prior, α: hyper-parameter, a_0 and b_0 are hyper prior parameters.

Fig. 5 Classified complementary Candidate Character Regions (red) and Text Regions (green)

Classified complementary candidate regions are considered for further grouping with posterior probability greater than or equal to 0.5. The color output of classified candidate character regions are shown in Figure 5 (*red*) while text regions after a direct comparison with the use Yin et al. method [1] and selection rules are shown in Figure 5 (*green*).

3.6 Complementary Candidate Characters Grouping

For complementary classified candidate characters, h_s and *Align* measures are adapted to eliminate the non-characters. The ranges for h_s and *Align* vary from *70%* to *90%* and *5%* to *15%* respectively. Despite this, there are a huge number of non-characters in complementary classified candidates. Therefore, horizontal centroid difference and vertical centroid difference are considered for grouping candidate characters to form text regions through Equation (6) and (7) respectively.

$$0.1\max\left(\frac{H_i}{2},\frac{H_j}{2}\right) < \left|C_{xi}-C_{xj}\right| < 2\max\left(\frac{H_i}{2},\frac{H_j}{2}\right) \tag{6}$$

$$\left|C_{yi}-C_{yj}\right| < 0.1\max\left(\frac{W_i}{2},\frac{W_j}{2}\right) \tag{7}$$

Furthermore, repulsion score of classified complementary candidate regions can be used to measure the degree (0.7 or higher) for grouping and calculated by Equation (8).

$$RS = 1 - min\left(1,\frac{\left\|C_i-C_j\right\|}{max\left(D_i,D_j\right)}\right) \tag{8}$$

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

Fig. 6 ZText Enhancements in Text Localization over Yin et al. [1]

After grouping, connected components are used to label complementary classified candidate character regions through 8-neighbour metric, with at least a sum of pixel 8. In this way, enhancements of additional text detected region are found as shown in Figure 6.

4. Experimental Results

An ICDAR 2015 Robust Reading Competition (Challenge 2: “Focused Scene Text”, Task 2.1: Text Localization) database [23] is used as a benchmark dataset for experiments and evaluation of text detection algorithms. The ICDAR 2015 database contains 233 test set images. Zoning process with regional features is applied on MSER-based extracted candidate characters and classifier [20, 21] is trained to improve performance. To measure the performance of ZText, text regions are constructed through [1] and additional text regions detected from complementary candidate character regions are combined with an aim to get the best performance. Thus, the complementary candidate characters are grouped to localize text by outlining the words and sentences using selection rules including horizontal and vertical centroid difference, and repulsion score. The localized text regions, in this way, are shown in Figure 7.

An ICDAR 2015 Robust Reading Competition (Challenge 2: “Focused Scene Text”, Task 2.1: Text Localization) database [23] is used as a benchmark dataset for experiments and evaluation of text detection algorithms. The ICDAR 2015 database contains 233 test set images. Zoning process with regional features is applied on MSER-based extracted candidate characters and classifier [20, 21] is trained to improve performance. To measure the performance of ZText, text regions are constructed through [1] and additional text regions detected from complementary candidate character regions are combined with an aim to get the best performance. Thus, the complementary candidate characters are grouped to localize text by outlining the words and sentences using selection rules including horizontal and vertical centroid difference, and repulsion score. The localized text regions, in this way, are shown in Figure 8.

