Global Scientific JOURNALS

GSJ: Volume 8, Issue 4, April 2020, Online: ISSN 2320-9186
www.globalscientificjournal.com

CHALLENGES OF ENGLISH TEXT RECOGNITION FROM NATURAL SCENES

[1]Muhammad Ahmed Zaki, [2]Urooba Zaki

[1]M. E Scholar, Mehran University of Engineering and Technology, Jamshoro, Department of Computer System Engineering, Pakistan.
[2]M. Phil Scholar, University of Sindh, Jamshoro, Institute of Information and Communication Technology, Pakistan

*Corresponding Author

E-mail: [1*]ahmed_zaki64532@yahoo.com, [2]zurooba6@yahoo.com

Keywords

Recognition Challenges, Character shapes, Complex Background, Lighting Exposure, Android App Development, Image processing, Tesseract, Artificial Intelligence.

ABSTRACT

Reading text in natural scenes called scene text recognition (STR), was an important task. The maturity of Optical Character Recognition (OCR) systems led to its successful application on cleaned documents, but STR tasks were not carried out by most traditional OCR methods because of the various text appearances in the real world and because of these scenes are captured in imperfect conditions. Photographs and digital multimedia texts are the best communication way in the universe. However, text in an item provides details regarding the world in a concise and attractive format, such as warning, traffic signals, company contents, etc. Text is any material can help us to understand more quickly the intent. Detection of text is quite a challenging issue especially for the pictures comprising of the natural scene because of multiple differences and uncontrollable variables as opposed to scanned record text detection. This paper reflects various issues and obstacles which may include multi-script text on an image, various font text with text-size variations, sun, dark, reflection, text color, background text, blur, low resolution, skew, and neon signboard. Furthermore, this study detects the text from natural scenes with the help of an Optical Character Recognition (OCR) algorithm and also shows the challenging images results faced during the complete process of detection.

1. INTRODUCTION

The identification of the image text began as a research area of artificial intelligence and perception of the machine [1]. The track-ing and recognition of scene text in natural images of unconstrained settings is still a computer vision challenge. Text in scenes offers highly realistic, semantic information, such as indoor navigation, outdoor navigation, content-based picture retrieval [2]. The ability to read text robustly in unconstrained scenes can help significantly with numerous real-world applications, e.g., visually impaired assistive technology, robot navigation, and geo-location [3]. Text details in an image can be used in a wide range of devices such as automated translation, scenario texts, and assistive interpretation for people with visual impairments [4]. Because of a number of applications, several researchers in natural scene images have given a considerable amount of attention to this issue [5]. For years, reliable and rigorous identification of machine-based scene texts has been a research problem, primarily owing to the vast number of variations in language, complicated picture context, scenery items, and many more [6]. Several other variable factors including text scale, design, and different contexts and lighting variance [7], [8].

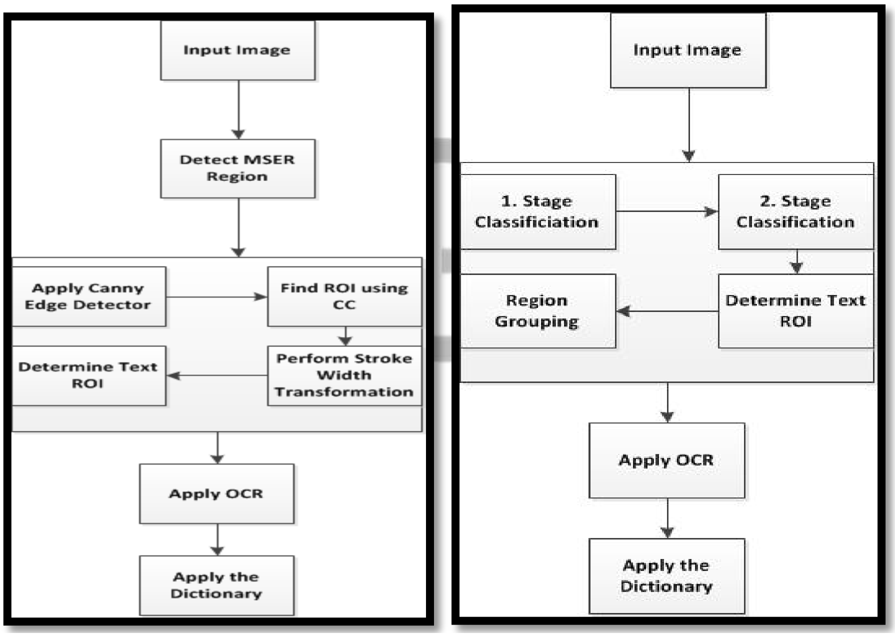
This study includes a brief analysis of the content recovery method and offers a comprehensive description of the issues and difficulties confronting nature data retrieval systems. The analysis is divided into different sections. Section 2, presents the related work. Section 3 provides a detailed description of Issues and challenges for text detection and Section 4 discusses the frame-work of the text recognition system. The conclusion is summarized in Section 5.

