



HAL
open science

A Review of various Text Localization Techniques

Niharika Yadav, Vinay Kumar

► **To cite this version:**

Niharika Yadav, Vinay Kumar. A Review of various Text Localization Techniques. 2016. hal-01326642

HAL Id: hal-01326642

<https://inria.hal.science/hal-01326642>

Submitted on 4 Jun 2016

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

A REVIEW OF VARIOUS TEXT LOCALIZATION TECHNIQUES

Niharika Yadav, Vinay Kumar

Electronics and Communication Department,
Thapar University, Patiala, Punjab 147004, India
y.niharika27@gmail.com,vinaykverma@gmail.com

ABSTRACT

Text information extraction is a growing area of research. Enormous work has been done to efficiently and robustly extract the text regions from scene text images. Various Text extractions models have been defined comprising of various stages. Amongst them text localization is an important stage and thus researchers mainly focus on this stage. This paper provides a review of various techniques of text localization. The paper also includes a comparison of the performance of these techniques on the parameters such as precision rate and recall rate.

Keywords: Scene images, Text localization, CC-based method, Texture based method.

I. INTRODUCTION

Text extraction is an arduous task in document image Analysis. Text in images contains information which is noteworthy and can be utilised for structuring of images or indexing databases. This also has its usefulness for many real world applications, such as navigation, licence plate detection, conversion of paper based collection to e-books using OCR's etc. [1].text embedded in the mages contains a systematic information which can be used as a content based access to digital images, if it is extracted and harnessed efficiently.

Text appearing in images is broadly classified in three categories: Document text, Scene text, Digital-born text [2]. Amongst these extracting texts from the scene text images is most challenging as this type of text can have variable orientations, font size etc. In addition they may also be distorted due to the perceptive projections and may have low contrasts. Figure 1 shows examples of scene text images.

A generalised Text Information Extraction (TIE) model has been shown in Figure 2. Regardless of the profound research to develop an efficient model there are various problems with a scene text image that make automatic text extraction a challenging task.

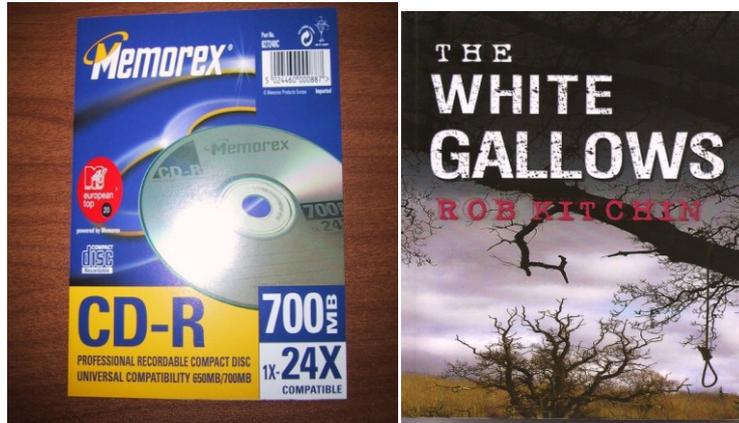


Figure 1: Examples: Scene text images

A variety of methods have been proposed by various researchers that help in retrieving the scene text from the images. This manuscript gives a broad study of various techniques proposed in the area of text detection and localization. These are the first two major steps in the TIE model and are the most difficult one to implement.

