# A Survey of Text Reading from Scene Images

Poonam B. Kadam
Dept of Computer Engineering
D.Y.P.I.E.T., Pimpri, Pune, India

Latika R. Desai
Dept of Computer Engineering
D.Y.P.I.E.T., Pimpri, Pune, India

## ABSTRACT
Text reading in images or video is important step to achieve content retrieval from images. The content retrieve from image or video content useful information it act as clue for many image based applications such as scene understanding, content based image retrieval text based image indexing, industrial automation. Locating or detecting text from complex background is a challenging task due to variation in size, orientation, style. This paper describes the techniques which try to solving this problem.

## Keyword
Scene image text detection, text localization, text recognition.

## 1. INTRODUCTION

Although many research done in historical document collection analysis and recognition has focused on detection and analyzing scanned document. Obstruct readability and decrease the performance because of large degradation in document text images. Documents text reading is become difficult due to aging of document, poor quality of ink, physical deterioration, blur. To avoid the researches must be choose advance techniques for historical document image text analysis. This document images text analysis not limited to only historical document analysis one can have license number plate image, street name, sign recognition and translation, electricity meters and so many area where text reorganization is essential.

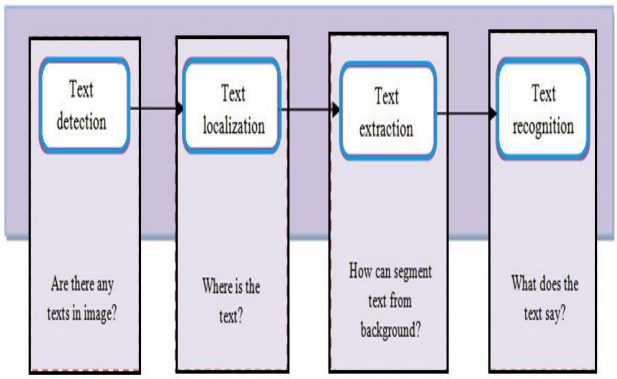


**Figure 1.Text information extraction system**

Already many researches done on scene image text detection and recognition and number of approaches for scene text detection /localization/ recognition tackle many problems (blur, unknown layout, low resolution, and multicolored character, variation in color, font, and size) it has mostly used in content based image search. As shown in Figure 1 text understanding system consist of four stages: text detection, text localization, text extraction, text recognition.

### 1.1. What is texts extraction from images?

Before precede further needs to understand problem of text information retrieval. This system receives an input in the form of an image or video. The images can be in color or gray scale, un-compressed or compressed, orientation. This problem can be divided into the following sub-problems: (i) detection, (ii) localization, (iii) extraction and enhancement, and (iv) recognition (OCR) (see Figure 1.)

We can use text detection, localization, and extraction interchangeably. In this paper differentiate between these terms. Text detection consist of determination of the presence of text in a given images. Text localization is the process of determining the location of text in the image. Although the specific location of text in an image can be getting, the text still needs to be segmented from the background to make possible its recognition. In stage of text extraction where the text components in images are segmented from the background. The text region usually has low-resolution that's why enhancement of the components is essential. After, the extracted text images can be converted into plain text using OCR technology.

## 2. TEXT EXTRACTION STAGES

As described in the previous section, TIE divided into four sub-stages: detection, localization, extraction and enhancement, and recognition. This section will review this sub-stage.

### 2.1. Text detection

Unless characters are expected to appear at pre-defined locations, the text must somehow be located. This involves a search for text in a binarized image. Detecting text in images or low resolution video is more difficult, yet these problems have experienced recently with contests sponsored by the International Conference on Document Analysis and Recognition in 2003 [1] and 2005 [2].
In this stage, since there is the existence or non-existence of text in the image must be determined. The text detection stage seeks to detect the presence of text in a image. Because of highlights, shadow, complex background image distortion and degrading accurate and fast text detection in scene images is still challenging task. Table I lists the possible effects on text in scene image.