2. RELATED WORK

Baran et al. [9] proposed natural scenery images in a modern and efficient way to automatically identify text and characters. The Maximally Stable Extremal Regions (MSER) feature was used for the text detection process. Later on, a variety of filters were added to the region of interest (ROI) for image segmentation. The words are recognized using the OCR device. Finally, the App has been presented to clarify the perspective of Internet-Mediated Communities of Practice (IMCOP) briefly label content detection and delivery stage.

Guzel [10] presents a natural scenario text recognition method utilizing two separate Maximally Stable Extremal Regions (MSER) and Class Specific Extremal Regions (CSER) methods. The machine takes first photos with the aid of MSER to identify an area of interest and then receives text which was transferred to the OCR engine for further processing as shown in Figure 1. The program then defines, classifies, and determines the area of significance through the CSER and then uses the last one to recognize through OCR. The findings demonstrate that the text area in natural scenes can be best identified by the CSER shown in Figure 1,2.

Huang et al. [11] recommend a new Convolution Neural Network (CNN) Text Detection System. Advanced apps can be trained to recognize text through the CNN network. In this method, sliding and MSER models are used. MSER operator decreases the scanning amount of windows and improves weak text identification. The sliding CNN window was correctly applied to distinguish the character connections in the variable. In the International Conference on Document Analysis and Recognition (ICDAR) 2011 bench-mark data collection, it is reported that the device has a function score of more than 78%.



(a) (b)

Figure 1(a) MSER-based signboard detection system (b) CSER-based signboard detection system [10]

Mishra et al. [12] present a strongly related histogram of oriented gradients (HOG)-based end-to-end approach and a Support Vector Machine (SVM) classifier. Patch-based approaches offer strong outcomes in-text localization even though they do not answer the problem of the text segmentation explicitly (separation of texts from the background).

Tian et al. [13] achieve the state of the art precision using a hybrid approach where a histogram of oriented gradients (HOG) features classification produces a text trust chart that offers an extractive area local binarization algorithm.

Neumann et al. [14] propose MSER-derived area representation where the character/noncharacter levels listed for increasing the Extreme zone that could be feasible. Within this group, other techniques are used in areas derived from the gradient of the picture edge or the hue.

3. ISSUES AND CHALENGES FOR ENGLISH TEXT DETECTION

For the recognition and analysis of content, landscape images (natural photos) are essential [15]. The algorithm for the recognition of printed text does not often succeed for either the recognition of natural scene text [16]. It's because of issues such as poor quality, natural lighting, dim light, and much more [17]. Words may often have different font sizes, patterns, and colors due to these characteristics, it is difficult to get text and the current OCR algorithm is not sponsored [18].

Text recognition is useful for image analysis, but it is challenging to recognize because of different reasons such as light intensity, text on buildings, and walls [19]. The importance of image processing and recovery framework can be improved by emerging smart app technology [20]. The text is scattered over scenes such as the name of the area, street signs, store names, banners, etc. This text is a significant predictor of image quality comprehension and supplying valuable and relevant scene knowledge [8]. The following subsections address the difficulties of text identification in greater depth Shown in Figure 2.

