



Camera Captured Natural Scene Gurmukhi Text Extraction and Recognition: Methodology and Challenges

Jaspreet Kaur

Assistant Professor, Department of Computer Science,
Khalsa College for Women, Sidhwan Khurd, Ludhiana, India

Abstract: Text present in natural scene images such as menu cards, traffic signboards, house or shop addresses, banners, route boards, milestones contains ample amount of information. Foreigners who could not understand local language feel difficulty in reading such information. Smart phone based systems, which provide the facility to capture and extract text from such textual images written in local scripts, so that it can be translated in their desired script, helps them to travel comfortably. This research work gives information about the basic OCR methodology and various challenges related to mobile camera captured natural scene text extraction and recognition.

Keywords: Smart phone, Script, OCR, Extraction, Recognition

I. INTRODUCTION

Text is indispensable component in today's world of science; this is a way to represent information; text is present in different forms around us such as in the form of handwritten or printed text. Text is also present in the form of textual images that could be document images or natural scene images. In the past, text was scanned using scanners like flatbed or sheet-fed. After that handheld digital cameras came into use, to click document images and scene text images as scanners could not be available all the time and everywhere. Nowadays smart phone cameras are used to capture the document images and scene text. The extraction of text present in natural scene images becomes more challenging as natural scene images often have cluttered backgrounds and multiple patterns for scene text. But there is an immense need to extract text from natural scene images because textual images contain ample amount of information such as to perform content based image retrieval, image/video based analytics, automatic indexing and industrial automation [1].

II. UNDERSTANDING NATURAL SCENE TEXT

Images of our surrounding environment which are clicked with no prior content restriction are natural scene images. Natural scene images may contain various kinds of objects including text. This text is called scene text.

Document text is usually scanned or camera captured, which follows simple foreground and background styles and

standard layouts of books and newspapers. Heading and paragraph font size are easily distinguishable in these images. On the other side, natural scene camera captured, textual images may contain large amount of non-text content. Text in single scene may contain cluttered background, different layout patterns and font style variations etc.

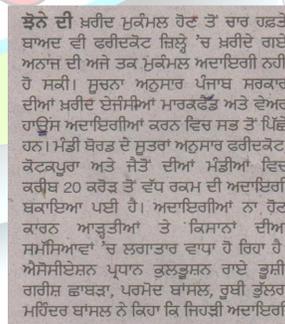


Fig. 1 Document images



Fig. 2 Natural scene images

III. CAMERA ACQUIRED NATURAL SCENE TEXT EXTRACTION AND RECOGNITION : METHODOLOGY

Text extraction from captured scene images means to acquire the textual portion of the scene by separating the text from the background. Text extraction from camera captured natural scene images is a challenging process which undergoes the following steps i.e. pre-processing, scene text detection, scene text localization, scene text verification, segmentation, recognition and enhancement [2].

A. Pre-processing

Camera captured images require certain amount of pre-processing operations because of the noisy nature of acquired images. Images captured through cameras often suffer from blur, noise and distortion etc. This could be due to uneven lightening while capturing image, addition of unwanted signals and/or the non-parallel axis of the imaging plane and camera. In this stage, such kinds of anomalies are removed from the image under process to gain better extraction outcome.

B. Text Detection

After the pre-processing operation next step is to determine the existence of any text in the captured image. This step is often not required in case of document images. But while considering scenic images, this step is vital as scenic images often contain text like objects and may or may not have text. Therefore, to proceed next, this is necessary to determine the presence of text. Text detection can also be performed using text localizers. Both terms are commonly used interchangeably by the researchers.

C. Text Localization

Text localization is a way to find the exact location of the detected text. This generates boundaries around the detected

text locations. The localization stage has an immense impact on the whole extraction process. If the localization of text is performed well, then the extraction would be better. Textual features, gradient features, colour, layout and font style of text acts as basic clues for the localization of the text. But this is critical to perform localization in scenic images due to the absence of strict text layouts and boundaries as compared to document images.

D. Text Verification

This step deals with the exclusion of false positives of localization outcome. Text localization often considers non-text objects as text due to some resemblance properties. Therefore, after performing localization, this is imperative to perform text verification to verify and remove wrong predictions about the text and non-text components. This step eliminates those non-text portions of images that are bounded in the localization step.

