ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 7, July 2017

# Review on Text Detection, Extraction and Recognition from Images and Videos

S.Shiyamala, Dr.S.Suganya,

Research Scholar, Department of CS, RVS Arts and Science College, Coimbatore, TN, India

Associate Professor, Department of CS, RVS Arts and Science College, Coimbatore, TN, India

**ABSTRACT:** Extraction and recognition of text present in images and videos have become a very popular research area in the last decade. Generally, text present in in image and video frames is of different size,orientation, style, etc. with complex backgrounds, noise, low resolution and contrast. These factors make the automatic text extraction and recognition in images and video frames a challenging task. Large number of techniques have been proposed by various researchers in the recent past to address the problem. This paper presents a review of various state-of-the-art techniques proposed towards different stages like detection, localization, extraction, etc., of text information processing in images anf video frames. Looking at the growing popularity and the recent developments in the processing of text in images and video frames, this review imparts details of current trends and potential directions for further research activities to assist researchers.

**KEYWORDS :** Text Detection, Text extraction and Text Recognition.

## I. INTRODUCTION

Text is born as an explicit carrier of high level semantics. This unique property makes text different from other generic visual cues, such as contour, color and texture. Therefore, detecting and recognizing texts in natural scenes have become important and vibrant research areas in computer vision. Text extraction is the task of automatically extracting structured information from unstructured and/or semi-structured machine-readable documents (text).The problem is challenging in nature due to variations in text properties and reflections. Text appearing in images is classified into three categories: document text, caption text, and scene text [6]. In contrast to caption text, scene text can have any orientation and may be distorted by the perspective projection therefore it is more difficult to detect scene text.

- **Document text**: A document image (Fig. 1) usually contains text and few graphic components. It is acquired by scanning journal, printed document, handwritten historical document, and book cover etc.
- **Caption text:** It is also known as overlay text or artificial text (Fig. 2 ). It is artificially superimposed on the imageat the time of editing, like subtitles and it usually describes the subject of the image content.
- **Scene text:** It occurs naturally as a part of the scene image and contain important semantic information such asadvertisements, names of streets, institutes, shops, road signs, traffic information, board signs, nameplates, food containers, street signs, bill boards, banners, and text on vehicle etc (Fig. 3 ).

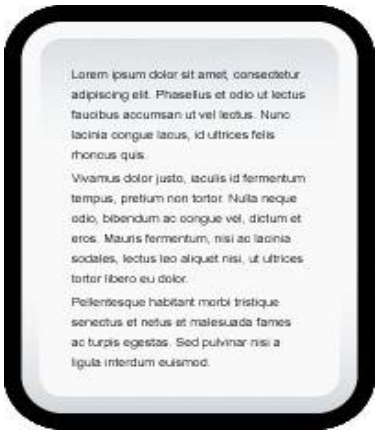
ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

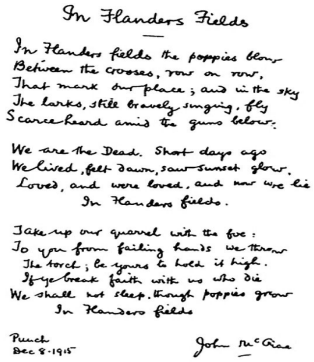
(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

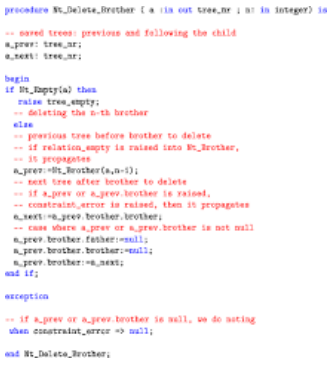
Vol. 5, Issue 7, July 2017




(a)




(b)



(c)

Fig. 1 Document Images (a) Gray-scale (b) Handwritten (c) Multi-color


Fig. 2 Caption text images


Fig. 3 Scene text images

**A. Properties of Text in Images:**

Texts usually have different appearance due to changes in font, size, style, orientation, alignment, texture, color, contrast, and background. These changes will make the problem of automatic text extraction complicated and difficult. Text in images exhibit variations due to the difference in the following properties:

- **Size**: The size of text may vary a lot.
- **Alignment:** Scene text may be aligned in any direction and have geometric distortions while caption text usually aligned horizontally and sometimes may appear as non-planar text.
- **Color:** The characters tend to have same or similar color but low contrast between text and background makes text extraction difficult.
- **Edge:** Most caption and scene texts are designed to be easily read, hence resulting in strong edges at the boundaries of text and background.
- **Compression:** Many images are recorded, transferred, and processed in compressed format. Th us, a faster text extraction system can be achieved if one can extract text without decompression.
- **Distortion:** Due to changes in camera angles, some text may carry perspective distortions that affect extraction performance.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 7, July 2017

## II. LITERATURE SURVEY ON TEXT EXTRACTION TECHNIQUES

In the past two decades, researchers have proposed numerous methods for detecting texts in natural images or videos. There are mainly three types of methods: Region based methods, texture based methods and morphological based methods. The various text extraction techniques are as follow:

**A. Region based Method**:

Region-based method uses the properties of the color or gray scale in the text region or their differences to the corresponding properties of the background. They are based on the fact that there is very little variation of color withintext and this color is sufficiently distinct from text's immediate background [20]. Text can be obtained by thresholdingthe image at intensity level in between the text color and that of its immediate background. This method is not robust to complex background. This method is further divided into two sub-approaches: connected component (CC) and edge based.

**i.) Connected Component (CC) based Method:**

CC-based methods use a bottom-up approach by grouping small components into successively larger components until all regions are identified in the image . A geometrical analysis is required to merge the text components using the spatial arrangement of those components so as to filter out non-text components and the boundaries of the text regions are marked. This method locate locates text quickly but fails for complex background.

**ii.) Edge based Method:**

Edges are a reliable feature of text regardless of color/intensity, layout, orientations, etc. Edge based method is focused on high contrast between the text and the background [5]. The three distinguishing characteristics of text embedded in images that can be used for detecting text are edge strength, density and the orientation variance. Edgebased text extraction algorithm is a general-purpose method, which can quickly and effectively localize and extract the text from both document and indoor/ outdoor images. This method is not robust for handling large size text.

**B. Texture based Method**

This method uses the fact that text in images have discrete textural properties that distinguish them from the background. The techniques based on Gabor filters, Wavelet, Fast fourier transform (FFT), spatial variance, SVM classifier and etc areused to detect the textual properties of the text region in the image [16]. This methord is able to detect the text in the complex background. The only drawback of this method is large computational complexity in texture classification stage.

**C. Morphological based Method:**

Mathematical morphology is a topological and geometrical based method for image analysis. Morphological feature extraction techniques have been efficiently applied to character recognition and document analysis. It is used to extract important text contrast features from the processed images. These features are invariant against various geometrical image changes like translation, rotation, and scaling. Even after the lightning condition or text color is changed, the feature still can be maintained. This method works robustly under different image alterations.

**Other approches:**

Epshtein et al. [8] proposed SWT, an image operator that allows for direct extraction of character strokes from edge map. Binarization techniques, which use global, local, or adaptive thresholding, are the simplest methods for text localization. These methods are widely used for document image segmentation, as these images usually include black characters on a white background, thereby enabling successful segmentation based on thresholding. This approach has been adopted for many specific applications such as address location on postal mail, courtesy amount on checks, etc., due to its simplicity in implementation [11].