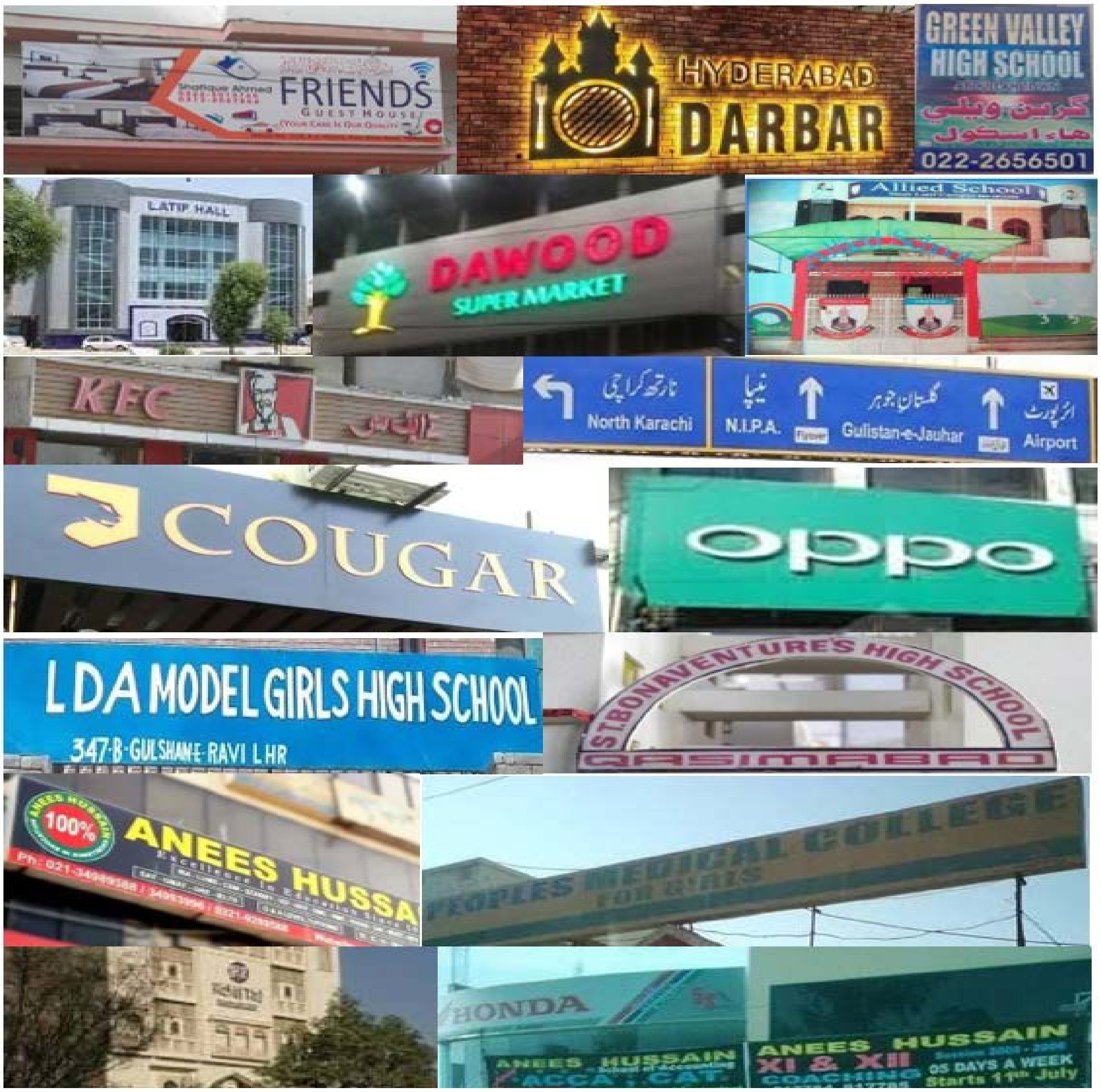


Figure 2 Text from Natural Scenes

3.1 TEXT IN MULTI SCRIPT

Images from the natural scene can have a variety of text. Images can sometimes have text and text boxes of other scripts and symbols, including numbers, letters, and words [21]. The text can be called a multilingual script of this type of image [22]. Figure 3 shows text diversity.

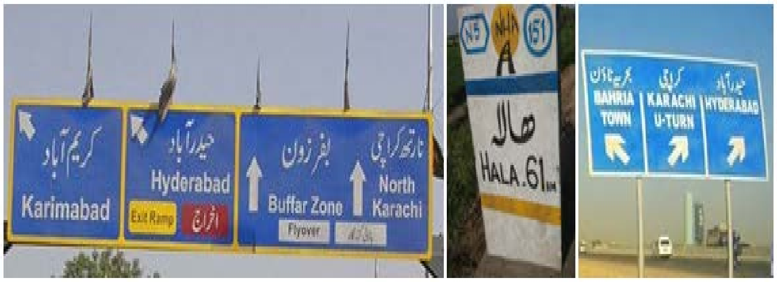


Figure 3 Multilingual Script

3.2 VARIETY IN TEXT SIZE AND FONT STYLES

The Latin script has a wide variety of text fonts. The font style and size varied from place to place if the text is written on boards, banners, and the wall. Figure 4 shows font variety and Figure 5 shows text and font style variations.

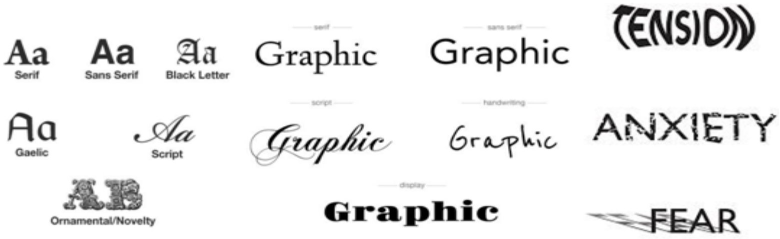


Figure 4 Font Variety

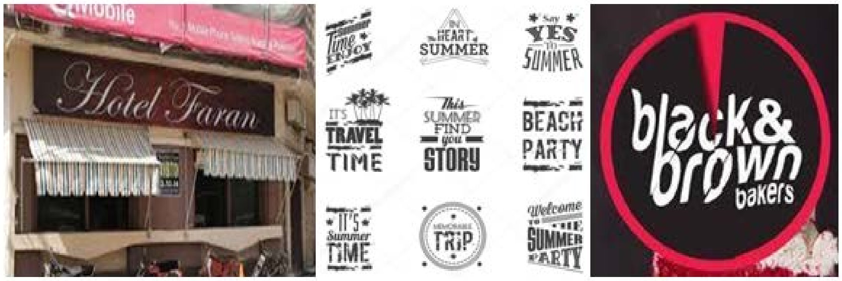


Figure 5 Font Variety in Wild / Natural

3.3 SKEWNESS

The border of the text image is called skew. Pay attention therefore to text scanning. It may cause errors if these pictures are segmented. Therefore, we have to detect the correct detection place. Figure 6 shows example photos of such cases from different angles.