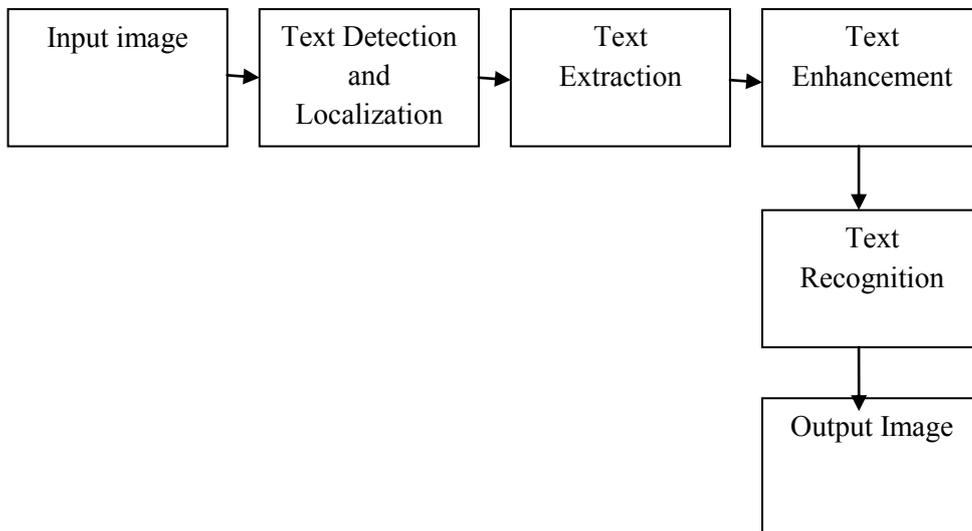


Figure 2: Text Information Extraction Model

II. TEXT LOCALIZATION TECHNIQUES

When an image is fed to the TIE, there is no prior information about the existence of the text in an image. Hence text detection techniques help identifying the presence of text in an image. The various features of a text such as the color, intensity, geometry etc, are utilized to classify the various localization techniques. Two broad classifications of localization

techniques are: Region based approaches and texture based approaches. Figure 3 shows the classification of text localization techniques.

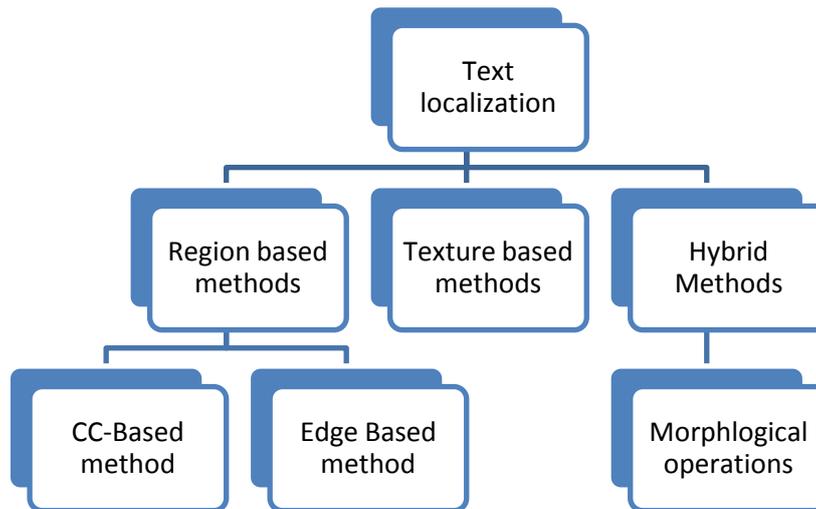


Figure 3: Classification: Text Localization Techniques

A. Region Based Approach:

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 [3]. They utilise the property that the variation of color within text is minimal moreover there is a sufficient distinction between the text and the background. Text can be obtained by thresholding the image at intensity level in between the text color and that of its immediate background. This approach is implemented in bottom-up fashion. Sub structures are identified and merged to form homogeneous region which are marked with bounding boxes. This method is further divided into two sub-approaches:

Connected component based (CC-based):

Small similar components of the input images are grouped to form successively large components using bottom up approach in the case of CC-based methods. This process iteratively continues until all the regions of the image are not identified. Filtering of the non text regions of the image is performed by considering a geometrical analysis which merges the components using the spatial arrangement of the components and mark the boundaries of the text regions.

Edge Based Methods:

Edges are those portions of the image that corresponds to the object boundaries. There is a major contrast in the text and the background and this drastic variation forms the edges.

The non text region of the images is filtered out by firstly identifying the edges of the text boundaries and then applying various heuristics [21]. Generally, an edge filter is incorporated for the identification of the edges in an image, and for the merging stage an averaging operation or a morphological operator is used.

B. Texture Based Approach

Texture-based methods exploit textural features of the text. These properties distinguished the text from its background [5]. Parameters such as energy, contrast, correlation, and entropy define the texture of an image. Texture based methods implement techniques such as FFT, wavelet spatial variance etc.

C. Morphological Approach

Mathematical morphology is a topological and geometrical based approach for image analysis [6]. Character recognition and document analysis uses the morphological operations because operations like translation, rotation and scaling do not have any effect on the geometrical shape of the image. This method works robustly under different image alterations.

A detailed study of various techniques based on the above approach has been done and review is given in the next section.