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 7, July 2017

## III. METHODOLOGY OF TEXT EXTRACTION

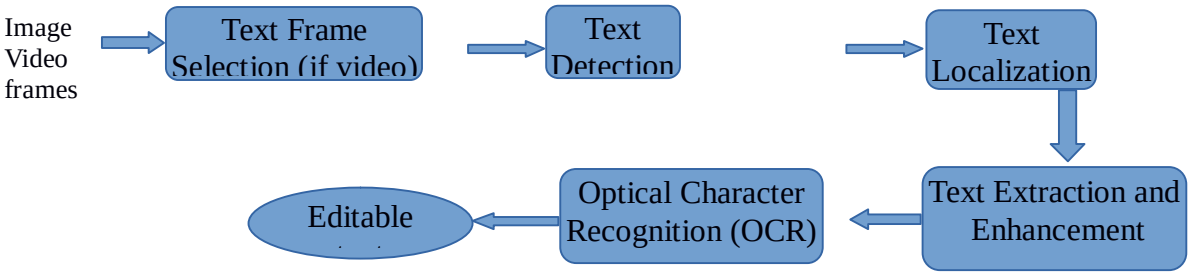



**Fig. 4 Block Diagram of Text Extraction**

The input image may be gray scale or color, compressed on uncompressed format. Text detection refers to the determination of the presence of text in the image while text localization is the process of determining the location of text and generating bounding boxes around it. After that, text is extracted i.e. segmented from the background. Enhancement of the extracted text is required as the text region usually has low-resolution and is prone to noise. Thereafter, the extracted text can be recognized using OCR. The block diagram of text extraction is shown in Fig. 4 .

### A . Applications of text extraction

Text extraction can be used for:

- Data entry for business documents, e.g. check, passport, invoice, bank statement and receipt
- Automatic number plate rcognition
- Automatic insurance documents key information extraction
- Extracting business card information into a contact list
- More quickly make textual versions of printed documents, e.g. book scanning
- Make electronic images of printed documents searchable
- Converting handwriting in real time to control a computer
- Assistive technology for blind and visually impaired users

## IV. DISCUSSION ABOUT PERFORMANCE EVALUATION

There are several difficulties related to performance evaluation in nearly all research areas in computer vision and pattern recognition (CVPR). The empirical evaluation of CVPR algorithms is a major endeavor as a means of measuring the ability of algorithms to meet a given set of requirements. Although various studies in CVPR have investigated the issue of objective performance evaluation, there has been very little focus on the problem of TIE in images and video. This section reviews the current evaluation methods used for TIE and highlights several issues in these evaluation methods.

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 frames and all the given frames containing text. Measuring the performance of text extraction is extremely difficult and until now there has been no comparison of the different extraction methods. Instead, the performance is merely inferred from the OCR results, as the text extraction performance is closely related to the OCR output Performance evaluation of text extraction is not simple. Some of the issues related to the evaluation of text localization methods have been summarized by Antani et al. [2]

**(i) Ground truth data :** Unlike evaluating the automatic detection of other video events, such as video shot changes, vehicle detection, or face detection, the degree of preciseness of TIE is difficult to define. This problem is related to the

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 7, July 2017

construction of the ground truth data. The ground truth data for text localization is usually marked by bounded rectangles that include gaps between characters, words, and text lines. However, if an algorithm is very accurate and detects text at the character level, it will not include the above gaps and thus will not have a good recall rate [1].

**(ii) Performance measure :** After determining the ground truth data, a decision has to be made on which measures to use in the matching process between localized results and ground truth data. Normally, the **recall and precision rates** are used. Additionally, a method is also needed for comparing the ground truth data and the algorithm output: pixel-by-pixel, character-by-character, or rectangle-by-rectangle comparison.
**(iii) Application dependence :** The aim of each text localization system can differ. Some applications require that all the text in the input image must be located, while others only focus on extracting important text. In addition, the performance also depends on the weights assigned to false alarm or false dismissal.

**(iv) Public database :** Although many researchers seek to compare their methods with others, there are no domain-specific or general comprehensive databases of images or videos containing text. Therefore, researchers use their own databases for evaluating the performance of the algorithms. Further, since many algorithms include specific assumptions and are usually optimized on a particular database, it is hard to conduct a comprehensive objective comparison.

Table 1 shows performance of various approaches in text detection and extraction.

