Survey on Text Detection, Segmentation and Recognition from a Natural Scene Images

Uma B. Karanje
Department of Computer Engineering,
Marathwada Mitra Mandals College of
Engineering, Pune.

Rahul Dagade
Department of Computer engineering,
Marathwada Mitra Mandals College of
Engineering, Pune.

ABSTRACT

Detecting text from an image is an important prerequisite for the content based image analysis process. To understand the contents of an image or the valuable information, there is need of analyzing the text appears in it. Various methods have been proposed over past years for text detection and extraction from different types of images, like scene image, born digital image and document image.

In this paper, we describe the existing methods of text detection, text segmentation and character recognition from natural scene images with their features, advantages and disadvantages.

General Terms

Pattern Recognition

Keywords

Text detection, text segmentation, character recognition, scene image

1. INTRODUCTION

In recent years, use of multimedia technology has increased tremendously. In multimedia technology image is one of the important part and image can have different contents in it, such as face, human, scene, text, etc. Among all contents in images, text is found to be one of the most important features to understand the image contents. Text in images can be used as indexing purpose.

The text information can be extracted in two stages : text detection and text recognition. Text detection detects the text regions as extremal regions of an image and in text recognition stage system retrieves the text information from these extremal regions[8].Retrieving the contents from images is very challenging because of image quality and background noise.

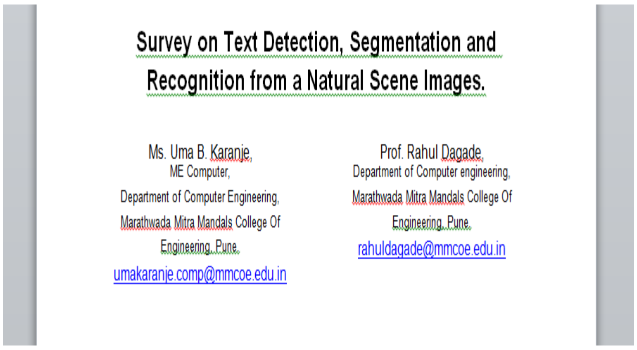
There are different kinds of images that have text as its part with background, such as document images, scene images and born-digital images. In which scene images are often taken by cameras. The digital cameras and camera phones enables acquisition of image and video materials containing scene text like street signs, advertisements, billboards, or restaurant menus, but these devices also introduce new imaging conditions such as sensor noise, viewing angle, blur, variable illumination, uneven lighting, lower resolution, etc.

Considering above problems and scene text properties, natural scene text detection and recognition is more difficult task in comparison with text in born-digital and document image.

The paper is structured as follows : Section 2 defines various types of images that contains text in it, with different factors that should be considered while detecting text from it. Section 3 defines various methods used for detecting, segmenting and recognizing text from the natural scene images. Section 4 gives some applications where text detection system is useful.

2. IN GENERAL, TEXT IMAGES CAN BE DIVIDED INTO THREE TYPES

A) Document images: Document images are nothing but image-format of the document[1]. Document images can have text and graphics. This type of images are generated by scanners or camera phones, which acquire printed documents, historical documents, handwritten documents, books, etc.[2]. In which, the image is transformed from paper-based documents into image-format for electric read. In the early stage of text extraction, there is only focus on document images.



Survey on Text Detection, Segmentation and Recognition from a Natural Scene Images.

Ms. Uma B. Karanje,
ME Computer,
Department of Computer Engineering,
Marathwada Mitra Mandals College Of
Engineering, Pune.
umakaranje.comp@mmcoe.edu.in

Prof. Rahul Dagade,
Department of Computer engineering,
Marathwada Mitra Mandals College Of
Engineering, Pune.
rahuldagade@mmcoe.edu.in

Figure 1. Document text image

B) Scene images: Scene images contain the text, such as the advertising boards, banners, which is captured by the cameras; therefore scene text appears with the background part of the scene[1]. These types of images are very challenging to detect and recognize, because the backgrounds are complex, containing the text in different sizes, styles and alignments. Also, scene text is affected by lighting conditions and perspective distortions. The current OCR software cannot handle complex background interferences and non-orienting text lines.