III. LITERATURE REVIEW

This section provides a detailed review of various localization techniques.

A. Connected Component Based Method

Sobattka et al. [7] proposed a CC-based method that implemented bottom up and top down analysis on images of book covers and journals. Histogram based clustering was implemented in the preprocessing step which reduced the color variation in the image significantly. For the text localization step, in bottom-up analysis homogenous regions were identified while in case of bottom-up analysis image was iteratively split horizontally and vertically. Finally the two results were compared to obtain the text regions of the image. The method gave a character recognition rate of 98%.

Hase et al. [8] proposed a CC-based method for color documents. A common assumption of the machine printed characters is that each character is printed with a single color. Color histogram was generated by converting the Pixel values to a new color space. In the next step the image is divided in various binary images and a multi-stage relaxation technique is

applied to extract strings .In the next step all the results obtained from individual images were merged and character strings were identified using likelihood using conflict resolution rules.

Jain and Yu [9] applied a model that included following stages. In the first stage, preprocessing, this included bit dropping. A 24-bit image was bit dropped to 6-bit image. In the second stage of color clustering, image was quantized after the decomposition of the image in multiple foregrounds; same localization technique is implied on each foreground image. The authors used a block adjacency graph to generate the CCs for each foreground image. The localized text components of all the individual foreground images were then merged to obtain an output image. The algorithm gave promising results for horizontal and vertical texts however text detection was poor in case of skewed text.

B. Texture Based Method:

Shiva kumara et.al [10] proposed a new method for detection of handwritten text from videos. The algorithm was divided in five steps. In the first stage. Averaging of the RGB planes was performed to enhance the text areas of image. Maximum Color Difference (MCD) criteria then applied to restore the contrast of the image and to highlight the test regions in the next stage, to cluster the text regions K-mean clustering was performed with $K = 2$. In the following stage the candidate region were bounded using sobel filter. In the next stage, a novel Boundary Growing Method (BGM) based on the nearest neighbour concept was used to adjust the boundary of the text. The method fixed the boundaries by considering the fact that in one direction there is regular spacing between words and in the final stage, the false positive edges obtained due to BGM were eliminated using the concept of intrinsic and extrinsic edges.

Shyama et.al [11] proposed a text segmentation technique that extracted text by projection algorithm. Color based segmentation was performed to link consecutive pixels that lied in the same direction. Light Edge Enhancement (LEE) was used to locate the regions having consecutive pixel and to enhance the edges. Next, the motion blur was removed with the help of heavy edge enhancement (HEE) from camera image

Kim et al. [13] proposed a method that could detect texts using texture properties. A support vector machine (SVM) was implemented to study the properties of textural basis of the image. The method used the raw pixels of the image that confided with the textural pattern rather than adding new external texture feature extraction model. This worked fine even for the high dimensional spaces. Next step of the method was to implement, continuously adaptive mean shift algorithm

(CAMSHIFT) that could locate the text regions by using the analysis of the texture properties. The combination of CAMSHIFT and SVMs produced very resourceful results as text detection was significant. The performance of the system was based on the criterion that the SVM was able to classify the text and non text regions and not on the final text detection. The author used a set of 100 images that were divided in two sets. The first set comprised of 70 images that were used for training and the other 30 were used to validate the results obtained.

Edge Based Methods:

Xiaoqing Liu [14] method consists of three stages: candidate text region detection, text region localization and character extraction. The first stage implements a multiscale edge detector where multiscale images are produced by Gaussian low pass filter which undergo down sampling to reduce the image in both horizontal and vertical direction. A feature map is then generated that separates the texture like regions from text regions. In the second stage text localization is performed implying clustering techniques. In the third stage, already proposed existing OCR engine where used, that were only efficient with printed documents having plane background and are unable to give promising results in case of text embedded in the image or in case of complex backgrounds.

Chen et al. [38] implemented canny edge operator that identified the high contrast regions of the image. In order to reduce the computational complexity a single edge point was used in one window for the estimation and orientation. The information obtained is then used to enhance the image by performing morphological operations on images that could connect the edges into clusters. Morphological operations were performed on the image to connect the edges into clusters. Two types of Gabor asymmetric filters were implemented to generate a general estimation: an edge-form filter and a stripe-form filter. The obtained information about the edge is then enhanced using a suitable scale. This resulted in the blurring of the area that had no specific scale. Image enhancement was then obtained using the localization techniques.