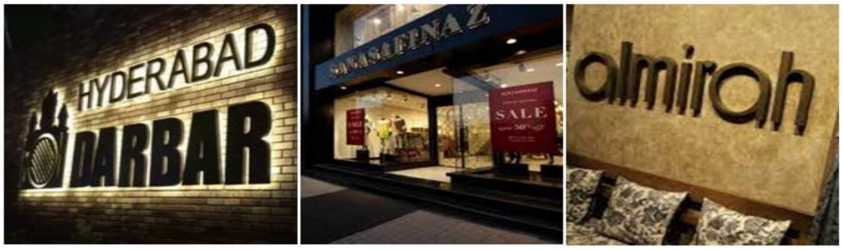


Figure 6 Different Angle Images

3.4 FONT COLOR AND CONTRAST

For text detection, font color with background color is most important. Both of these are interconnected. If font color is incorrect, it may cause text detection and recognition difficulties shown in Figure 7.

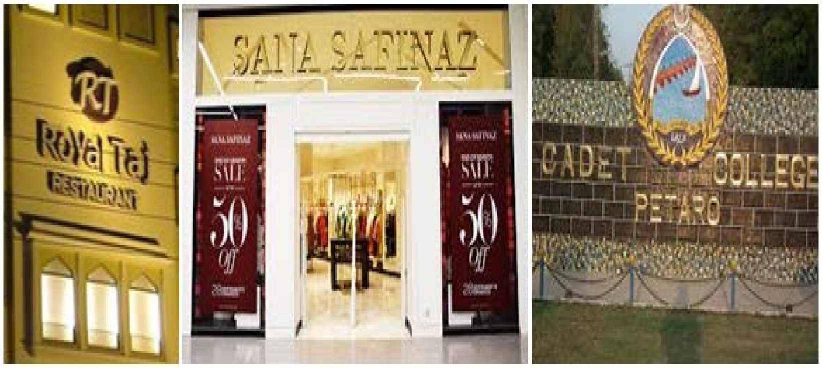


Figure 7 Font Color

3.5 COMPLEX BACKGROUND

Complex context also causes text detection complexity. Therefore, the right background is very necessary. Figure 8 illustrates some complex backgrounds.

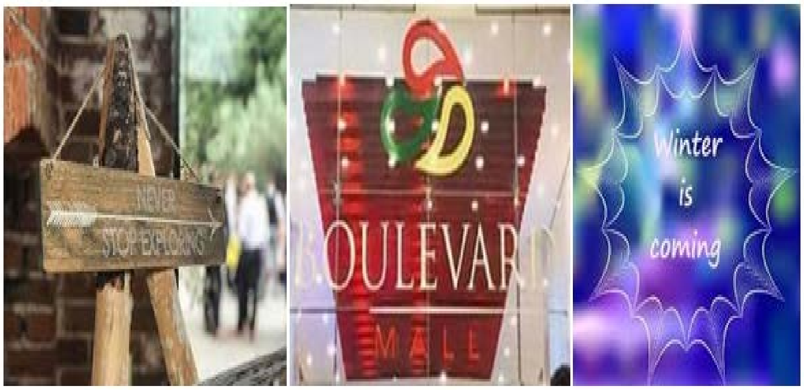


Figure 8 Complex Background

3.6 NEON SIGNBOARD

The Neon signboard radiates the electric signals through long, diluted neon or other gas-based, luminous gas discharge pipe. The signs are useful for distant viewing of the text but are a tough task. The neon signboard image in Figure 9 is shown below.

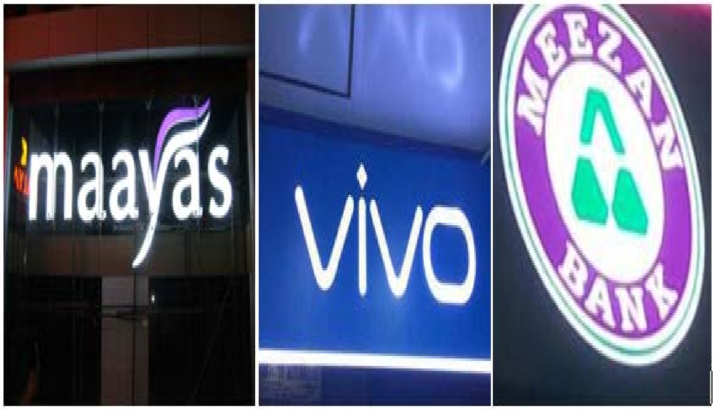


Figure 9 Neon Signboard

3.7 DISTANT OR BLUR SIGNBOARD

Photo details and clarity are defined in the resolution. If the image appears fluid or blurred and is not clearly visible, this is an image with low resolution. The motion may result in blurred images during acquisition. Any other kind of fog-like misleading focus could impair the quality of the image. It is a challenge to their impact on text detection. Examples of the blurry pictures are shown in Figure 10.