(a) ZText Enhancements in Text Detection by ZText over Yin et al. [1]

(b) ZText and Yin et al. [1]Text Detection

Fig. 7 Text Detection by ZText in Natural Scene Images

Table 1. Performance Comparison of ZText Method

Method	*Recall*	*Precision*	*F-Measure*
Proposed Method ZText	**0.6351**	**0.8473**	**0.7260**
BayesText [22]	0.6351	0.8430	0.7244
Text_Detector_CASIA [30,31]	0.6285	0.8470	0.7216
I2R NUS FAR	0.69	0.7508	0.7191
I2R NUS	0.6617	0.7254	0.6921

TH-TextLoc	0.6519	0.6996	0.6749
Text Detection [15, 32]	0.5342	0.7415	0.6210
Baseline	0.3474	0.6076	0.4421
Inkam	0.3527	0.3120	0.3311

After showing the detected text in Figure 8, ZText is also evaluated in comparison with the recently published state-of-the-art text detection methods. According to the evaluation procedure of ICDAR 2015 for ICDAR 2013, all images are provided with the ground truths for each scene image text as words. The detection is evaluated through precision, recall and f-measure. First, precision is the fraction of correctly classified complementary text windows of ZText and [1]. Second, recall is the fraction of all correctly indentified text windows of ZText and [1]. Lastly, f-measure is the harmonic mean of precision and recall. The optimal performance of precision, recall and f-measure should be unity. However, there are certain problems in obtaining the appropriate evaluation. The problems in rational evaluation includes the inconsistency in ground truths, granularity of annotation and bumpy weights for text regions by taking average of recall and precision of each image [15].

We evaluated ZText on ICDAR 2015 benchmark dataset [23] and experimental results are presented in Table 1. ZText with zones and candidate regional features resulted an f-measure of 72.60% which is an improved value as compared to the recently published BayesText approach [22]. Figure 8 illustrates several successful results on ICDAR 2015 database scene images. Figure 8(a) gives a performance improvement over Yin et al. method [1]. Figure 8(b) reflects the combined performance of [1] and zoning based features text localized results in scene images.

5. Conclusion

In this paper, MSER-based extracted complementary candidate character regions are splitted into 9 zones with the identification of features of respective zones including regional features in conjunction with effective text localization method based on geometric features. A text classifier (Bayesian Logistic Regression) is trained on complementary candidate characters using zone features and regional features. Consequently, we build an effective scene text localization ZText method that shows a significant relative performance with the latest state-of-the-art techniques. Besides this, ZText can further be improved by the recognition of more robust features of MSER-based extracted complementary candidate character regions. Therefore, our future target can be to train a new classifier based on zone based features as well as to add some more features to enhance performance of ZText to retrieve information from natural scene images.

References

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

[2] Wolf, C., Jolion, J., and Chassaing, F., "Text localization, enhancement and binarization in multimedia documents," in *Pattern Recognition, 2002. Proceedings. 16th International Conference on*, 2002, pp. 1037-1040.

[3] Kavallieratou, E., Balcan, D. C., Popa, M. F., and Fakotakis, N., "Handwritten text localization in skewed documents," in *Image Processing, 2001. Proceedings. 2001 International Conference on*, 2001, pp. 1102-1105.

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

[5] Arth, C., Limberger, F., and Bischof, H., "Real-time license plate recognition on an embedded DSP-platform," in *Computer Vision and Pattern Recognition, 2007. CVPR'07. IEEE Conference on*, 2007, pp. 1-8.

[6] Xiao, Y. and Yan, H., "Text region extraction in a document image based on the Delaunay tessellation," *Pattern recognition*, vol. 36, pp. 799-809, 2003.

[7] Wahl, F. M., Wong, K. Y., and Casey, R. G., "Block segmentation and text extraction in mixed text/image documents," *Computer graphics and image processing*, vol. 20, pp. 375-390, 1982.

[8] Lienhart, R. and Effelsberg, W., "Automatic text segmentation and text recognition for video indexing," *Multimedia systems*, vol. 8, pp. 69-81, 2000.

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

[10] Epshtein, B., Ofek, E., and Wexler, Y. , "Detecting text in natural scenes with stroke width transform," in *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*, 2010, pp. 2963-2970.

[11] Pan, Y. F., Hou, X., and Liu, C. L., "A hybrid approach to detect and localize texts in natural scene images," *Image Processing, IEEE Transactions on*, vol. 20, pp. 800-813, 2011.

