TEXT DECTECTION IN NATURAL IMAGES

- Kedarnath P (1501029)

Indian Institute of Information Technology, Guwahati

Department of Computer Science Engineering

(under the guidance of Dr.Nilkanta Sahu)

March 9, 2017

Abstract

Text extraction in images is an important research field in visual content understanding and retrieval, automatic explanation and structuring of images. Text extraction from image involves detecting the text from given image, finding the presence of text location, extraction, enhancement and recognition of text from the given image. Text detection in natural images has gained much attention in the last years as it is a primary step towards text recognition as well as it describes the content of an image. It needs to be accurate, efficient and robust against the variations of text in images. A lot of work has been done for detecting text in images and a lot has to be done. A large number of techniques have been proposed to address this problem and the purpose of this paper is to classify and review various text extraction algorithms and discuss challenges for future research.

1 INTRODUCTION

Text Extraction from image is concerned with extracting the relevant text data from a collection of images. As stated in rapid advancement of digital technology has resulted in digitization of all categories of materials. Lot of resources is available in electronic medium. Many existing paper-based collections, books, journals, etc. are converted to images. These images present many challenging research issues in text extraction and recognition. As stated by Jung, Kim and Jain in, text data is becoming region of interest, because text can be used to easily and clearly describe the contents of an image. Since the text data present in image or video in different variations, the problem of extracting the text region from images becomes a challenging one. Among them, text within an image is of particular interest as (i) it is used to describe the contents of an image; (ii) it can be easy to extract as compared to other semantic contents, and (iii) it enables applications such as keyword-based image search text-based image indexing etc.

A. Text in Images

A large number of approaches to text information extraction (TIE) from images have been proposed for specific applications and including page segmentation, address block location, license plate location, and content-based image/video indexing. As stated in text in images can be classified into three categories: documented text, caption text(also known as overlay text) is artificially superimposed on the video/image at the time of editing and it usually describes the subject of the image/video content. While, other type is scene text that appears within the scene is captured by the recording device. It is difficult to detect and extract scene text then caption text as it may appear in a virtually unlimited number of variations in font style, size, orientation, alignment and background complexity.

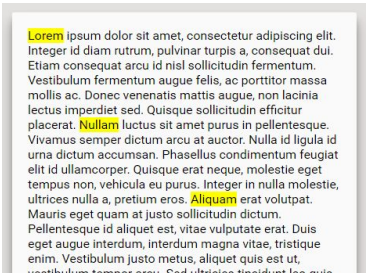


fig 1a: Documented text image

fig 1b: Colored text image

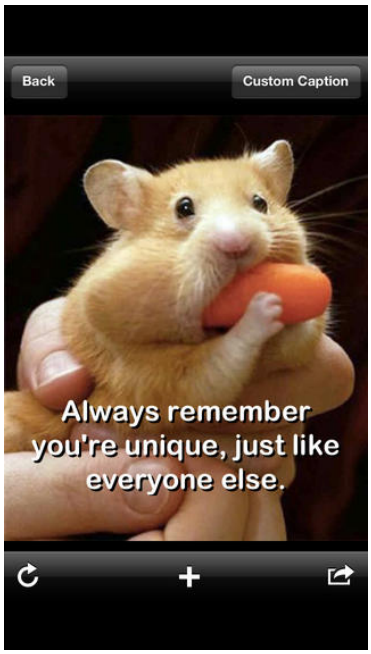


fig 2: Caption text image

fig 3: Scene text image

B. Properties of Text in Images

Due to the variations in geometrical properties of text and background complexities the problem of automatic text extraction extremely complicated and difficult. Text in images can exhibit many properties with respect to the following geometric properties:

- Size: The text size can be of variable size.
- Alignment: The caption texts appear in clusters and usually lie horizontally. This

does not apply to scene text, which has various perspective distortions. Scene text can be aligned in any direction.

- Edge: An edge in the images is the most reliable feature of text regardless of color/intensity, orientations, etc.

- Color: In a simple image the characters in a text usually have the same or similar colors. This property makes it possible to use a connected component-based approach for text detection. However, video images and other complex color documents usually contain text strings with more than two colors.

The rest of the paper is organized as follows: section 2 lists few applications ,Section 3 describes Text Information Extractions (TIE). Recent techniques for text information extraction are reviewed in Section 4. Section 5 describes various approaches, Section 5.1 presents the implemented approach, and Sections 6 and 7 present challenges,conclusion and future work respectively.

2 APPLICATIONS

There are numerous applications of a scene text information extraction system, including:

- vehicle license plate extraction.

