# A Study of Text Localization Algorithms for Complex Images

Niti Syal[1], Naresh Kumar Garg[2]

M.Tech Student, Dept. of CSE, GZS PTU Campus, Bathinda, India[1]

Assistant. Professor, Dept. of CSE., GZS PTU Campus, Bathinda, India[2]

**ABSTRACT**: With the advancement of the digital technology, the multimedia databases are increasing every day. The databases usually contain images and videos in addition to the textual information. Several papers have been reported in the literature and numerous algorithms are proposed for text localization in natural images and videos. All these algorithms are proposed considering the different properties of text that helps to distinguish the text regions from the other regions in the natural scenes. In this paper, review of text localization is given. It will help the beginner's to start their work in the field of text localization.

**KEYWORDS**: Region Based, Texture Based, Support Vector Machine

## I.    INTRODUCTION

Text localization and recognition in real-world (scene) images is an open problem which has been receiving significant attention since it is a critical component in a number of computer vision applications like searching images by their textual content, reading labels on  businesses in map applications (e.g. Google Street View) or assisting visually impaired as shown in Figure 1 [18].
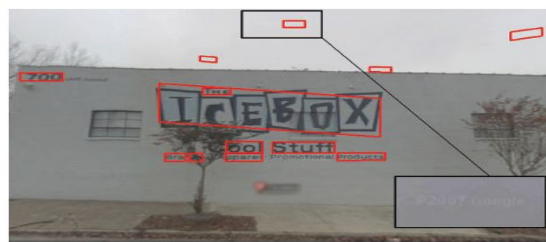


**Figure 1: Text detection in the SVT dataset. All "false positives" in the image are caused by watermarks embedded into the dataset. This demonstrates robustness of the proposed method against noise and low contrast of characters (in the bottom right corner the area of interest is enlarged and contrast artificially increased, "2007 Google" is readable).**

The textual information is very useful semantic information because it describes the image or video and can be used to fully understand images and videos. Text localization can be done in three kinds of images namely:

1. Document image

2. Scene text image

3. Caption text image

Document images may be in the form of scanned book covers, CD covers or video images. Text in images or videos is classified as scene text and caption text. Scene text is also called as graphics text. Natural images that contain text are called scene text. The name of caption text is artificial text and it is one in which text is inserted or superimposed in the image [2]. Localizing text in an image is potentially a computationally very expensive task as generally any of the $2^N$ subsets can correspond to text (where N is the number of pixels). Text localization methods deal with this problem in two different ways.

Most published methods for text localization and recognition [3]-[4]-[5] are based on sequential pipeline processing consisting of three steps - text localization, text segmentation and processing by an OCR for printed documents. In such approaches, the overall success rate of the method is a product of success rates of each stage as there is no possibility to made decisions by previous stages. Some authors have focused on subtasks of the scene text recognition problem, such as text localization [3]-[4] [6]-[7] individual character recognition [8]-[13] or reading text from segmented areas of images [10]. Whilst they achieved promising results for individual subtasks, separating text localization from text recognition inevitably leads to loss of information, which results in degradation of overall text localization and recognition performance. Text localization can be useful for many applications including tourism, navigation for the blind, robot guidance, and intelligent transportation systems. In order to be easily seen from the road, most signs contain text printed on a homogenous background.

The objective of text localizer is to output one text rectangle per text word present in the image as ICDAR database [10] is annotated at word level. The output of text detector is rectangles of varying sizes covering text regions and some non-text regions.

## II.    APPROACHES FOR TEXT LOCALIZATION

There are two different approaches have been used for text localization from complex images namely region based approach and texture based approach.

### A.    Region Based Approach

This approach uses the properties of the color or gray scale in a text region or their differences regarding the background. This method is basically divided in two sub categories: edge based and connected component (CC) based methods. The edge based method is mainly focus on the high contrast between text and background. In this method, firstly text edges are identified in an image and are merged. Finally, some heuristic rules are applied to discard non-text regions. Connected component based method considers text as a set of separate connected components, each having distinct intensity and color distributions. The edge based methods are robust to low contrast and different text size where as CC based methods are somewhat simpler to implement, but they fail to localize text in images with complex backgrounds [11].

### B.    Texture Based Methods

As we know that, text in images has distinct textural properties which can be used to differentiate them from the background or other non text regions [12]. This method is based on the concept of textural properties. In this method, Fourier transforms. Discrete cosine transform and wavelet decomposition are generally used. The main drawback of this method is that it is highly complex in nature but, in other hand, it is more robust than the CC based methods in dealing with complex background.