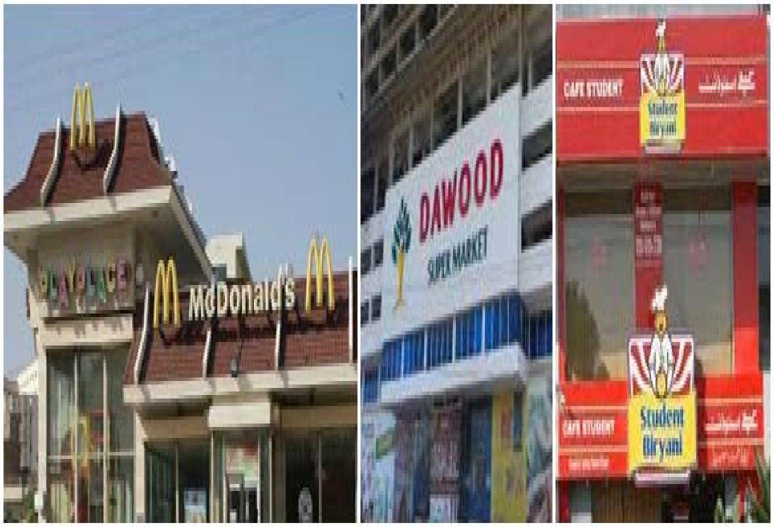


Figure 2 Distant and Blur Images

3.8 MIRROR IMAGE

It is especially difficult to recognize mirror-returned letters and words as shown in Figure 11.

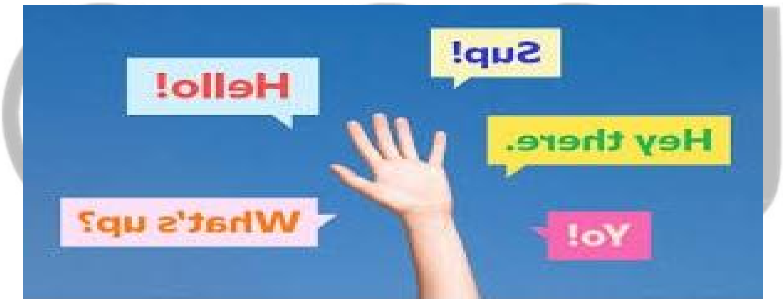


Figure 3 Mirror Image

3.9 LIGHTING EXPOSURE

If the photographs are taken at night, the detection and recognition corrections may be significantly reduced. The image is taken from The National Highways Authority (NHA) of India as shown in Figure 12.

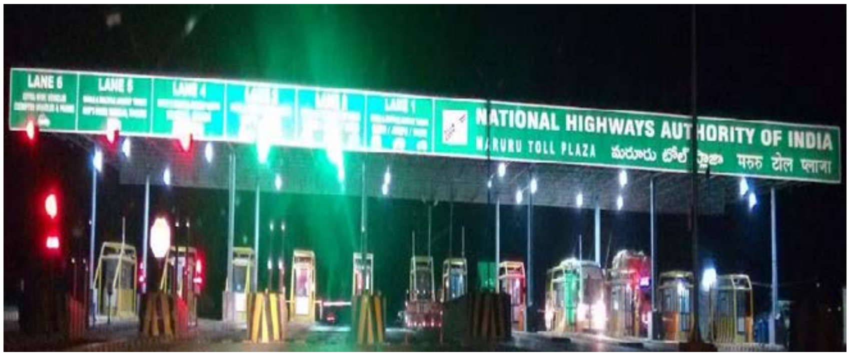


Figure 4 Lighting Exposure

The font size, fog, texture, background color, light problems and much more make it difficult to find text according to various research. The summary of the overall problems and challenges is given in Table 1.

Table 1: Summary of issues and challenges

Issues & Challenges	Explanation
Text in Multi script	Images are sometimes available in single script, and sometimes written in more script.
Varity in Text Size and Font Styles	The script has a various fonts style and sizes
Skewness	Skew makes identity more difficult.
Font Background Color and Contrast	Colors contrast also a big challenge it makes detection difficult task.
Neon Signboards	These types of signboards sometimes mixed the text so that text not recognized.
Distant and Blur images	Blur images with less resolution cause difficulty for text detection
Mirror Images	Mirror-reversed letters and words are especially difficult to *recognize.*
Lighting problem	Lighting shadows and reflection also a barrier between detection and recognition of text.

4. ENGLISH TEXT RECOGNITION FROM NATURAL SCENES

In a wide range of industrial applications reading text in natural scenes, called scene text recognition (STR), was an important task [23]. The maturity of Optical Character Recognition (OCR) systems led to its successful application on cleaned documents, but STR tasks were not carried out by most traditional OCR methods because of the various text appearances in the real world and be-cause these scenes are captured in imperfect conditions [24]. Every day, the popularity of smart handsets, tablets, and other mobile devices is growing [25]. These devices are applicable for government and commercial services that often require data from printed documents to be entered [26]. Hence, many datasets are available to detect English text from the wild [22]. Some of the examples of the dataset shown in Table 2.