- visual search system have been developed for applications such as product recognition,landmark recognition.

- aids for visually impaired people

- translators for tourists .

- electricity meters reading.

3 TEXT INFORMATION EXTRACTION(TIE)

A Text Information Extraction system (TIE) receives an input image and output the relevant text data. As stated in the images can be in gray scale or color, compressed or uncompressed. The Text Information Extraction (TIE) as shown in Fig.4. includes the following stages as follow:

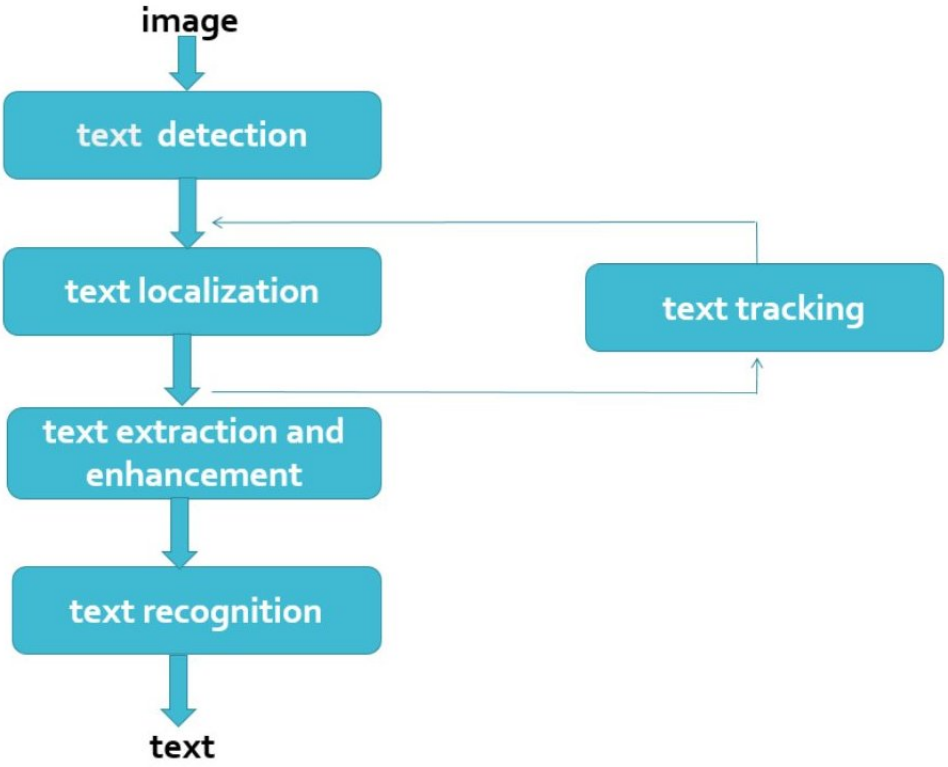


Fig.4 Text Information Extraction (TIE)

- *Text detection* refers to the process of determining the presence of text in a given image/frame.
- *Text localization* refers to the determination of the location of text in the image/frame and generating bounding boxes around the text.
- *Text tracking* is used to reduce the processing time for text localization.
- *Text extraction* is the stage where the text needs to be segmented from the background to facilitate its recognition.
- *Text Enhancement* is required because when the text region is extracted from the background it usually has low resolution and is more likely to suffer from noise. Thereafter, the extracted text images can be transformed into plain text using OCR technology.

In spite of extensive studies, it is not easy to design a general-purpose Text Information Extraction (TIE) system. This is because there are many possible sources of variation when text is being extracted from the complex background, from low contrast or complex images or from the images having variations in geometric properties of text. These variations make the problem of automatic Text Information Extraction (TIE) extremely difficult.

4 TEXT EXTRACTION TECHNIQUES

Text extraction process mainly consists of five important phases: text region detection, text localization, tracking, character extraction, text recognition. From which first two

(text region detection, text localization) stages are more important and also they are more difficult to implement. The output of text information extraction is mainly dependent on the output of these two phases. According to the features utilized, the techniques used for text information extraction falls in three categories as follows:

A. *Region-Based Techniques*

Region-based methods use the properties of the color or gray-scale in a text region or their differences with the corresponding properties of the background. This method uses a bottom-up approach by grouping small components into successively larger components until all regions are identified in the image. Therefore, filter out non-text components and mark the boundaries of the text regions.