[12] Lee, J. J., Lee, P. H., Lee, S. W., Yuille, A. L., and Koch, C., "AdaBoost for Text Detection in Natural Scene," in *ICDAR*, 2011, pp. 429-434.

[13] Matas, J., Chum, O., Urban, M., and Pajdla, T., "Robust wide-baseline stereo from maximally stable extremal regions," *Image and vision computing*, vol. 22, pp. 761-767, 2004.

[14] Yin, X., Huang, K., and Hao, H., "Robust text detection in natural scene images," 2013.

[15] Fabrizio, J., Marcotegui, B., and Cord, M., "Text detection in street level images," *Pattern Analysis and Applications*, vol. 16, pp. 519-533, 2013.

[16] Pan, J., Chen, Y., Anderson, B., Berkhin, P., and Kanade, T., "Effectively leveraging visual context to detect texts in natural scenes," in *Proc. Asian Conf. Comp. Vis*, 2012.
[17] Gaurav, D. D. and Ramesh, R., "A feature extraction technique based on character geometry for character recognition," *arXiv preprint arXiv:1202.3884,* 2012.
[18] Yin, X. C., Hao, H. W., Sun, J., and Naoi, S., "Robust vanishing point detection for MobileCam-based documents," in *Document Analysis and Recognition (ICDAR), 2011 International Conference on*, 2011, pp. 136-140.
[19] K. Mikolajczyk, *et al.*, "A comparison of affine region detectors," *International journal of computer vision*, vol. 65, pp. 43-72, 2005.
[20] Iqbal, K., Yin, X. C., Yin, X., Ali, H., and Hao, H. W., "Classifier comparison for MSER-based text classification in scene images," in *Neural Networks (IJCNN), The 2013 International Joint Conference on*, 2013, pp. 1-6.
[21] Hosmer Jr, D. W., Lemeshow, S., and Sturdivant, R. X., *Applied logistic regression*: John Wiley and Sons, 2013.
[22] Iqbal, K., Yin, X. C., Hao, H. W., Asghar, S., and Ali, H., "Bayesian network scores based text localization in scene images," in *Neural Networks (IJCNN), 2014 International Joint Conference on*, 2014, pp. 2218-2225.
[23] Karatzas, D., Shafait, F., Uchida, S., Iwamura, M., Mestre, S. R., Mas, J., Mota, D. F., Almazan, J. A., and de las Heras, L.P., "ICDAR 2013 robust reading competition," in *Document Analysis and Recognition (ICDAR), 2013 12th International Conference on*, 2013, pp. 1484-1493.

Khalid Iqbal received his PhD degree in Applied Computer Technology from University of Science and Technology Beijing in 2014, and the B.Sc., and MS(CS) degrees from University of the Punjab, Lahore and SZABIST Karachi respectively. He was awarded a fully funded scholarship by the China Scholarship Council for the entire duration of his PhD studies. He also won the excellent researcher (or excellent international research student award) from University of Science and Technology Beijing in the year 2012-13. He is currently an Assistant Professor in the Department of Computer Science, COMSATS Institute of Information Technology, Attock Campus. He is also the director of Pattern Recognition, Images and Data Engineering (PRIDE) research group where graduate students are working under his supervision in different fields such as Data Mining, Social Networks, Image Processing and simulation based Models. He has worked on Bayesian network application for Privacy Preserving of XML association rules and Text Localization in Scene Images. His research work has been published in several international conference proceedings and journals. His research interests include Pattern recognition, Machine Leaning, Data Mining and Social Networks. He is the recipient of the CSC scholarship and QCRI/Boeing Travel grant.