E. Segmentation and Enhancement

Segmentation means to segment the textual portions of image which are localized and verified as text, from the background. Here, text portions are extracted from the image for recognition. Segmentation is performed based on line, word and further character component. Line segmentation, segments the region into sub-regions each containing single lines. Word segmentation divides lines into isolated words and character segmentation results into separate sub-regions, each containing single character or sub-character. Further, enhancement is used to correct distorted text to enhance resolution for better recognition [1].

F. Text Recognition

Text recognition is performed with the help of feature extraction and classification techniques. Feature analysis is performed to determine, all the descriptors or the features sets that are used to describe all the characters. Classification is performed by comparing the input character image with predefined templates for each character class. Each comparison provides similarity measure between the input character and template.

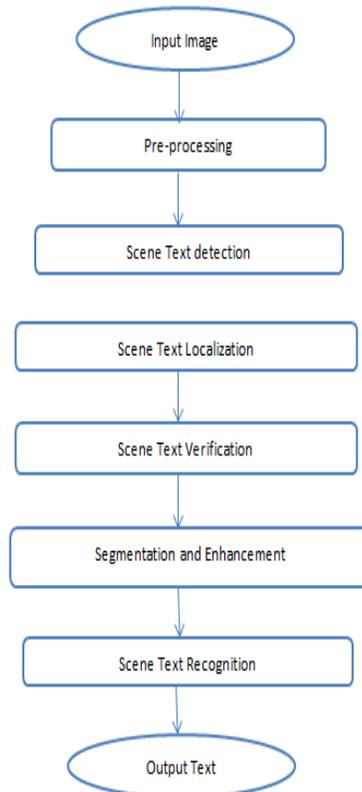


Fig. 3 Natural scene text extraction and recognition: methodology

IV. CHALLENGES ISSUES REGARDING NATURAL SCENE TEXT EXTRACTION AND RECOGNITION USING SMART PHONES

While dealing with text present in natural scene image captured using smart phone camera, there would be certain challenges related with the smart phones, natural scene and the text as compared to the scanned images or document images. Major challenges occur due to the diversity of scene text, flexibility in acquisition and complexity in environments. The obstacles that occur during natural scene text extraction that makes the extraction process more challenging are listed as follows:

A. Challenges due to the Use of Smart Phone Camera

- *Blurring and Noise*

Capturing images using camera, is not always noise free. Most of the times images captured using smart phone camera suffer from blurring and unwanted signals i.e. noise. Text present in blurred and noisy images is difficult to detect.

- *Non-Uniform Lightening*

When capturing photos with smart phone camera, uneven lightening conditions become a hindrance in the quality of the image. Due to the uneven luminance, uncontrolled lightening conditions, other sub-problems occur such as color distortion and feature deterioration. Uneven lightening conditions generate the problem of false text detection, localization and segmentation [3].

- *Skewness*

Camera captured images suffers from skewness problem due to the angle of the camera while clicking the photograph. Skewness means when the horizontal alignment of text is affected by the camera angle. Skewness makes the detection and extraction processes more challenging.

- *Perspective Distortion*

Perspective distortion occurs when imaging plane and text plane are not parallel to each other. Recognition models are trained using rectangular boundary text samples, but due to perspective distortion, text distorts and thus decreases the performance of recognition model [1].

- *Zoom and Blur*

In present era, digital cameras and smart phone cameras provide the zoom and focus facility during clicking pictures. As sharp edge response is required for the better text detection and recognition but large aperture and small distance may cause uneven focusing. Hence, results into poor detection and recognition results [2].

B. Challenges regarding Natural Scene Text Extraction

- *Aspect Ratio (Size)*

In natural scene image aspect ratio of text can be different depending upon the type of the scene. It could be of small size on a signboard and large on a banner. Different size and font style can be present in same scene.

- *Cluttered Background*

In natural scene, there exist multiple man-made objects such as paintings, bricks, window, grid, symbols that resembles with the text formations. In such situations it becomes cumbersome to distinguish between the text and non-text elements. This becomes challenging task to extract text form natural scene images [3].