Table 2 Dataset of English Text [21]

Dataset Language	Dataset	Captured Images	Segmented Words	Segmented Characters
English	ICDAR 2k3	509	999	11,615
	Chars74k	7,705	---	12,503
	SVT	350	647	3,796
	MSRATD 500	500	---	---
	IIIT5K-word	5,000	---	---
	CIFAR-10	60,000	---	---
	ICDAR 2k15	2,670	---	---

OpenCV stands for the android platform for Open Source Computer Vision [27]. It is a real-time computer-view programming library [28]. More than 2000 optimized algorithms are available in this library and are widely used worldwide. Android programmers can use this to implement numerous digital image processing algorithms on the Android phone platform [15]. Google's mobile vision Application Program Interface APIs introduced. These APIs provide a very easy to use interface with programming, through which faces, bar codes, QR codes, and text can be scanned without writing a great deal of code. It can also be used offline [29]. This feature is available on Android via Google Play Services [30]. ABBYY FineReader and Tesseract OCR, both the most frequently utilized one's systems. ABBYY FineReader is a commercial product and Tesseract OCR is a free OCR engine [31].

The study offers a flexible way in which texts are extracted through basic image processing techniques from difficult images. This system constructs an application that can recognize the English content in a Smartphone image and display the recognized text on your mobile screen as an editable format [32]. Figure 13 shows the Flow chart for the procedure. For randomly-selected test images, Figure 14 represents good qualitative results to demonstrate the application's effectiveness. Figure 15 represents the results of challenging natural scene images. We calculate the performance rate by using the formula introduce by Zaki et al. [32]. Figure 16 shows the performance rate of good and poor images.

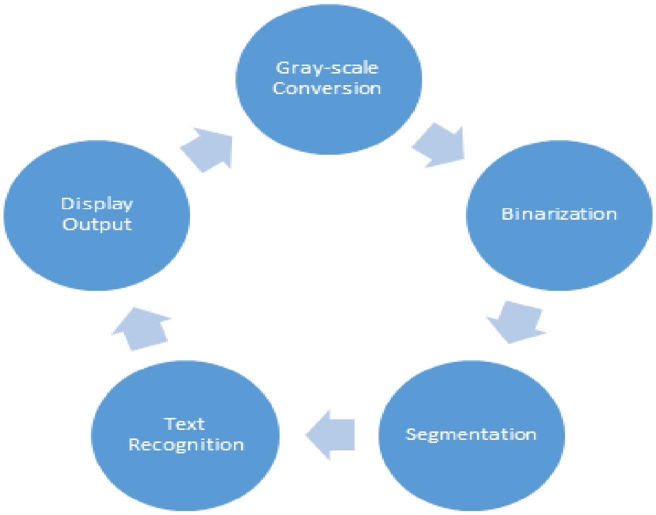


Figure 5 Block Diagram of System [32]

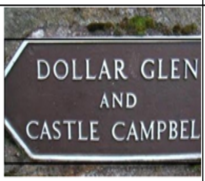
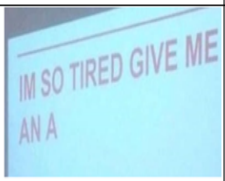
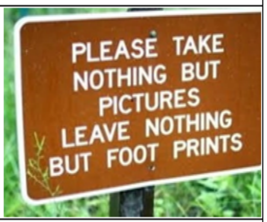
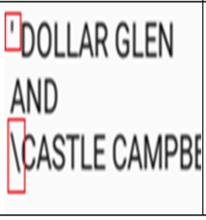
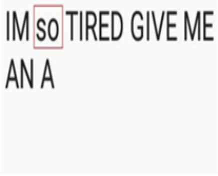
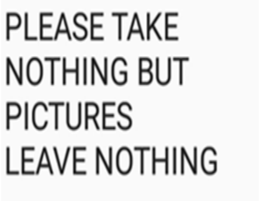
S. No:	1	2	3
Test Images	 	 	 PLEASE TAKE NOTHING BUT PICTURES LEAVE NOTHING BUT FOOT PRINTS
Results	 'DOLLAR GLEN AND \CASTLE CAMPBE	 IM so TIRED GIVE ME AN A	 PLEASE TAKE NOTHING BUT PICTURES LEAVE NOTHING

Figure 6 Good Qualitative Results

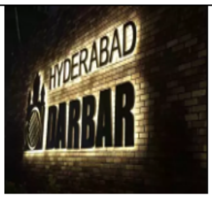
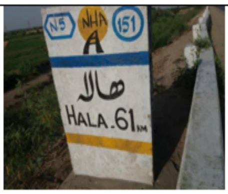
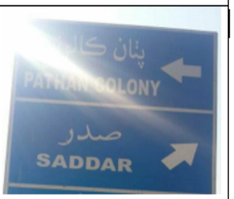
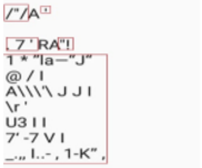
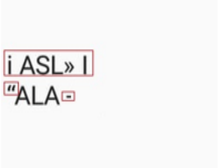
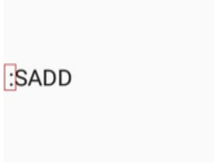
S. No:	1	2	3
Test Images	 	 	
Results	 /"/A' , 7 ' RA"! 1 * "la—"J" @ / I A\\\\'\ J J I \r ' U3 I I 7' -7 V I _.„ I..- , 1-K" ,	 i ASL» I "ALA-	 :SADD