**Table 1** Various Text Extraction Techniques

| Author, year | Technique Used | Images | Parameters | Remarks |
|---|---|---|---|---|
| Yao et al.[17], 2007 | CC andSupport Vector Machine (SVM) | Complex background images | PR=64% RR=60% | Pixels of each character assumed to have similar color. |
| Lai et al. [13], 2008 | Edge detection and K-means clustering | Signboard Images | | Efficient for uneven illumination. |
| Zhang et al. [20], 2008 | Discrete Wavelet Transform (DWT), k-means clustering, morphology Operations | Background images with different languages, fonts and sizes | DR= 94.5%, FAR= 13.6% | Text character Color independent. |
| Song et al. [16], 2008 | Histogram Projection and color based K-means clustering | Chinese text | PR=77.05% RR=75.63% | K=3 gives best performance. |
| Dinh et al. [7], 2008 | Edge detection and Histogram Projection | Signboard Texts | | Low complexity algorithm. |
| Fan et al.[9], 2009 | Stroke features and connected component | Caption text images | PR=95.2% RR= 94.5% | Color information is not fully used. |
| Audithan et al.[5], 2009 | Haar DWT, Morphological Dilation operator, logical AND operator, Dynamic thresholding | Document images | DR =94.8 % | Independent of contrast. |

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 7, July 2017

| | | | | |
|---|---|---|---|---|
| Anoualet al.[3], 2010 | Edge detection, texture features, connected component analysis | Complex background images | PR=95% RR=89% | Robust and effective. |
| Kumar et al [12], 2010 | CC Analysis | ICDAR 2003 scene images | PR=90% RR=89% | Capable of Multilingual Text extraction. |
| Hassanzadeh et al.[10], 2011 | Morphological operator, Decision classifier | Logo detection in document images | PR=95.6% Accuracy=86.9 % | A novel and fast method for logo detection. |
| Zaravi et al. [18], 2011 | DWT, Dynamic thresholding, Region of Interest (ROI) | Colored book and journal cover sheets | DR=91.20% | Robust to noise. |
| Zhang et al. [19], 2012 | Edge Enhancement and CC | Web images and caption text images | DR=92.4% | Not sensitive to various types of background noises. |
| Seeri et al. [15], 2012 | Median filter, Sobel edge detectoreconnected component labeling, order static filter. | Kannada text images | PR=84.21% RR=83.16% Accuracy = 75.77% | Fails to extract very small characters. |
| Azadboni et al. [6], 2012 | FFT Domain Filtering , SVM Classification, K-means clustering | Scene text images | DR= 98.10% | Text characters having uniform colour. |
| Anupama et al.[4], 2013 | Morphology operators, Histogram Projection ( X and Y histogram) | Handwritten Telugu document images. | DR=98.54%, Accuracy =98.29% | Fail in case of touching characters and over-lapping lines. |
| Raj et al. [14], 2014 | CC based | Natural Scene Images (Devanagari text) | PR= 72.8%, RR=74.2 % | Fails for small slanted/curved text. |

## V.CONCLUSION

This paper provided a comprehensive survey of text information extraction in images and video. As a result a text-image-analysis is needed to enable a text information extraction system to be used for any type of image, including scanned document images, real scene images through a video camera, caption text images and videoimages. The above study shows that, most of the methods fall into any one of the above techniqus and also that there is a limitation in each technique to give a better detection rate with fewer false alarms without any constraints for text region extraction in different type of images. But still, to need a completely robust and generalized technique for text segmentation, it is difficult to provide appropriate input to the optical character recognition (OCR) system. So, a combined method has been proposed for automatic text content extraction from different images and video which is independent of the various characteristics of text.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 7, July 2017