- **Variation in Text Layouts, Multi- script Text**
Text present in natural scenes can be both multi-script and in the form of versatile layouts to catch the attraction of viewers. Such kind of layout is cognitively perceivable by the human readers but add complexity in the recognition and segmentation process. Multi-script text can also be present along with the versatile layouts that cause more complexity [1].
- **Color, Space Complexity**
Document images often exist in uniform color patterns along with the uniformity in the spacing between the lines. Natural scene images often contain multiple colors and non-uniform spacing between the texts [4].
- **Multi-oriented/Curved Text**
Text in natural images may exist in different orientations. Text can be non-horizontal and curved also. Straight forward techniques cannot be applied for such text orientations. Thus, it is cumbersome to detect multi-oriented text.

C. **Challenges regarding Gurmukhi Text Extraction**
Available methods work well for the roman scripts such as English due to their simple structure. But the structure of Indic scripts such as Gurmukhi is complex and different as compared to the Roman scripts. Following are the major differences that make Gurmukhi language cumbersome to handle as compared to Roman scripts [2].

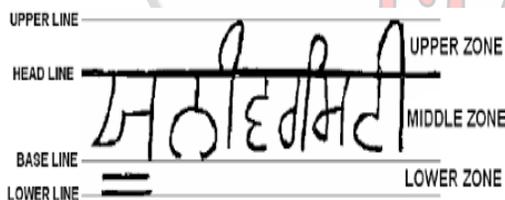


Fig. 4 Basic Structure of Gurmukhi Text

1. Gurmukhi word is represented in three different zones i.e. lower zone, middle zone and upper zone.
2. As all the Roman characters reside only in the middle zone, therefore, minimum height is fixed according to the height of the characters, but this cannot be done for Gurmukhi script because of presence of characters in all the three zones [5].
3. No vertical inter character gap exists between the letters of the word, because most of the Gurmukhi characters are connected to a headline. The headline is a horizontal line at the upper part.

4. Roman word can form a separate component (e.g. the individual characters of the word 'PUNJABI' are separated as ('P','U','N','J','A','B','I'), due to the inter character vertical gap present in the letters of English language. But Gurmukhi word can be considered as one single component because inter character vertical gap is not present for e.g. ('ਪੰਜਾਬੀ').
5. In Gurmukhi script, some vowels and half vowels may appear in different zones that can be upper zone, middle zone or lower zone only or some may overlap higher and middle zones. For example in the word "ਜੂਨੀਵਰਸਿਟੀ", sihari and bihari are overlapped.

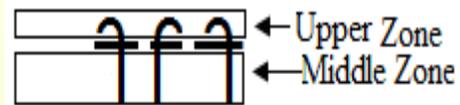


Fig. 5 Vowels Overlapping Upper and Middle Zone

V. CONCLUSION

Textual images of our surrounding contain very meaningful information that is required to be translated into another person's understandable language for their better understanding. Images captured by mobile's camera suffer from certain shortcomings such as blurring, noise, low resolution, and perspective distortion etc. Extraction of text present in natural scene images becomes more challenging as natural scene images often have cluttered backgrounds and multiple patterns for scene text. So this is a challenging area that still needs attention of the researchers.

REFERENCES

- [1] Q. Ye and D. Doermann, "Text Detection and Recognition in Imagery: a Survey," *TPAMI, IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. c, pp. 1480-1500, 2015.
- [2] A. Kaur, R. Dhir, and G. S. Lehal, "A survey on camera-captured scene text detection and extraction: towards Gurmukhi script," *Int. J. Multimed. Inf. Retr.*, vol. 6, no. 2, pp. 115-142, 2017.
- [3] E. Kim, S. Lee, and J. Kim, "Scene text extraction using focus of mobile camera," *Proc. Int. Conf. Doc. Anal. Recognition*, pp. 166-170, 2009.
- [4] T. E. de Campos, B. R. Babu, and M. Varma, "Character Recognition in Natural Images," *Visapp (2)*, pp. 273-280, 2009.
- [5] S. Singh, D. Sharma, "Gurmukhi Text Detection and Localization in Natural Scene Images," *ijarcse*, vol. 5, no. 12, pp. 426-428, 2015.