B. *Edge-Based Techniques*

An edge in the images is the unique features of text regardless of color/intensity, orientations, etc. Edges are considered as the distinguishing characteristics of text embedded in images, as it is the main features for detecting text. Edge-based text extraction algorithm is a general-purpose method, for effectively localize and extract the text from both indoor/outdoor images.

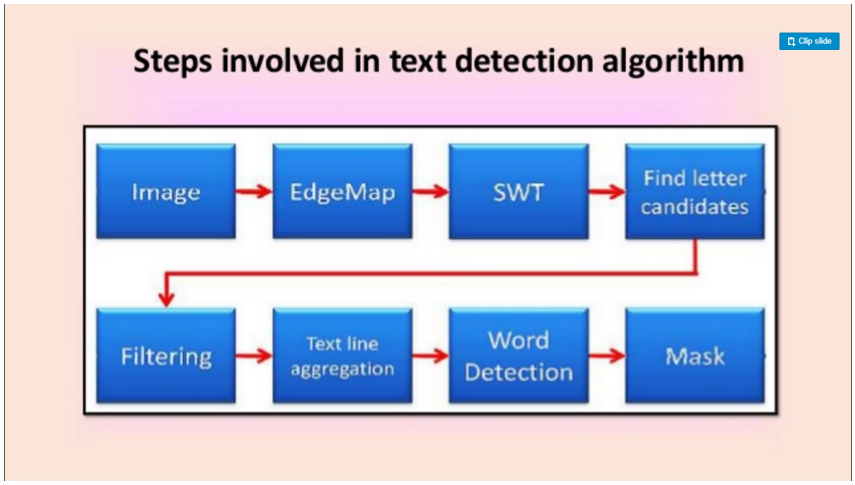
C. *Texture-Based Techniques*

Texture-based methods use the observation that text in images has distinct textural properties that differentiate them from the background. The methods based on Gabor filters, Wavelet, etc. can be used to detect the textural properties of a text region in an image.

5 VARIOUS APPROACHES

Various methods are used for the extraction of text from colored journal images, camera captured images, video images, printed document, degraded document images, handwritten historical document, graphical and color document images, low resolution images, book cover and web pages.

SWT method proposed by **Epshtein et.al**[4] is a *region-based* text detection method. It follows the stroke width constancy assumption, which states that stroke widths remain constant throughout individual text characters. After obtaining an edge map of an input image, SWT method locates pairs of parallel edge pixels in the following fashion: for each edge pixel p a search ray in the edge gradient direction is generated, and the first edge pixel q along the search ray is located. If p and q have nearly opposite gradient directions, an edge pair is formed and the distance between p and q (called stroke width) is computed. All pixels lying on the search ray between p and q (including p and q) are assigned a corresponding stroke width. After assigning stroke widths to all image pixels, the SWT method groups pixels with similar stroke widths into connected components and filters out those that violate geometrical properties of the text. When the edge threshold is sufficiently low, SWT typically finds all characters in the image or at least small portions of each of them. However, it often fails to detect whole characters and leaves parts of them undetected . Another SWT drawback is the detection of non-text structures with nearly parallel edges.



P. Nagabhushan et.al [5] proposed a *edge-based* approach to extract the text in complex background color document images. The proposed method used canny edge detector to detect edges. When dilation operation was performed on edge image, it created holes in most of the connected components that corresponds to character strings. Connected components without hole(s) were eliminated. Other non-text components were eliminated by computing and analyzing the standard deviation of each connected component. An unsupervised local thresholding was devised to perform fore-ground segmentation in detected text regions. Finally the noisy text regions were identified and reprocessed to further enhance the quality of retrieved foreground.

Mao et al. [6] proposed a *texture-based* text localization method using Wavelet transform. Harr Wavelet decomposition is used to define local energy variations in the image at several scales. Binary image, which is acquired after thresholding the local energy variation, is analyzed by connected component-based filtering using geometric attributes such as size and aspect ratio. All the text regions, which are detected at several scales, are merged to give the final result.

5.1 IMPLEMENTED APPROACH

We propose an *edge-enhanced MSER* text detection algorithm referred from [7](shown in Fig.5).At the input of the system, the image intensities are linearly adjusted to enhance the contrast. Subsequently, MSER regions are efficiently extracted from the imageand enhanced using Canny edges obtained from the original gray-scale image. As a next step, the resulting CCs are filtered using geometric constraints on properties like aspect ratio and number of holes. The stroke width information is robustly computed using a distance transform and objects with high variation in stroke width are rejected. Text candidates are grouped pairwise and form text lines. Finally, words within a text line are separated, giving segmented word patches at the output of our system.