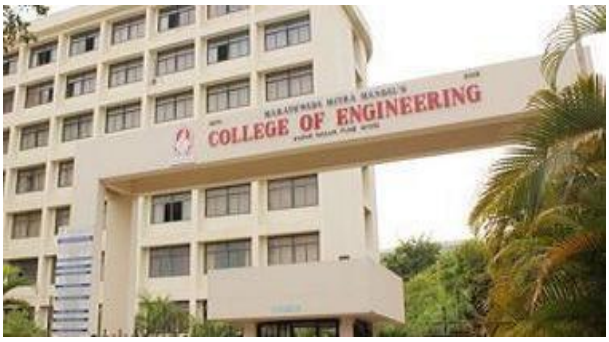


Figure 2. Scene text image

C) Born-digital images: Born-digital images are generated by computer software and are saved as digital images. Compared with document images and scene images, there are more defects in born digital images, such as more complex foreground/background, low resolution, compression loss, and severe edge softness. Therefore, during text extraction, it is difficult to distinct the text from the background[1].

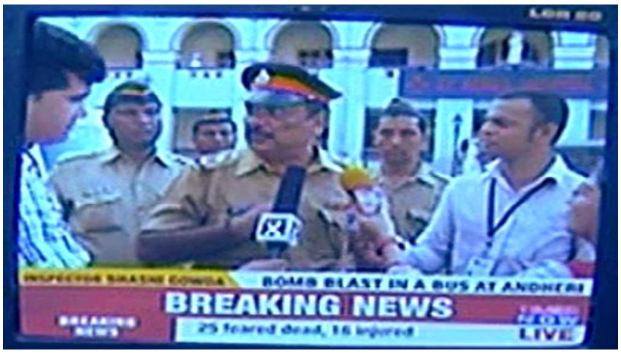


Figure 3. Born-digital text image

D) Heterogeneous text images : This image have all kinds of images such as scene text images, caption text, document image and born-digital image[2].

2.1 Factors that should be considered while Detecting Text from an Image

a) Font style, size(height, width) and thickness (stroke width);

b) Co-ordinates (X, Y) or position in image;

c) Background as well as foreground color and texture;

d) Camera position which can introduce geometric distortions;

e) Orientation;

f) Alignment;

g) Symbols, integers and non-text contents;

h) Illumination ;

i) Language ;

j) Resolution ;

k) Contrast;

l) Blur and noise.

These contents are related to the textual information appearing in them, which can be divided into two groups [5] :

1. The text appeared in image that does not represent any important contents related to image, that referred as scene text.
2. The text which produced separately from the image is good key to understand the image, is called as an artificial text.

In contrast to scene text, artificial text is not only an important source of information but also a significant entity for indexing and retrieval purposes. So it is very challenging task to detect, segment, recognize and retrieve text from an image with accuracy and robustness of the image contents.

3. THE CURRENT RESEARCH ON SCENE IMAGES

Many methods for text detection from scene images have been proposed over the past years; in this section, we will briefly review methods for text detection, text segmentation and character recognition with their advantages and disadvantages.

3.1 Existing Methods for Scene Text Detection

3.1.1 Sliding window based methods, also known as region-based methods. This method uses a sliding window to search for possible texts in the image and then use machine learning techniques to identify the text.

Disadvantages: These methods are slow, because the image has to be processed in multiple scales. These methods limits the search of text to a subset of image rectangles. So, it reduces the number of subsets checked for the presence of text [11].

3.1.2 Connected component (CC) based methods, extracts the character candidates from an image by connected component analysis, followed by grouping character candidates into text; additional checks may be performed to remove false positives.

Advantages: Achieves state of the art result. The complexity does not depend on the properties of text like orientation, font style etc. [11].

Disadvantages: Fails in some natural scene images which have very poor contrast text and strong illumination.

The methods for drawing connected components from images can be categorized in three groups:

A) Method based on Edges [1]:

This method is based on the factor like edge of character; edge is reliable feature of the text regardless of color/intensity, layout, orientations, etc. As the text region has high contrast to it's background, the edges of character can be easily detected. There are two steps used in this method: first, an edge extraction algorithm (such as canny edge detector) is used to get the edges and second, smoothing algorithm or morphology is used for edges connections and obtaining a full character boundary.

