World Journal of Advanced Research and Reviews

eISSN: 2581-9615 CODEN (USA): WJARAI

Cross Ref DOI: 10.30574/wjarr

Journal homepage: https://wjarr.com/

(Review Article)

Camera-based OCR scene text detection issues: A review

Francisca O Nwokoma [1], *, Juliet N Odii [1], Ikechukwu I Ayogu [1] and James C Ogbonna [2]

[1] *Department of Computer Science, School of Information and Communication Technology, Federal University of Technology Owerri, Imo State, Nigeria.*

[2] *Department of Mathematics and Computer Science, Clifford University Owerrinta, Abia State, Nigeria.*

World Journal of Advanced Research and Reviews, 2021, 12(03), 484–489

Publication history: Received on 13 November 2021; revised on 21 December 2021; accepted on 23 December 2021

Article DOI: https://doi.org/10.30574/wjarr.2021.12.3.0705

Abstract

Camera-based scene text detection and recognition is a research area that has attracted countless attention and had made noticeable progress in the area of deep learning technology, computer vision, and pattern recognition. They are highly recommended for capturing text on-scene images (signboards), documents with a multipart and complex background, images on thick books and documents that are highly fragile. This technology encourages real-time processing since handheld cameras are built with very high processing speed and internal memory, are quite easy and flexible to use than the traditional scanner whose usability is limited as they are not portable in size and cannot be used on images captured by cameras. However, characters captured by traditional scanners pose fewer computational difficulties as compared to camera captured images that are associated with divers' challenges with consequences of high computational complexity and recognition difficulties. This paper, therefore, reviews the various factors that increase the computational difficulties of Camera-Based OCR, and made some recommendations as per the best practices for Camera-Based OCR systems.

Keywords: Camera-Base; Scene Images; Image Acquisition; OCR; Text Detection; Text Recognition

1. Introduction

Fundamentally, the original goal of OCR is to process document images acquired by desktop scanners. This is attributed to the fact that the resolutions of scanned images are sufficient. Recently, Handheld Cameras started gaining persistent demand for OCR document image acquisition. One of the major limitations of the traditional scanners is that they only deal with text on paper (print document) while the target of handheld digital cameras is not limited to just print documents; they can handle scene images as well [1].

Camera-based OCR technology is therefore concerned with recognizing characters captured by portable devices with an attached camera be it handheld or wearable cameras. It is best suited for scene images, documents with multipart and complex backgrounds since it is flexible to use than the traditional Scanner whose usability is limited as they are not portable in size. Moreover, the shot speed of a scanner is slower than that of a digital camera. Hand-held digital cameras such as high-end cell phones, Personal Digital Assistants (PDA), smartphones: iPhones, iPods, Android phones, Windows phones, etc. possess a very high processing speed and internal memory which encourages real-time processing [2]. However, characters captured by a camera are different from those in scanned documents and their recognition is very difficult [1]. Text detection and recognition in documents with multiple parts, complex and heterogeneous backgrounds, mixed fonts or other irregular characteristics are much more challenging than in regular, standard font, plain printed text documents [3]. The difficulty of this task is further aggravated when the document

* Corresponding author: Francisca O Nwokoma
Department of Computer Science, School of Information and Communication Technology, Federal University of Technology Owerri, Imo State, Nigeria.

image is poorly captured either due to operator error, incompetence, low device quality or poor environmental conditions such as lighting when a camera is used to capture the image [4, 5], because of the non-contact nature of digital cameras attached to handheld or wearable devices, the output of acquired images mostly suffer from skew and perspective distortion. It is affected by complex background, blur, low resolution, non-uniform lightning, skew and perspective distortion.

2. Applications of Camera-Based OCR

This technology as noted by [1, 6] has gained popularity in the field of computer vision and its importance is evident in the following areas:

- Real-time Recognition of print documents of standard, multipart and complex background, scene images;
- Mobile text recognizer and speech generator for the visually impaired;
- A camera-based OCR technique to develop a Portable Camera-Based Assistive Text Reading from Hand-Held Objects for Blind Person which was proposed by [7];
- Business cards are popular targets of camera-based OCR on mobile phones;
- Thick books and precious historical books are also a target [8];
- Smart desktop systems, where a camera will capture documents or texts on the desk and understand their location and contents;
- Image search where a personal or large public image database is searched for images including a keyword specified [9];
- Mobile Dictionary [10];
- Sign Detection and Translation: This is the ability to detect and recognize text using PDAs or cellular phones. [11] Evaluated a prototype system for sign detection and translation of text from natural scenes;
- License Plate Reading: automatic license plate reading (ALPR) is applied in many applications especially in traffic control systems. Many systems have been developed for ALPR and many commercial products are making practical use of it such as in toll collecting monitoring, security management, parking lot billing, and road law enforcement [8, 12];
- Document Archiving: With the portability and flexible nature of camera-based OCR, users can conveniently carry such a device anywhere for instant recording of record interesting document pages.