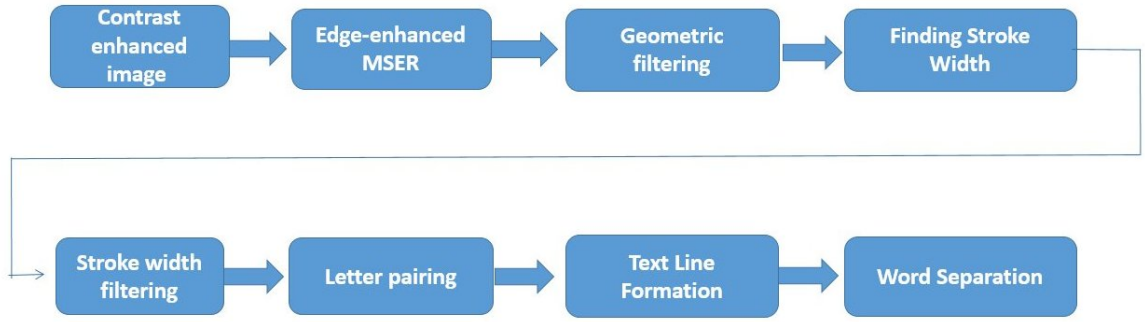


fig5: edge-enhanced MSER flowchart

6 CHALLENGES

Although a lot of approaches have been developed on text detection in real applications as discussed in section 5. But a fast and robust algorithms for detecting text under various conditions need to be further investigated. To develop a fast and robust text detection algorithm is a nontrivial task since there exists such difficulties as:

- Text may be embedded in complex background.
- It is difficult to find effective features to discriminate text with other text-like things, such as leaves, window curtains or other general textures.
- Text pattern varies with different font-size, font-color and languages.
- Text quality decreases due to noise.
- It is difficult to detect text of arbitrary orientations.

7 CONCLUSION AND FUTURE WORK

A large number of techniques have been proposed in the past but the detection of scene text with high precision and recall rate is still a challenging problem because of additional complexities such as varying lighting, variable font sizes, style, color, variance of orientation and complex background. The purpose of our paper is to classify and review various approaches and to point out challenges for future research.
Future work is to get the path of bounding boxes and correctly align the text of different orientation in natural scene images horizontally to improve recognition rate. In recognition phase, classification errors can be found and those errors can be caused due to ambiguous characters, such as [L, I], [O, D], [h, n], [e, c] etc. Therefore, further improvements can be made to recognize these characters correctly, so that accuracy can be increased.

References

[1] C.P.Sumathi,T.Santhanam and G.Gayathri Devi, “A survey on various approaches of text extraction in images”, Proceedings of International Journal of Computer Science Engineering Survey (IJCSES), Vol. 3, pp. 27-42, August 2012.

[2] Partha Sarathi Giri, “Text information extraction and analysis from images using digital image processing techniques”, Proceedings of Special Issue of International Journal on Advanced Computer Theory and Engineering (IJACTE), Vol. 2, pp. 66-71, 2013.

[3] Y.Y.Tang, S.W.Lee and C.Y.Suen, “Automatic document processing: a survey”, Pattern Recognition, pp. 1931-1952, 1996.

[4] Boris Epshtein, Eyal Ofek, Yonatan Wexler, "Detecting Text in Natural Scenes with Stroke Width Transform".

[5] P. Nagabhushan, S. Nirmala(2009) ,”Text Extraction In Complex Color Document Images For Enhanced Readabi.

[6] W. Mao, F. Chung, K. Lanm, and W. Siu, Hybrid Chinese/English Text Detection in Images and Video Frames, Proc. of International Conference on Pattern Recognition, 2002, Vol. 3, pp. 1015-1018.

[7] Huizhong Chen, Sam S. Tsai, Georg Schroth, David M. Chen, Radek Grzeszczuk and Bernd Girod1,"Robust text detection in Natural Images with edge-enhanced Maximal stable extremal regions".Image Processing (ICIP), 2011 18th IEEE International Conference on. IEEE, 2011.

[8] D. Nist´er and H. Stew´enius, “Linear time maximally stable extremal regions,” in ECCV, 2008, pp. 183–196.

[9] D.S.Kim and S.I.Chien, “Automatic car licence plate extraction using modified generalized symmetry transform and image warping”, Proceedings of International Symposium on Industrial Electronics, Vol. 3, pp. 2022-2027, 2001.

[10] Adrian Canedo and Jung H. Kim” English to Spanish Translation of Signboard Images from Mobile Phone Camera” SOUTHEASTCON 2009 IEEE.