**Table 1.Effect on text in an image caused by various conditions**.

| Properties of Text / Image in an Images condition | color | size | highlight |
|---|---|---|---|
| Standard Orientation | Yes | Yes | Yes |
| Lighting | Maybe | No | Yes |
| Low resolution | No | Yes | Yes |
| Distance | Maybe | Yes | Yes |

The approach used in [3, 4] uses a support vector machine (SVM) classifier to segment text from an image or video frame. Initially text is detected in multi scale images using edge based techniques, morphological operations and projection profiles of the image [4].These detected text regions are then verified using wavelet features and SVM. The algorithm is robust when variance in color and size of font. Smith et al. [5] text detected by using vertical edges with a predefined template, then grouping vertical edges into text regions using a smoothing process. If a text localization module (Section 2.2) can operate in real time; it can also be used for detecting the presence of text. Zhong et al. [6] and Antani et al. [7] performed text localization on an image, which resulted in a better and faster performance. Therefore, their text localization also used for text detection. The text detection stage is closely related to the text localization stages, which will be discussed in Sections 2.2, respectively.

## 2.2. Text localization

Research on text detection and localization is carried out since from decades and numerous text detection/localization algorithms are proposed we can use it interchangeably also. All these approaches are majorly classified into two categories i) region based (edge-based, connected-component based) and ii) texture based techniques.

### *2.2.1. Region based Methods*

Region-based methods use the properties of the color or gray scale or alignment in a text region or their differences in properties of the background. This method can be further divided into two sub-approaches: edge-based and connected component (CC)-based. Note that some approaches use a combination of both edge-based and CC-based methods.
Although the existing methods have reported promising localization performance, there still remain several problems to solve. For region-based methods, the speed is comparatively slow and the performance is sensitive to text orientation. To overcome this difficulty, some approaches edge-based and CC-based methods to robustly detect and localize texts in natural scene images.

#### *2.2.1.1. Edge-based methods*

The edges of the text boundary are identified and merged, and then several heuristics are used to alter out the non-text regions. Among the several textual properties edge based method focus on ‘high contrast between the background and text. Usually, an edge filter (e.g., a Canny edge detector) is used for the edge detection, and a smoothing operation or a morphological technique is used which are intensive to skew, noise, text orientation. Garcia and Apostolid is [9] use edge detectors and morphological operations to remove noise and fill in dense edged areas. Gao et al. [16, 17] have an approach involving edge detection, color modeling, adaptive search and layout analysis. For superimposed horizontal text in video, the technique of Wolf et al uses combination of horizontal derivatives followed by morphological processing. Kim’s Method [10] this method propose a new text detection algorithm for localizing text region in a mobile phone. The maximally stable extremal regions (MSER) approach used to text extraction and the horizontally neighboring characters are merged of similar sizes and colors. The candidate region detection enables fast and robust text localization, while it also detects a huge amount of non-text regions as candidate regions. In order to minimize the false positives, gradient features obtained from oriented gradient images are used. A cascade classifier is used to discriminate text from non-text. The Adaboost learning method used to localization.
Hasan and Karam [11] uses a morphological approach for text extraction. The RGB components of a color image are combined to give intensity image Y as follows:
Y = 0:30R + 0:59G + 0:12B;
Where R, G, and B are the red, green, and blue components, respectively. Then convert color image to gray scale and after the color conversion, the edges are identified using a morphological gradient operator. The resulting edge image is then threshold.

#### *2.2.1.2. CC-based methods*

Connected component based method use bottom up approach in which grouping of smaller components are done into larger components until all regions are identified in the image. After this a geometrical analysis used to identify text components and group them using the spatial arrangement of the components to localize text regions.CC-based methods are widely used. CC-based methods divided into four processing stages: (i) pre-processing includes noise reduction and color clustering, (ii) CC retrieval, (iii) connected component grouping.
A CC-based method could segment a character into multiple CCs, especially in the cases low-resolution and noisy video images. Kim et al. [12] used cluster-based templates for altering out non-text components for multi-segment component.
A similar approach was also reported by Ohya et al. [13]. Cluster-based approach used along with geometrical information, such as size, area, and alignment. They are constructed using a K-means clustering algorithm from input text images. Jiang et al. [14] took the size and shape of the text characters denoted by connected components (CC) and used SVM learning classifiers to detect text from scene images. Chucai Yi and YingLi Tian[15] this paper use two stages for of text detection .First image partition to find text character components based on local gradient based partition and color clustering and second, connected component grouping to detect text strings based on features of text characters in each text string such as distances between neighboring characters, as character size differences and character alignment.