Hazrat Ali is Assistant Professor at Department of Electrical Engineering, COMSATS Institute of Information Technology Abbottabad. He completed his Ph.D. degree in Communication and Information Systems from the University of Science and Technology Beijing, Beijing, China. Earlier, he received his B.S.

and M.S. degrees from the University of Engineering and Technology, Peshawar, Peshawar, Pakistan, in 2009 and 2012, respectively. He was a Visiting Researcher at City University, London, UK, for a period of 12 months. From 2010 to 2012, he worked as a Faculty Member at the Department of Electrical Engineering, University of Engineering and Technology, Peshawar, Peshawar, Pakistan. He has served as a Reviewer for the International Journal of Artificial Intelligence Tools, Reviewer for Springer Multimedia Tools and Applications, sub-reviewer for AAAI Conference on Weblogs and Social Media, Review Committee co-chair and co-organizer for City Informatics PhD Symposium (CIPS 2014). He is a PC Member of the Workshop on Security and Privacy in Cybermatics (SPiCy 2015). He has published several papers in leading journals and conferences and has presented his research work in Netherlands, Qatar, UK, and China. For his excellent academic performance, he was awarded the USTB Excellent International Student Award (2nd Prize) for the session 2013/2014. He also received HEC Scholarship, IEEE EMBS Travel grant (2015), TERENA/CISCO Travel grant, and QCRI/Boeing Travel grant (2014, 2015, 2016 and 2017).

Qazi Mudassar Ilyas is currently working as an Assistant Professor at King Faisal University, Saudi Arabia. He received his Bachelors and Masters degrees from University of Agriculture,
Faisalabad, Pakistan and PhD degree from Huazhong University of Science and Technology, China. His research interests include Semantic Web, Knowledge Engineering and Machine Learning.

Irfan Mehmood received BS degree in Computer Science in 2010 from National University of Computer and Emerging Sciences, Pakistan. From 2010 to 2011, he served as a Software Engineer in Talented Earth Organization, where he provided various services such a mobile application development. At the end of 2011, he joined Intelligent Media Laboratory (IM Lab) as a research associate while enrolling as a PhD student in Sejong University, Seoul, South Korea. In IM Lab, he worked on various projects related to video summarization and prioritization, image super-resolution, image quality assessment, and mixed reality. In 2016, he joined Sejong University, South Korea and serving as an assistant professor in computer science and engineering department. He is also head of the Visual Analytics Lab, where students are working on research projects such social data mining and information retrieval, steganography, digital watermarking, and medical imaging analysis etc. under his supervision. His current research activities include social data mining, multi-modal medical data analysis, information summarization and multi-modal surveillance frameworks.

Hikmat Ullah Khan is serving as an Assistant Professor in Department of Computer Science, COMSATS Institute of Information Technology, Wah Campus. He has received his PhD degree from Computer Science from International Islamic University, Islamabad. He is a member of the technical committee of a number of top IEEE International conferences and regular reviewer of various Impact Factor Journals.His research interests include Social Web Mining, Social Network Analysis and Mining, Semantic Web, Information Retrieval and Islamic Applications.

Zahoor-ur-Rehman has experience in academia, software development, project management, open source software development and promotion. He has received his educational and academic training at university of Peshawar, Foundation University Islamabad and UET Lahore, Pakistan. Dr. Rehman earned PhD in Computer Science from University of Engineering and Technology Lahore in year 2014 with focus on data stream mining, context-awareness and ubiquitous data mining. During the period 2008-2012, he was the part of Al-Khawarizmi Institute of Computer Science (KICS) UET Lahore (CS research center) where he gained essential analytical, developmental, data mining and machine learning, and process thinking skills. Dr. Rehman joined COMSATS institute of information technology as assistant professor in the early 2015. He is responsible to teach both at graduate and undergraduate levels. Along with teaching responsibilities, he is an active researcher and developing research proposals and also producing research articles. He is an active researcher and has authored/co-authored many research articles that are indexed in Scopus and ISI web of knowledge and published in prominent journals and well reputed national and international conferences. Being in the reviewer of panel of several international journals and IEEE conferences is an honor for him.