3. Tasks of Camera-Based OCR

Text recognition in scene images, multipart documents, documents with complex backgrounds and video is comprised of three major tasks, Text image Acquisition/data collection, text localization and text recognition. These tasks are shown in Figure 1. This paper will be dwelling on the first task which is text image acquisition since all the challenges associated with Camera-Based OCR arises from this phaseand methods.

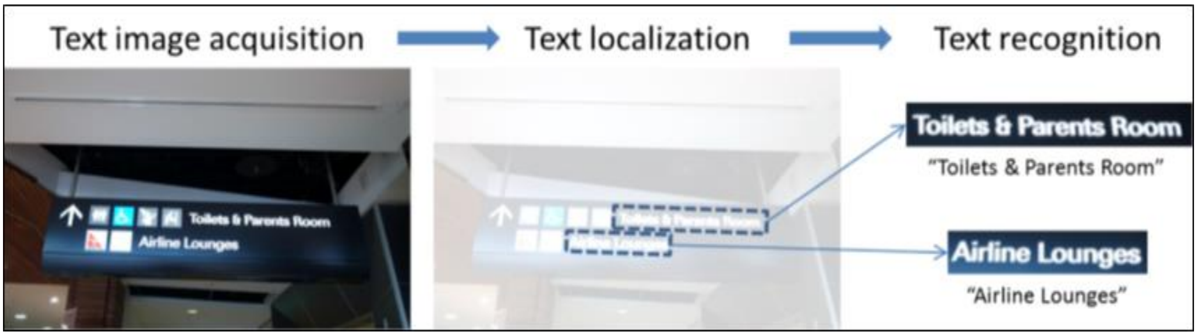


Figure 1 Camera-Based OCR Tasks. Source [1]

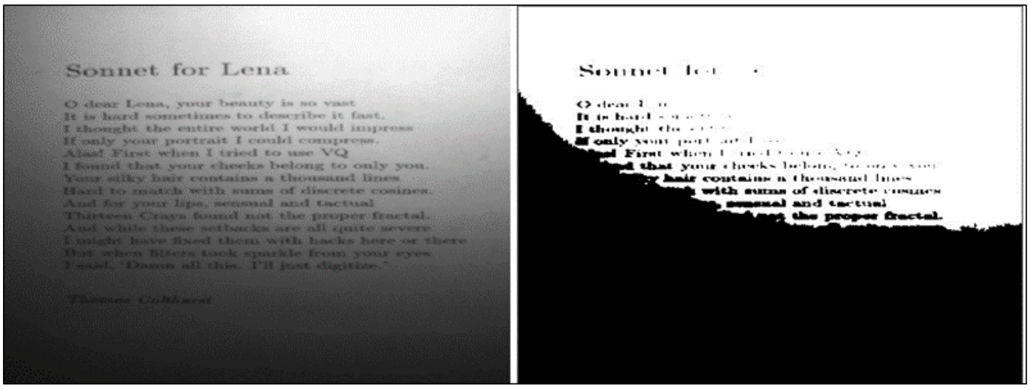
3.1. Text Image Acquisition/Data Collection

This is the first step of Camera-based OCR. It involves image capturing and preprocessing (rectification and deblurring) for improving the quality of the captured image. The essence of this is to make the next text, which is a localization and recognition task easier. As noted by [1], many difficulties of camera-based OCR such as low resolution, motion blurring, perspective distortion, non-uniform lighting, specular by flash and occlusion are caused by this acquisition step [13, 14].

3.2. Preprocessing

The next phase after capturing is preprocessing. The main objective of this phase is to make it easy and possible for the OCR to distinguish character/word from the background. This can be achieved using the traditional machine learning approach such as Binarization, or the deep machine learning approach such as CRAFT algorithm, which incorporates the following Binarization, skew correction, Noise removal, Thinning and Skeletonization techniques.