Jain and Chugh [16] proposed a method based on the nearest neighbour approach. The algorithm has four stages i) Color Processing ii) Detection of Edges iii) Text Area Localization iv) Gap Filling and Enhancement. In the preprocessing step color processing is performed by first converting the image to YUV format. In the next stage an edge image is generated to highlight the text regions of the image. To localize the text areas, height-wise and width-wise analysis is done through the horizontal and vertical histogram for the target

text areas. A geometric proportion between the length and width is modelled experimentally to remove the probable non-text regions. Areas with ratio of less than ten are marked text regions. The algorithm was evaluated for a dataset of 25 images with variation in font size, perspective and alignment under different lightning conditions.

C. Morphological Based Method

Jui-Chen Wu [17] proposed a method that consists of three stages a) feature extraction, b) Text localization, and c) Text verification. In the first stage, the high contrast feature of the text region was extracted using a novel set of morphological operations. In the second stage, the possible text regions were marked using a labelling technique. The skewness of the text lines was handled using a moment based method. The potential text segments were extracted using a rule based scheme. Furthermore the text lines were merged by defining similarity measures. In the final stage, an x-projection technique was proposed that was able to extract all the regularities from the text lines.

Luz et.al [18] implemented morphological filters to extract the candidate text regions from the scene images. Closing and opening operations were performed on the connected components. A decision tree classifier was used to differentiate between the text and non text regions.

IV. COMPARISON AND PERFORMANCE ANALYSIS

Analysis of the various techniques for text localization is shown in Table I. In order to calculate the robustness and performances of the proposed algorithms two methods are defined. The Precision rate is defined as the ratio of correctly detected words to the sum of correctly detected words plus false positives.

$$\text{Precision rate} = \frac{\text{correctly detected characters}}{\text{correctly detected characters} + \text{false positive}} \times 100 \quad (1)$$

The Recall rate is defined as the ratio of correctly detected words to the sum of correctly detected words plus false negatives.

$$\text{Recall rate} = \frac{\text{correctly detected characters}}{\text{correctly detected characters} + \text{false negative}} \times 100 \quad (2)$$

Table II shows the performance analysis of the text localization techniques.

Table I Comparison of various techniques

Technique	Author	Approach	Remark
CC- Based Approach	Sobattka et al.[7]	Implemented bottom-up and top down analysis	Robust to size of text
	Hase <i>et al.</i> [8]	Implemented likelihoods using conflict resolution rules.	Works for curved and shadowed strings.
	Jain and Yu [9]	Implemented Bit dropping	Poor performance when color variance is less.
Texture based Approach	Shiva kumara <i>et.al</i> [10]	Based on Maximum Color Difference (MCD) and Boundary Growing Method (BGM)	Insensitive to contrast.
	Shyama et.al [11]	Based on colour based segmentation, Light edge enhancement, Heavy edge enhancement	Size and orientation does not affect the performance.
Edge based approach	Xiaoqing Liu [14]	Based on multiscale images using Gaussian pyramids.	Robust and effective for complex images
	Chen <i>et al.</i> [15]	Based on canny edge detector	Blurs or eliminates regions that do not have specified spaces.
	Jain and Chugh [16]	Based on nearest neighbour approach	Robust to sizes and styles, orientations, alignment, uneven illumination, and reflection effects.
Morphological	Jui-Chen Wu [17]	Novel set of morphological Operations and an x-projection	Detects complete line from its segments

based approach		techniques	
	Luz <i>et.al</i> [18]	Morphological filters , Decision tree classifier	Insensitive to position of the text

Table-II Performance Analysis

Technique	Precision rate	Recall rate
Sobattka et al. [7]	83.5	85
Hase <i>et al.</i> [8]	86.9	85
Shiva kumara <i>et.al</i> [10]	65	68
Shyama <i>et.al</i> [11]	89	84
Kim <i>et al.</i> [12]	85.3	82.2
Xiaoqing Liu [13]	91.2	96
Chen <i>et al.</i> [14]	82	87
Jain and Chugh [15]	70	90
Jui-Chen Wu [16]	95.1	96
Luz <i>et.al</i> [17]	86.9	85

V. CONCLUSION

Content retrieval from images for various proposes has been performed for many years. This paper provides a broad study of various techniques propounded over the years of research. The paper presents a comparison of approaches that utilize diverse features of scene text images. A performance analysis of all the studied method is also includes in this paper. From the study done it is observed that depending upon the raw image and the application different features are exploited, hence implementing one of the techniques will not offer a high efficiency. This is mainly due to the fact that naturally captured images have varying features. Thus a hybrid approach is required technique that can efficiently identify the text regions using varied text features.