The main disadvantage of this method is that small image regions and stroke may be misidentified. Therefore this method needs to be verified using other methods.

B) Method based on Color [1]:

In this method, color clustering is done by categorizing the pixels with same or similar colors and forming a candidate region. Then the candidate regions are analyzed and the CC is estimated.

The main challenge of this method is the degree of clustering. If the data is over clustered, the background and text region may be mixed together. And if the data is under clustered, the number of clustering will be increased and the system performance will be degraded.

C) The Method based on the Combining of Edges and Color [1]:

Some methods combine Method 1 and Method 2, which detects both edges and colors of the text. This method has achieved better results by combining both features together than using these features separately.

3.1.3 Hybrid methods, this method uses a region detector to detect text candidates and extracts connected components as character candidates by local binarization; non-characters are eliminated with a Conditional Random Fields model, and characters can finally be grouped into text.

Advantages: Region based information is very helpful for text component segmentation and analysis; The CRF model differentiates text components from non-text components better than local classifiers.

Disadvantages: Hard-to-segment the texts.

3.1.4 Method based on Texture [1]:

This method deals with text regions as a special texture. The region is identified as text region or not according to the extracted relevant texture of the candidate regions. To overcome the disadvantages mentioned above, hybrid approach is presented, which takes the advantages of both texture-based and CC-based methods, to robustly detect and localize texts in natural scene images. In this method, a text region detector is designed which is based on the texture. This can be used to estimate the probabilities of the position and the scale of the text and then it is analyzed to be text region or not.

3.1.5 Method based on Corner [1]:

This approach is inspired by the observation that the characters in the text, usually contains multiple corner points. The method is to describe the text regions formed by the corner points using several discriminative features. The research on the method based on corners is still in the early stage. Compared with texture based method, this method is faster but the performance is less satisfied.

3.1.6 Method based on Semiautomatic Ground Truth Generation [6]:

The semiautomatic ground truth generation system for text detection and recognition includes text with different orientation and language. In this method, the system allows user to manually correct the ground truth if the automatic method produces incorrect results. This method uses eleven attributes at the word level, namely: line index, word index, coordinate values of bounding box, area, content, script type, orientation information, type of text (caption/scene), condition of text (distortion/distortion free), start frame, and end frame to evaluate the performance of the method.

3.1.7 Method used to Querying Images using Scene Text [7] :

Here two stages are used as text localization and text verification. Text localization consists of two steps: pre-processing and generation of candidate text regions. Next, edge features and morphological dilation are employed to locate image blocks. Then Stroke Width Transform is applied with some modification to generate candidate letters. These letters are paired to identify text lines, which are subsequently separated into words. At text verification stage, Histogram of Oriented Gradient (HOG) features used to train a Support Vector Machines (SVM) based classifier to determine whether a candidate word is text or not. Then the text regions are extracted by a novel binarization algorithm.

3.1.8 Method based on Extremal Region of Character [11] :

In this method, character candidate are extracted by calculating the probability of each extremal region being a character and ERs with locally maximal probability are selected.

3.2 Existing Methods for Scene Text Segmentation

In the existing research, A. Mishra defined the pixels in a document image as random variables in an Markov Random Field (MRF) and introduced a new energy (or cost) function on these variables, each of which takes a foreground or background label, and the quality of the binarization is determined by the value of the energy function. Then the energy function is minimized, i.e. to find the optimal binarization using an iterative graph cut scheme[1].

M.S.Cho defined that single feature of the text, such as the color, edge and stroke is not useful for text extraction, by combining these features and relationships together for creating the Conditional Random Field (CRF) framework it is beneficial[1].

X.F.Wang proposed a novel method for embedded text segmentation. This method is based on two assumptions of embedded texts: i) the color of text pixels follows Gaussian distribution; ii) the local part of the embedded text has the same color distribution with the global part. By these two assumptions, he developed a two-step text segmentation approach: in the first step, coarse segmentation step, he utilized a 1-D Gaussian function to create a model for the color distribution of text pixels. Then the confident text region was extracted using a stroke operator to acquire the model parameters and the parameters are estimated from a developed heuristic process[1].