Binarization: means converting a coloured image into an image that consists of only black and white pixels (Black pixel value=0 and White pixel value=255). As a basic rule, this can be done by fixing a threshold (normally threshold=127, as it is exactly half of the pixel range 0–255). If the pixel value is greater than the threshold, it is considered as a white pixel, else considered as a black pixel [15]. However, this technique fails when lightning is not uniform in the image as shown in Figure 2



Sonnet for Lena

O dear Lena, your beauty is so vast
It is hard sometimes to describe it fast.
I thought the entire world I would impress
If only your portrait I could compress.
Alas! First when I tried to use VQ
I found that your cheeks belong to only you.
Your silky hair contains a thousand lines
Hard to match with sums of discrete cosines.
And for your lips, sensual and tactual
Thirteen Crays found not the proper fractal.
And while these setbacks are all quite severe
I might have fixed them with hacks here or there
But when filters took sparkle from your eyes
I said, 'Damn all this. I'll just digitize.'

Thomas Colthurst

Figure 2 Binarization using threshold captured under non-uniform lighting. Source [15]

Challenges and Factors that Contributes to the Difficulties of Camera-Based OCR and their Diverse Effects

The various challenges associated with camera-based OCR, which emanates from image acquisition are listed and explained as follows:

3.3. Scene impediment

Text detection and recognition in documents with multiple parts, complex and heterogeneous background, mixed fonts or other irregular characteristics is much more challenging than in regular, standard font, plain printed text documents [3, 11, 16]. The difficulty of this task is further aggravated when the document image is poorly captured due to either operator error, incompetence, low device quality or poor environmental conditions such as lighting when a camera is used to capture the image [4, 5].

3.4. The multiplicity of Scene Text

Research has shown that processing text elements in documents with a non-regular or complex mix of elements and quality issues still pose computational difficulties to OCR systems because texts can be written in different fonts, appear in different resolutions, appear with irregular spacing, overlap and bear artistic effects [17]. This is contrary to Text in document images, which normally appears in regular font, single colour, and uniform layout [18].

3.5. Intrusive Factors

Furthermore, camera-based OCR is susceptible to lighting conditions and other operator-dependent exposure settings that often result in uneven lighting, blurred or perspective-distorted images and other forms of degradation such as warping and shadow overlay [4, 5].

3.6. Low-Resolution Images

Images captured with cameras usually have low resolutions.

3.7. Image Zooming and Focusing

Focus is usually a drawback for many digital devices as they are designed to operate over a variety of distances. Slight perspective changes can cause uneven focus even at short distances [19].

3.8. The High Dimensionality

The high dimensionality of the feature space is one of the very prominent challenges for OCR techniques on the problem scenario described in the preceding paragraphs. The problem of high dimensionality is mostly occasioned by the presence of noise in the form of a subsuming background or low contrast background, which reduces the accuracy of text detection and recognition by OCR systems. Another major challenge is the aspect ratio. Some texts are short while others are lengthier. Texts detection in this scenario involves a process with high computational complexity, since the location, scale and length of these texts need to be considered within the search process. Thus, the computational time is increased without a corresponding improvement in the recognition rate.

3.9. Complex Background

Text printed on a background that has a very solid colour, separation is rather easy like ordinary scanned documents. In scene Images, characters and their background often have very low contrasts and as such, separation is a little more difficult even when a solid coloured background is used. Captions are mostly without solid backgrounds because they are placed on top video frames directly. This case is one of the most difficult cases of text localization.

3.10. Curved Scene Images

Many patterns seem like characters as a result of confusing and ambiguous shapes in the scene. "Y" shaped edges are found on the corner of a room. Around leaves of trees, we also find dense and fine (i.e., high-frequency) edge structures which look like dense text lines. Equally, it is also obvious that some decorated characters seem like a branch of a tree. This difficulty invites a very important question; what are character patterns? More precisely, what is the difference between character patterns and non-character patterns? We never read the corner as "Y". Therefore, the results of localization will be given as a set of rectangles (or, more generally, arbitrarily shaped connected regions) each of which contains a single character, or a single word. For single-character rectangles, neighbouring character rectangles will be grouped (i.e., concatenated) to form a word. Normally, scene images are often curved, rotated or even slanted. Figure 3 shows the diagrams of some of the listed scene text challenges.