### *2.2.2. Texture-based methods*

The Texture based method is a feature based approach which involves the construction of gray-level co-occurrence matrix,

Gabor Alters, Wavelet, FFT, spatial variance. This matrix is used to calculate the features like homogeneity, contrast, dissimilarity and which are the results for feature extraction in texture based method. Used to detect the textural properties that distinguish a text region in an image from the background. A texture-based text localization method using Wavelet transform. Harr Wavelet decomposition is used to define local energy variations in the image at different scales. After thresholding the local energy variation we get binary image is analyzed by connected component-based on geometric attributes such as size and aspect ratio. All the text regions, which are detected at several scales, are merged to text from images. Since the utilization of texture information for text localization is also sensitive to the character font size and style, it is difficult to manually generate a texture alter set for each possible situation.

## 2.3. Text Extraction

Text extraction from image or video is used in many applications. Detection of vehicle license number plate [15] it includes two subsystems first extract object (number of a car) after detection and second is recognition of detected object. Text extraction can be use binarization or original image. Some of text extraction algorithms are connected component based or histogram based. In [11] morphological text extraction text region successfully extracted in different orientation, size, with graphical noise.
Based most of the images and videos are now a day's stored, processed, and transmitted in a compressed format, text extraction methods that apply on images in JPEG or MPEG compressed form. Only decoding helps to make algorithm faster. Moreover, the DCT coefficients and motion vectors in a video are also useful in text detection.

## 2.4. Text Recognition

This technique used for converting textual content from an image into machine readable form. The computer actually recognizes the characters in the document through a revolutionizing technique called Optical Character Recognition. Several methods like OCR using correlation method and neural networks. Many previous methods done text recognition in complex background images or video worked on improving the segmentation and preprocessing method which include binarization and gray color before applying an OCR module.

# 3. PERFORMANCE EVALUATION

Computer vision and pattern recognition (CVPR) facing many difficulties in performance valuation in all research areas. The performance measure used for text detection, which is easier to define than for localization and extraction, is the detection rate, defined as the ratio between the number of detected text blocks and all the given block containing text. Performance evaluation of text localization is not simple as compare text detection and recognition.

## 3.1. Performance measures

Performance evaluation performed by using two basic metrics, precision *p* and recall *r*. Here, precision is the ratio of area of the successfully extracted text regions to area of the whole detected region, and recall is the ratio of area of the successfully extracted text regions to area of the regions. The area of a region is the number of pixels inside it. Low precision means overestimate while low recall means underestimate.
To combine *p* and *r*, *f* a standard measure is defined by

$$f = 1/\left(\frac{\alpha}{p} + \frac{(1-\alpha)}{r}\right)$$

Where α represents the relative weight between the two metrics.

## 3.2. Datasets

Using datasets in Robust Reading Dataset1 from ICDAR 2005 and 2011.
Images extracted from different types of HTML documents (Web pages, spam and ham emails)
•Minimum image size: 600 x 450 to 1280 x 960
•Word images (cut-out) provided separately
•Minimum word size: 3 characters
Different ground truth provided for the three tasks
Task 1: Bounding box positions of individual words
Task 2: Pixel-level classification to text / non-text

Task 3: Word images with transcription

We selected 450 images which are compatible with the assumption that a text contains at least three characters with variety of color and fonts with complex background and various orientations.
Dataset contain various complex images which increase performance of system design.

# 4. CONCLUSION

Due to the variety of orientation and complex backgrounds, text reading from natural scene images is still an unsolved problem. To locate text embedded in those images we used text detection, localization, extraction and recognition interchangeably. However, this paper differentiate between these terms. The terminology used in this paper is mainly explained by Antani et al. Text detection refers to the determination of the presence of text in a given frame (normally text detection is used for a sequence of images). Text localization is the process of determining the location of text in the image. Thereafter, the extracted text in an image can be transformed into plain text using OCR technology. There are number of techniques have been proposed to solve this problem, and the purpose of this paper is to classify and discuss these algorithms and performance evaluation, and to point out promising directions for future research.