REFERENCES

- [1] K. Jung, K. I. Kim, and A. K. Jain, "Text information extraction in images and video: a survey," *Pattern Recognition*, vol. 37, no. 5, pp. 977-997, 2004.
- [2] K. Wang, B. Babenko, and S. Belongie, "End-to-end scene text recognition," in *Computer Vision (ICCV), 2011 IEEE International Conference on*, pp. 1457-1464, 2011.
- [3] M. Sawaki, H. Murase, and N. Hagita, "Automatic acquisition of context-based images templates for degraded character recognition in scene images," in *Pattern Recognition, 2000. Proceedings.15th International Conference on*, vol. 4, pp. 15-18, IEEE, 2000.
- [4] Y. Zhong, K. Karu, and A. K. Jain, "Locating text in complex color images," in *Document Analysis and Recognition, 1995, Proceedings of the Third International Conference on*, pp. 146-149, 1995.
- [5] J. Gao and J. Yang, "An adaptive algorithm for text detection from natural scenes," in *Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on*, vol. 2, pp. II-84, IEEE, 2001.
- [6] S. R. Sternberg, "Grayscale morphology," *Computer vision, graphics, and image processing*, vol. 35, no. 3, pp. 333-355, 1986.
- [7] K. Sobottka, H. Kronenberg, T. Perroud, and H. Bunke, "Text extraction from colored book and journal covers," *International Journal on Document Analysis and Recognition*, vol. 2, no. 4, pp. 163-176, 2000.
- [8] H. Hase, T. Shinokawa, M. Yoneda, and C. Y. Suen, "Character String Extraction from Color Documents," *Pattern Recognition*, vol. 34, no. 7, pp. 1349-1365, 2001.
- [9] A. K. Jain and B. Yu, "Automatic text location in images and video frames," *Pattern Recognition*, vol. 31, no. 12, pp. 2055-2076, 1998.
- [10] X. Liu and J. Samarabandu, "Multiscale edge-based text extraction from complex images," in *Multimedia and Expo, 2006 IEEE International Conference on*, pp. 1721-1724, 2006.
- [11] Shivakumara P, A Dutta, U Pal And C L Tan, "A New Method For Handwritten Scene Text Detection In Video", *International Conference On Frontiers In Handwriting Recognition*, pp: 16-18, 2010.
- [12] Y.F Pan, X. Hou, C. -Lin "Text Localization In Natural Scene Images Based On Conditional Random Field," *ICDAR*, pp 6-10, 2009.

- [13] K. I. Kim, K. Jung, and J. H. Kim, "Texture-based approach for text detection in images using support vector machines and continuously adaptive mean shift algorithm," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 25, no. 12, pp. 1631-1639, 2003.
- [14] S. P. Chowdhury, S. Dhar, K. Rafferty, A. K Das, and B. Chanda, "Robust Extraction of Text from Camera Images using Colour and Spatial Information Simultaneously." *Journal of Universal Computer Science*, vol. 15, no.18, pp.3325-3342,2009.
- [15] D. Chen, K. Shearer, and H. Bourlard, "Text enhancement with asymmetric filter for video OCR", in *Image Analysis and Processing, 2001. Proceedings.11th International Conference on*, pp. 192-197, 2001.
- [16] Y. K. Jain, S. Chugh, "Character Localization From Natural Images Using Nearest Neighbours Approach," *International Journal of Science & Engineering Research* , vol. 2, no. 12, 2011.
- [17] J.-C. Wu, J.-W. Hsieh and Y.-S. Chen, "Morphology-based text line extraction," *Machine Vision and Applications*, vol. 19, no. 3, pp. 195-207, 2008.
- [18] W. A. L. Alves, R. F. Hashimoto, "Text Regions Extracted from Scene Images by Ultimate Attribute Opening and Decision Tree Classification", *Proc. 23rd SIBGRAPI Conference on Graphics, Patterns and Images*, ISBN: 978-1-4244-8420-1, pp. 360 - 367, 2010.