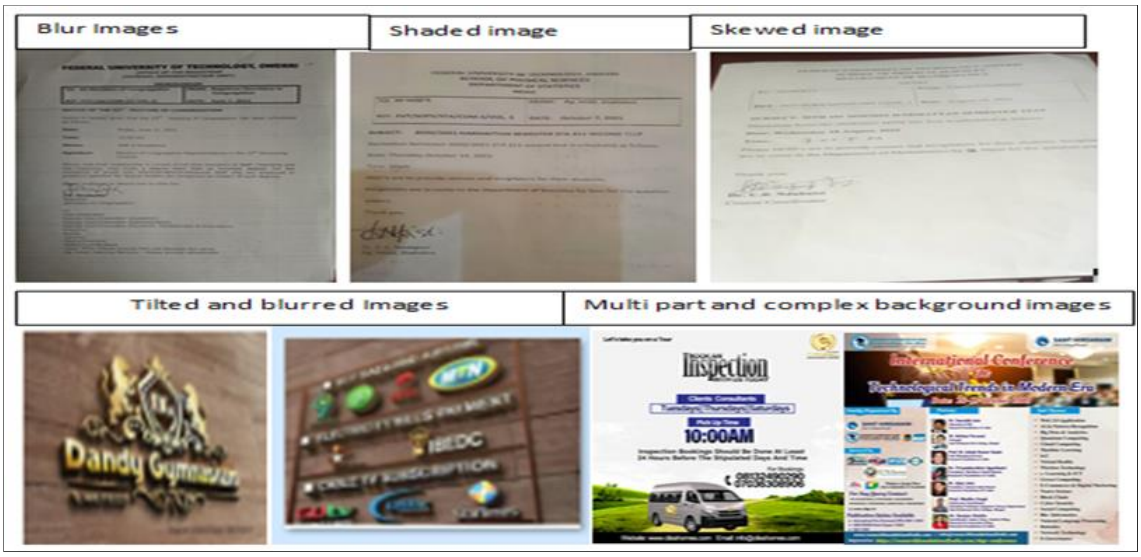

Blur Images
Shaded image
Skewed image
Tilted and blurred Images
Multi part and complex background images

Figure 3 Some Challenges of Camera-based OCR Images

4. Camera-Based OCR System Best Practices

The various challenges associated with camera-based OCR, which emanates from image acquisition are listed and explained as follows:

4.1. Image Acquisition

- To acquire high-resolution images at some distance, we need to perform zooming to the area of interest [19]
- To avoid low resolution, do not capture all the text in one frame.
- The mosaic technique is needed to put pieces of text images together as a large high-resolution image.

- Auto zooming should be used in the background around an object that has low variance compared to the main object. In this case, the variance in the observation window could be used as an indicator of best zoom [20].
- The best focus is achieved when edges are strongest in the image.
- The sum of the differences between neighbouring pixels, the sum of gradient magnitude, or the sum of a Laplacian filter's output could be used as the overall edge strength measure. [21]
- To avoid perspective distortion, the camera should directly face the document.

4.2. Image Pre-processing

- Image Conversion: The captured image should be converted to have only black and white pixels (Black pixel value=0 and White pixel value=255);
- Image Cropping: The text area of the captured image should be cropped. Firstly, the text area should be predefined depending on the type of image so that the unwanted areas from the image will be eliminated [22];
- Image Brightness Correction: The brightness levels of the image should be varied to remove the dark areas. The new brightness can be calculated using any of the brightness methods such as the Root Mean Square (RMS) pixel method [22];
- Gamma Correction: Gamma values of the image are varied depending on its brightness level. A lookup-mapping table should be built to map the input pixel values to the output gamma-corrected values;
- Skew Correction: The geometric distortion caused by the camera position is corrected hereby. According to [23], many techniques have been developed to handle this.

5. Conclusion

In this paper, we reviewed the difficulties associated with camera-based OCR and the likely factors that contributed heavily to its computational complexity, and recommend best practices for extracting scene texts from complex images. As was clearly stated, the major challenges of every camera-based OCR system originate from the first stage, which is image acquisition. When the document image is poorly captured it results to skew, tilted and misalignment of a document. For a pure environmental condition or uneven lightning, blurred images are obtained. Therefore, text detection and recognition in documents with multiple parts, complex and heterogeneous backgrounds, poorly captured images, mixed fonts or other irregular characteristics are much more challenging than in regular, standard font, plain printed text documents, which the traditional desk OCR is very comfortable with. However, with the recommended best practices, a camera-based OCR scene text detection system can easily be designed and deployed to extract texts from complex images.