# 5. REFERENCES

[1] ICDAR 2003 robust reading competitions. In Proc. Intl. Conf. on Document Analysis and Recognition (2003), vol. 2, pp. 682–687.

[2] Lucas, Simon M. Text locating competition results. In Proc. Intl. Conf. on Document Analysis and Recogniton (2005), pp. 80–85.

[3] Qixiang Ye, Qingming Huang, Wen Gao and Debin Zhao, Fast and Robust text detection in images and video frames, Image and Vision Computing 23, 2005.

[4] Qixiang Ye, Wen Gao, Weiqiang Wang and Wei Zeng, A Robust Text Detection Algorithm in Images and Video Frames, IEEE, 2003.

[5] M.A. Smith, T. Kanade, Video skimming for quick browsing based on audio and image characterization, Technical Report CMU-CS-95-186, Carnegie Mellon University, July 1995.

[6] Y. Zhong, H. Zhang, A.K. Jain, Automatic caption localization in compressed video, IEEE Trans. Pattern Anal. Mach. Intell.22 (4) (2000) 385–392.

[7] S. Antani, U. Gargi, D. Crandall, T. Gandhi, R. Kasturi, Extraction of text in video, Technical Report, Department of Computer Science andEngineering, Pennsylvania State University, CSE-99-016, August 30, 1999.

[8] Xiaoqing Liu and Jagath Samarabandu," An Edge-based text region extraction algorithm for Indoor mobile robot navigation", Proceedings of the IEEE, July 2005.

[9] Garcia, C., and Apostolidis, X. "Text detection and segmentation in complex color images". In Proc. Intl. Conf. on Acoustics, Speech, and Signal Processing (June 2000)

[10] J. Park, G. Lee, E. Kim, J. Lim, S. Kim, H. Yang, M. Lee and S. Hwang, "Automatic detection and recognition of korean text in outdoor signboard images," Pattern Recognition Letters, vol. 31, no. 12, pp. 1728–1739, Sep. 2010.

[11] Y.M.Y. Hasan, L.J. Karam, "Morphological text extraction from images, IEEE Trans. Image Process. 9 (11) (2000) 1978–1983.

[12] E.Y. Kim, K. Jung, K.Y. Jeong, H.J. Kim, Automatic text region extraction using cluster-based templates, Proceedings of International Conference on Advances in Pattern Recognition and Digital Techniques, Calcutta, 1999.

[13] J. Ohya, A. Shio, S. Akamatsu, Recognizing characters in scene images, IEEE Trans. Pattern Anal. Mach. ntell. 16 (2)(1994) 214–224

[14] R. Jiang, F. Qi, L. Xu, and G.Wu, "A learning-based method to detec and segment text from scene images," J. Zhejiang Univ., vol. 8, pp.568–574, Apr. 2007.

[15] Aria Pezeshk and Richard L. Tutwiler." Automatic Feature Extraction and Text Recognition From Scanned Topographic Maps", in 2011.

[16] Gao, Jiang, and Yang, Jie. An adaptive algorithm for text detection from naturalscenes. In Proceedings of the 2001 IEEE Conference on Computer Vision and Pattern Recognition (December 2001), vol. 2, pp. 84–89.

[17] Gao, Jiang, Yang, Jie, Zhang, Ying, andWaibel, Alex. Text detection and translation from natural scenes. Tech. Rep. CMU-CS-01-139, Carnegie Mellon University,School of Computer Science, 2001.

[18] C. Yi and Y. Tian, "Text string detection from natural scenes by structure-based partition and grouping," vol. 19, no. 12, 2011.

[19] Chen and A. L. Yuille, "Detecting and reading text in natural scenes," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2004, pp. 366-373.

[20] K. C. Jung, K. I. Kim, and A. K. Jain, "Text information extractionin images and video: A survey," Pattern Recognit., vol. 5, pp.977-997, May 2004.

[21] Datong Chen, Jean-Marc Odobez, Herve Bourlard "Text detection and recognition in images and video frames" Pattern Recognition, vol.37, 2004, pp 595-608.