## REFERENCES

1. S. Antani, D. Crandall, A. Narasimhamurthy, V. Y. Mariano, and R. Kasturi, Evaluation of Methods for Detection and Localization of Text in Video, Proc. of the IAPR workshop on Document Analysis Systems, Rio de Janeiro, December 2000, pp. 506-514.
2. S. Antani, Reliable Extraction of Text From Video, PhD thesis, Pennsylvania State University, August 2001.
3. H. Anoual, D. Aboutajdine, S.E. Ensias, A.J. Enset, Features Extraction for Text Detection and Localization,5th International Symposium on I/V Communication and Mobile Network, IEEE, 2010, pp. 1-4.
4. N. Anupama, C. Rupa, E.S. Reddy, Character Segmentation for Telugu Image Document using Multiple Histogram Projections, Global Journal of Computer Science and Technology, Vol. 13, 2013, pp. 11-16.
5. S. Audithan, RM. Chandrasekaran, Document Text Extraction from Document Images using Haar Discrete Wavelet Transform, European.
6. M.K. Azadboni, A. Behrad , Text Detection and Character Extraction in Color Images using FFT Domain Filtering and SVM Classification, 6th International Symposium on Telecommunications. IEEE, 2012, pp. 794-799.
7 . T.N. Dinh, J. Park and G.S. Lee, Low-Complexity Text Extraction in Korean Signboards for Mobile Applications, IEEE International Conference on Computer and Information Technology, 2008, pp. 333-337.
8. B. Epshtein, E. Ofek, and Y. Wexler. Detecting text in natural scenes with stroke width transform. In Proc. of CVPR, 2010.
9. W. Fan, J. Sun, Y. Katsuyama, Y. Hotta, S. Naoi, Text Detection in Images Based on Grayscale Decomposition and Stroke Extraction, Chinese Conference on Pattern Recognition, IEEE, 2009, pp. 1-4.
10. S. Hassanzadeh, H. Pourghassem, Fast Logo Detection Based on Morphological Features in Document Image, 2011 IEEE 7th International Colloquium on Signal Processing and its Applications, 2011, pp. 283-286.
11. L. Kang, Y. Li, and D. Doermann.Text detection using binarization in natural images. In Proc. of CVPR, 200ssssss4.
12. M. Kumar, Y.C. Kim and G.S. Lee, Text Detection using Multilayer Separation in Real Scene Images, 10th IEEE International Conference on Computer and Information Technology, 2010, pp. 1413-1417.
13. A.N. Lai, G.S. Lee, Binarization by Local K-means Clustering for Korean Text Extraction, IEEE Symposium on Signal Processing and Information Technology, 2008, pp. 117-122.
14.H. Raj, R. Ghosh, Devanagari Text Extraction from Natural Scene Images, 2014 International Conference on Advances in Computing, Communications and Informatics (ICACCI), IEEE, 2014, pp. 513-517.
15. S.V. Seeri, S. Giraddi and Prashant. B. M, A Novel Approach for Kannada Text Extraction, Proceedings of the International Conference on Pattern Recognition, Informatics and Medical Engineering, 2012, pp. 444-448.
16. Y. Song, A. Liu, L. Pang, S. Lin, Y. Zhang, S. Tang, A Novel Image Text Extraction Method Based on K-means Clustering, Seventh IEEE/ACIS International Conference on Computer and Information Science, 2008, pp. 185-190.
17. J.L. Yao, Y.Q. Wang, L.B. Weng, Y.P. Yang, Locating Text Based on Connected Component And SVM, International Conference On Wavelet Analysis And Pattern Recognition, Vol. 3, 2007, pp. 1418 - 1423.
18. D. Zaravi, H. Rostami, A. Malahzaheh, S.S Mortazavi, Text Extraction using Wavelet Thresholding and New Projection Profile, World Academy of Science, Engineering and Technology, Vol. 5, 2011, pp. 528-531.
19. Y. Zhang, C. Wang, B. Xiao, C. Shi, A New Text Extraction Method Incorporating Local Information, International Conference on Frontiers in Handwriting Recognition, 2012, pp. 252-255
20.X.W. Zhang, X.B. Zheng, Z.J. Weng, Text Extraction Algorithm Under Background Image Using Wavelet Transforms, IEEE Proceedings of International Conference On Wavelet Analysis And Pattern Recognition, Vol. 1, 2008, pp. 200-204.