Compliance with ethical standards

Acknowledgments

The authors are thankful to the anonymous reviewers of this work for their reviews.

Disclosure of conflict of interest

The authors declare that no known competing interest exists.

References

[1] Uchida S. 25 – Text Localization and Recognition in Images and Video. 2020; 0–42.

[2] Mollah AF, Majumder N, Basu S, Nasipuri M. Design of an Optical Character Recognition System for Camera-based Handheld Design of an Optical Character Recognition System for Camera-based Handheld Devices. IJCSI Int J Comput Sci Issues. 2011; 8(4): 1694–0814.

[3] Gupta N, Banga VK. Localization of Text in Complex Images Using Haar Wavelet Transform. Int J Innov Technol Explor Eng. 2012; 1(6): 111–5.

[4] Trémeau A, Godau C, Karaoglu S, Muselet D. Detecting Text in Natural Scenes Based on a Reduction of Photometric Effects : Problem of Color Invariance. In: In International workshop on computational color imaging. 2011. 214–29.

[5] Mhaske AV, Sadavarte MS. Portable Camera Based Assistive Text Reading from Hand Held Objects for Blind Person. Int J Adv Res Comput Commun Eng. 2016; 5.

[6] Cao D, Zhong Y, Wang L, He Y, Dang J. Scene Text Detection in Natural Images: A Review. Symmetry (Basel). 2020; 12(12): 1956.

[7] Sandhiya D, Shynimol KM, Alagi TT, Vidhya RM. Portable Camera Based Assistive Text Reading From Hand Held Objects for Blind Persons. Digit Signal Process. 2018; 10(8): 136–8.

[8] Liang J, Doermann D, Li H. Camera-based analysis of text and documents : a survey. Int J Doc Anal Recognit. 2005; 7: 84–104.

[9] Xu Liu, Doermann D. Mobile Retriever: Access to digital documents from their physical source. Int J Doc Anal Recognit. 2008; 11(1): 19–27.

[10] WatanabeK Y, YokomizoYS, Okada. Translation camera on mobile phone. In: Multimedia and Expo. 2003.

[11] Yang J, Chen X, Zhang J, Zhang Y, Waibel A. Automatic detection and translation of text from natural scenes. In: 2002 IEEE International conference on acoustics, speech, and signal processing. 2002. II--2101.

[12] Chang S, Chen L, Chung Y, Chen S. Automatic license plate recognition. IEEE Trans Intell Transp Syst. 2004; 5(1): 42–53.

[13] Vit Comparison of various edge detection technique. Int J Signal Process Image Process Pattern Recognit. 2016; 9: 143–58.

[14] Nixon MS, Aguado AS. Feature Extraction and Image Processing for Computer Vision Fourth Edition. United States: Elsevier Ltd. 2020. 1–818.

[15] Sunsmith R. A basic explanation of the most widely used preprocessing techniques by the OCR system. 2019.

[16] Mahmood H. Text-detection and recognition from natural images. Loughborough University. 2020.

[17] Fei L, Wang K, Lin S, Yang K, Cheng R, Chen H. Scene text detection and recognition system for visually impaired people in real world. In: SPIEDigitalLibrary.org/conference-proceedings-of-spie. 2018; 10.

[18] Hanaa Fathi Mahmood. Text Detection and Recognition from Natural Images. 2020.

[19] Doermann D, Liang J, Li H. Progress in camera-based document image analysis. In: Seventh International Conference on Document Analysis and Recognition, 2003 Proceedings. 2003; 606–16.

[20] Mirmehdi M, Palmer PL, Kittler J. Towards optimal zoom for automatic target recognition. In: Proceedings of the Scandinavian Conference on Image Analysis. 1997; 447–54.

[21] Zandifar A, Duraiswami R, Chahine A, Davis LS. A video based interface to textual information for the visually impaired. In: Proceedings Fourth IEEE International Conference on Multimodal Interfaces. 2002; 325–30.

[22] Zacharias E, Teuchler M, Bernier B. Image Processing Based Scene-Text Detection and Recognition with Tesseract. April 2020; 1–6.

[23] Jin D, Park W, Jeong S-G, Kim C-S. Harmonious Semantic Line Detection via Maximal Weight Clique Selection. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021; 16737–45.