Figure 7 Poor Performance

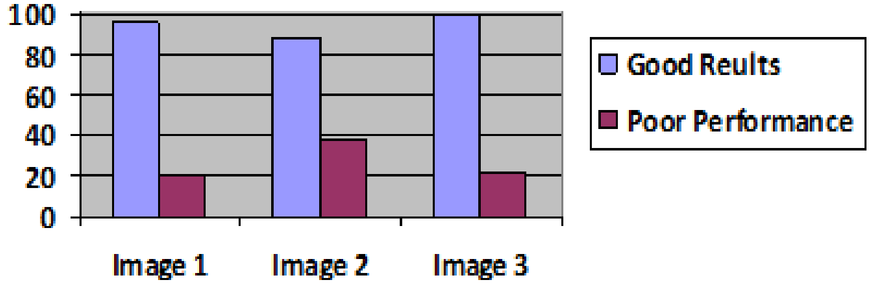


Figure 8 Performance Rate [32]

CONCLUSION

In recent years, text extraction and its acknowledgment of the natural environment have become an active and attractive field. This paper has presented various challenges and issues while detecting text especially from the images of natural scenes such as Multi-writing text, text fonts, text size, light shade, reflections, color text, contrast, missing words, fade, low resolution, neon signs were all challenging. Furthermore, the main steps of detecting scene text images are also presented in this paper with some results.

ACKNOWLEDGEMENT

Authors would like to encompass their sincere appreciation to Mehran University of Engineering and Technology, Jamshoro, Pakistan, for all their support and influence on this study.

REFRENCES

[1] B. H. Rudall, “Image recognition system,” *Kybernetes*, vol. 28, no. 1, pp. 30–31, 1999, doi: 10.1108/k.1999.06728aaa.004.

[2] U. Zaki, D. N. Hakro, K.-R. Khoumbati, M. A. Zaki, and M. Hameed, “Issues & Challenges in Urdu OCR,” *Univ. Sindh J. Inf. Commun. Technol.*, vol. 3, no. 1, pp. 42–49, 2019.

[3] J. Yang, X. Chen, J. Zhang, Y. Zhang, and A. Waibel, “Automatic detection and translation of text from natural scenes,” *ICASSP, IEEE Int. Conf. Acoust. Speech Signal Process. - Proc.*, vol. 2, pp. 2101–2104, 2002, doi: 10.1109/ICASSP.2002.5745049.

[4] J. Baek *et al.*, “What Is Wrong With Scene Text Recognition Model Comparisons? Dataset and Model Analysis,” pp. 4715–4723, 2019.

[5] K. J. Velmurugan and M. A. Dorairangaswamy, “Tamil character recognition using Android mobile phone,” *ARPN J. Eng. Appl. Sci.*, vol. 13, no. 3, pp. 1119–1123, 2018.

[6] T. E. De Campos, B. R. Babu, and M. Varma, “Character recognition in natural images,” *VISAPP 2009 - Proc. 4th Int. Conf. Comput. Vis. Theory Appl.*, vol. 2, no. Visigrapp, pp. 273–280, 2009, doi: 10.5220/0001770102730280.

[7] A. Veit, C. Science, and C. Tech, “COCO-Text: Dataset and Benchmark for Text Detection and Recognition in Natural Images.”

[8] U. ZAKI *et al.*, “Implementation Challenges in Information Retrieval System,” *SINDH Univ. Res. J. (SCIENCE Ser.*, vol. 4, no. 2, pp. 339–344, 2019, doi: http://doi.org/10.26692/sujo/2019.6.55.

[9] R. Baran, P. Partila, and R. Wilk, “Automated Text Detection and Character Recognition in Natural Scenes Based on Local Image Features and Contour Processing Techniques,” in *Intelligent Human Systems Integration*, 2018, pp. 42–48.

[10] M. S. Guzel, “A Novel Framework for Text Recognition in Street View Images. International Journal of Intelligent Systems and Applications in Engineering, 5(3), 140-144.,” *Int. J. Intell. Syst. Appl. Eng.*, vol. 5, pp. 140--144, 2017, doi: 10.1039/b000000x.

[11] X. Huang, T. Shen, R. Wang, and C. Gao, “Text detection and recognition in natural scene images,” *Proc. 2015 Int. Conf. Estim. Detect. Inf. Fusion, ICEDIF 2015*, pp. 44–49, 2015, doi: 10.1109/ICEDIF.2015.7280160.

[12] A. Mishra, K. Alahari, and C. V Jawahar, “Enhancing energy minimization framework for scene text recognition with top-down cues ☆,” *Comput. Vis. Image Underst.*, vol. 145, pp. 30–42, 2016, doi: 10.1016/j.cviu.2016.01.002.

[13] S. Tian, Y. Pan, C. Huang, S. Lu, K. Yu, and C. L. Tan, “Text Flow: {A} Unified Text Detection System in Natural Scene Images,” *Proc. ICCV*, pp. 4651–4659, 2015, doi: 10.1109/ICCV.2015.528.