T.Wakaharad generated multiple binary sub-images after K-mean clustering in HIS image space and calculated the probability of character images from the coarse segmented sub-images based on the network features and SVM classification calculation[1].

3.3 Existing Methods for Scene Character Recognition

Qi Zheng proposed recognition of characters using segmentation algorithm with multiple-size sliding sub-windows. The customized program generate template images first and the extracted SIFT features are matched to the template images. After using the segmentation algorithm multiple single-character-areas are identified and the results are verified by a voting and geometric verification algorithm [1].

Masakazu Iwamura, propose a character recognition method using local features with several desirable properties. In this method, arrangement of local features is important to recognize multiple [1].

Tu et al., used insights from natural language processing and present a Markov chain framework for parsing images [12].

Jin and Geman,introduced composition machines for constructing probabilistic hierarchical image models which accommodate contextual relationships. This approach allows re-usability of parts among multiple entities and non-Markovian distributions [12].

Weinman and Learned Miller, proposed a method that fuses image features and language information (such as bi-grams and letter case) in a single model and integrates dissimilarity information between character images [12].

3.4 Comparison of different Text Detection Methods based on the Maximally Stable Extremal Regions

Maximally Stable Extremal Region(MSER) is one of recent method used for text detection from natural scene images. It gives advantages over other region detectors like Harris-affine, Hessian-affine, edge-based regions, intensity extrema, and salient regions. MSER detects near about 2600 regions for image and gives greater results in region size, blur, viewpoint change, scale change and light change.

MSER based methods are categorized under connected component based methods. MSER-based methods have reported promising performance on the widely used ICDAR 2011 Robust Reading Competition database [4].

Table below gives comparison of various text detection methods based on MSER with its advantages and disadvantages.

Table 1. Various MSER methods with advantages and disadvantages

Method	**Advantages**	**Disadvantages**
MSER ++ [4]	Exploits higher order properties of text, Uses exhaustive search for pruning.	Absence of an effective text candidate construction algorithm.
2 Stage Algorithm for ERs Pruning [4]	Estimate class conditional probabilities of ERs using trained classifier.	Absence of an effective text candidate construction algorithm. Requires Tuning parameters.
MSER as Character Candidate [4]	Detects char in low quality, low resolution, strong noises, low contrast.	Most of detected MSERs are repeating with each other.
Graph cut model with MSER [14]	non-text MSERs are efficiently removed	some text are not detected as MSERs due to the instability of the color caused by the disturb of the illumination condition

4. APPLICATIONS

Text detection, segmentation and extraction from complex images can be applied to a variety of fields where the information needs to be analyzed and understood. Some of these applications are given below:

1) ***Image understanding:*** When images can be automatically understood and indexed by computer, the efficiency of running digital libraries and video database system will be greatly improved[1].

2) ***Content-based image filtering:*** In content based filtering, image spam can be detected and pornography, reactionary and fraud words can be easily filtered[1].

3) ***Super map:*** Text extraction technology can be applied to detect scene text from images taken with laptops, phones and other equipments, so as to be applied to maps, navigation, automatic translation, foreign-related tour guides, walking robots and intelligent monitoring system[1] and also used as visual impaired peoples assistance[9].

4) ***Vehicle testing:*** Vehicle license and scene subtitles have many features in common, so text extraction can be used to supervise the traffic in real time. After text extraction from highway video flow, the traffic situation can be overseen and vehicle licenses can be recognized easily from traffic accidents, which can improve the efficiency of the transportation systems[1].

5) ***Optical character reading:*** Reads text from paper and translates images into a form that computer can manipulate (for example, into ASCII codes). An OCR system enables to take a book, feed it directly into an electronic computer file, and then edit the file using a word processor.

6) ***Automatic localization of postal addresses on envelopes and Automatic Geo coding:*** Postal automation tries to get the mail from the sender to the recipient quickly, in a reliable and economical process.

7) ***Text extraction in video sequences:*** Caption text or superimposed text provides valuable information about contents in images and video sequences.