[14] L. Neumann and J. Matas, “Text localization in real-world images using efficiently pruned exhaustive search,” *Proc. Int. Conf. Doc. Anal. Recognition, ICDAR*, pp. 687–691, 2011, doi: 10.1109/ICDAR.2011.144.

[15] A. Rohira, R. Shah, O. Sadarangani, M. Shinde, and S. Therese, “Word Detection and Translation,” *SSRN Electron. J.*, pp. 1–5, 2019, doi: 10.2139/ssrn.3372202.

[16] M. Jaderberg, K. Simonyan, A. Vedaldi, and A. Zisserman, “Synthetic Data and Artificial Neural Networks for Natural Scene Text Recognition,” pp. 1–10, 2014, doi: 10.1109/ICCV.2013.76.

[17] S. Ahmed, S. Naz, M. Razzak, and R. Yusof, “Arabic Cursive Text Recognition from Natural Scene Images,” *Appl. Sci.*, vol. 9, no. 2, p. 236, 2019, doi: 10.3390/app9020236.

[18] J. Gao and J. Yang, “An adaptive algorithm for text detection from natural scenes.,” *Proc. 2001 IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognition. CVPR 2001*, vol. 2, pp. II-84-II–89, 2001, doi: 10.1109/CVPR.2001.990929.

[19] A. Mishra, K. Alahari, and C. V. Jawahar, “Top-down and bottom-up cues for scene text recognition,” *Proc. IEEE Comput. Soc. Conf. Comput.*

Vis. Pattern Recognit., pp. 2687–2694, 2012, doi: 10.1109/CVPR.2012.6247990.

[20] S. B. I. N. Ahmed, S. Naz, M. I. RAZZAK, and R. B. T. E. Yusof, “A Novel Dataset for English-Arabic Scene Text Recognition (EASTR) -42K and Its Evaluation Using Invariant Feature Extraction on Detected Extremal Regions,” vol. 7, pp. 19801–19820, 2019.

[21] L. Gomez and D. Karatzas, “Multi-script text extraction from natural scenes,” *Proc. Int. Conf. Doc. Anal. Recognition, ICDAR*, pp. 467–471, 2013, doi: 10.1109/ICDAR.2013.100.

[22] U. ZAKI *et al.*, “Dataset of Urduud1k from Natural Scenes,” *Inst. Inf. Commun. Technol. Univ. Sindh, Jamshoro, Pakistan*, vol. 51, no. 04, pp. 595–600, 2019, doi: http://doi.org/10.26692/sujo/2019.12.95.

[23] F. Zhan and S. Lu, “ESIR: End-to-end scene text recognition via iterative image rectification,” *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 2019-June, pp. 2054–2063, 2019, doi: 10.1109/CVPR.2019.00216.

[24] M. Liao *et al.*, “Scene Text Recognition from Two-Dimensional Perspective,” *Proc. AAAI Conf. Artif. Intell.*, vol. 33, pp. 8714–8721, 2019, doi: 10.1609/aaai.v33i01.33018714.

[25] T. Kobchaisawat and T. H. Chalidabhongse, “Thai text localization in natural scene images using Convolutional Neural Network,” *2014 Asia-Pacific Signal Inf. Process. Assoc. Annu. Summit Conf. APSIPA 2014*, 2014, doi: 10.1109/APSIPA.2014.7041775.

[26] A. Canedo-Rodríguez, J. H. Kim, S. Kim, and Y. Blanco-Fernández, “English to Spanish translation of signboard images from mobile phone camera,” *Conf. Proc. - IEEE SOUTHEASTCON*, pp. 356–361, 2009, doi: 10.1109/SECON.2009.5174105.

[27] Z. Zhu and Y. Cheng, “Application of attitude tracking algorithm for face recognition based on OpenCV in the intelligent door lock,” *Comput. Commun.*, vol. 154, pp. 390–397, 2020.

[28] R. du Toit, G. Drevin, N. Maree, and D. T. Strauss, “Sunspot Identification and Tracking with OpenCV,” in *2020 International SAUPEC/RobMech/PRASA Conference*, 2020, pp. 1–6.

[29] T. B. Hossain, Y. A. Akter, and M. A. Rahman, “Voice mail application for visually impaired persons,” *Recent Res. Sci. Technol.*, pp. 15–18, 2020.

[30] G. Zhou, Y. Liu, Q. Meng, and Y. Zhang, “Detecting multilingual text in natural scene,” *Proc. 2011 1st Int. Symp. Access Spaces, ISAS 2011*, pp. 116–120, 2011, doi: 10.1109/ISAS.2011.5960931.

[31] Y. S. Chernyshova, A. V. Sheshkus, and V. V. Arlazarov, “Two-step CNN framework for text line recognition in camera-captured images,” *IEEE Access*, vol. 8, pp. 1–1, 2020, doi: 10.1109/access.2020.2974051.

[32] M. A. Zaki, S. Zai, M. Ahsan, and U. Zaki, “Development of an android app for text detection,” *J. Theor. Appl. Inf. Technol.*, vol. 97, no. 20, pp. 2485–2496, 2019.