8) ***Wearable applications:*** Wearable devices such as goggles, phones, cameras are created for detecting text elements and can be converted into voice for blind peoples[13].

9) ***Online electric goods search:*** Online shopping applications using mobile phone allows customer to type the name of goods and get required information about it with images and discriptions[13].

5. CONCLUSION

This paper presents review on existing methods for text detection, segmentation and recognition with their feature. Also this paper summarizes the key ideas, advantages, disadvantages and applications of text detection technique. Detecting and recognizing text from natural scene image is more difficult task than all other types of images. It have various affecting factors like light effects, orientation, font styles, blur, etc. Even though there are many algorithms, no single unified approach can fits for all the applications. So there is lot of scope to work with the text detection, extraction, segmentation and recognition from natural scene images. Also there is scope for detecting text from various languages, which have different characteristics than English.

6. REFERENCES

[1] Jian Zhang, Renhong Cheng, Kai Wang, Hong Zhao, “Research on the text detection and extration from complex images”, Fourth International Conference on Emerging Intelligent Data and Web Technologies. Vol. 10, 2013, Page no. 708-713.

[2] C.P. Sumathi, T. Santhanam, G.Gayathri Devi, “A Survey On Various Approaches Of text Extraction In Images”, International Journal of Computer Science & Engineering Survey (IJCSES). Vol.3, August 2012, Page no. 27-42.

[3] Datong Chen, Juergen Luettin, Kim Shearer, “A Survey of Text Detection and Recognition in Images and Videos”. Institute Dalle Molle d’Intelligence Artificielle Perceptive Research Report, August 2000, Page no. 00-38.

[4] Xu-Cheng Yin, Xuwang Yin, Kaizhu Huang, and Hong-Wei Hao, “Robust Text Detection in Natural Scene Images”, IEEE transaction on Pattern Analysis And Machine Intelligence, 2013, Vol. 36, Page no. 970 – 983.

[5] K. Wang, B. Babenko, and S. Belongie, “End-to-end scene text recognition”. International conference on computer vision ICCV 2011, vol. 10, Page no.1457 – 1464.

[6] Trung Quy Phan, Palaiahnakote Shivakumara, Souvik Bhowmick, Shimiao Li, Chew Lim Tan, Umapada Pal, “Semiautomatic Ground Truth Generation for Text Detection and Recognition in Video Images”. IEEE transactions on circuits and systems for video technology, VOL. 24, NO. 8, AUGUST 2014, Page no. 1277-1287.

[7] Thuy Ho, Ngoc Ly, “A Scene Text-Based Image Retrieval System”, IEEE international symposium on Signal Processing and Information Tech., pp. 79-84, 2012.

[8] Xiaobing Wang, Yonghang Song, Yuanlin Zhang, “Natural scene text detection in multi-channel connected component segmentation”, 12th International conf. on Document Analysis and Recognition, pp. 1375-1379, 2013.

[9] Shehzad Muhammad Hanif, Lionel Prevost, “Text Detection and Localization in Complex Scene Images using Constrained AdaBoost Algorithm”, 10th International Conference on Document Analysis and Recognition, pp.1-9, 2009.

[10] Teofilo E. de Campos, Bodla Rakesh Babu, Manik Varma, “Character Recognition In Natural Images”, International conf. on Intelligence Science and Big data Engg., pp. 193-200, 2011.

[11] Lukas Neumann, Jirı Matas, “Real-Time Scene Text Localization and Recognition”, IEEE Conf. on Computer Vision and Pattern Recognition, 2012, pp. 3538–3545.

[12] Teofilo E. de Campos, Bodla Rakesh Babu, Manik Varma, “Character Recognition In Natural Images”, International conf. on Intelligence Science and Big data Engg., pp. 193-200, 2011.

[13] Honggang Zhang, KailiZhao, Yi-ZheSong, JunGuo, “Text extraction from natural scene image: A survey", Elsevier journal on Neurocomputing ,pp.310-323, 2013.

[14] Cunzhao Shi, Chunheng Wang, Baihua Xiao, Yang Zhang, Song Gao, “Scene text detection using graph model built upon maximally stable extremal region". vol 34, issue 2, 2013, page no. 107-116.