## III.    TEXT LOCALIZATION

### A.    MSER Detection

Since the original search space induced by all regions (R) of image (I) is huge, certain approximations were applied in our approach. Assuming that individual characters are detected as External Regions (ER) and taking computation complexity into consideration, the search space was limited to the set M of Maximally Stable External Regions (MSER) [14], which can be computed in linear time in number of pixels [15]. The set of MSERs detected in certain scalar image projections (intensity, red channel, blue channel, green channel) defines the set of vertices of the graph G, i.e. V (G) =M. The edges of the graph G are not stored explicitly, but they are induced on the fly.

### B. Character and Non-Character Classification

In this module, each vertex of graph G is labeled as a character or a non character using a trained classifier which creates an initial hypothesis of text position, because character vertices are likely to be included in some path p representing a text as shown in Fig. 2 [10].



**Figure 2:Character/non-character MSER classifier: (a) Character and (b) non-character training samples (MSER boundaries marked yellow). Initial MSER classification for (c) Armenian, (d) Russian and (e) Kannada script (character regions marked green, Non character regions marked red). Note: The training set contains only 1227 character samples of Latin script and 1396 non-character samples**

The features used by the classifier as shown in Table.1 are scale invariant to detect all characters sizes, but they are not rotation invariant, which implies that characters at different rotations had to be included in the training set.

| | |
|---|---|
| Aspect ratio | Relative segment height |
| Compactness | Number of holes |
| Convex hull area to surface density | Character color consistency |
| Background color consistency | Skelton length |

**Table 1: Features used by the character classifier**

A standard Support Vector Machine (SVM) [16] classifier with Radial Basis Function (RBF) kernel [17] was used. The classifier was trained on a set of 1227 characters and 1396 non-characters, results obtained by manually annotating MSERs extracted from real-world images downloaded from Flicker. The classification error obtained by cross-validation was 5.6%. The training set was relatively small and certainly does not contain all possible fonts, scripts or even characters, but extending the training set with more examples did not bring any significant improvement in the classification success rate. This indicates that features used by the character classifier are insensitive to fonts and alphabets.

### A. Textline Hypothesis Formation

In real-world images a font rarely changes inside a word, which implies that certain character measurements (character height, aspect ratio, spacing between characters, stroke width, etc.) are either constant or constrained to a limited interval. Based on this observation, an approximation $h \char94 (p, v)$ of function $h (p;v)$ was implemented using a SVM classifier with polynomial kernel, whose feature vector is created by comparing average character measurements of the existing path p to the character measurements of given vertex v.

## IV. TEXT LOCALIZATION IN COMPLEX IMAGES

For text localization in complex images following steps are used: First, a wavelet-based edge extraction scheme is applied on gray-level image. Second the gray level edge image is converted into binary image by using suitable global thresholding. After that, a new filter is applied on binary image to remove noise and non text areas. The text locations

are determined using new projection profile. Several heuristic methods are applied to improve the system performance. Bounding boxes are generated two last steps. Figure.3 shows the structure of the system.
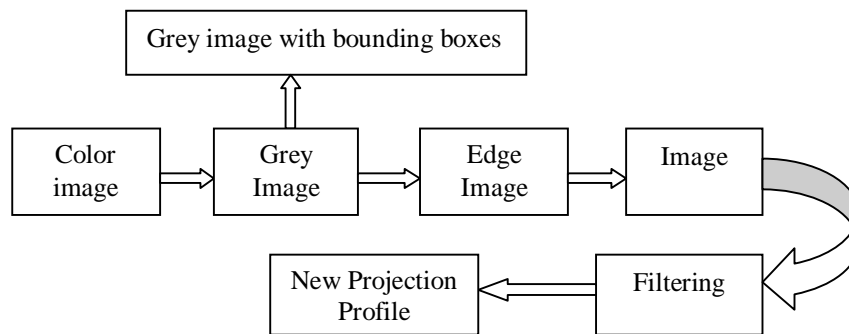


**Figure 3. Text Localization In Complex Images**

## V.    CONCLUSION

Text data in image present useful information for automatic annotation, indexing and structuring of images. Many text localization approaches use features related to text character properties. Edge detection is a common first step in many such algorithms. Texture is another widely used feature for text detection and localization. Some approaches assume that the text is written in the horizontal or vertical direction. A number of methods discussed above, focus on identifying text so that the text can be extracted and interpreted. This paper will be very useful for the beginners to start their work in the field of text localization.

### REFERENCES

1.      Neha Gupta, V.K Banga, "Localization of Text in Complex Images Using  Haar Wavelet Transform," International Journal of Innovative Technology and Exploring Engineering (IJITEE), Volume-1, pp.111, 2012.
2.      Chitrakala Gopalan and D. Manjula, "Contourlet Based Approach for Text Identification and Localization from Heterogeneous Textual Images", International Journal of Electrical and Electronics Engineering , Volume-2, pp. 491-500, 2008.
3.      Wu, V., Manmatha, R., Riseman, Sr., E.M., "An automatic system to detect and recognize text in images",IEEE Trans. Pattern Anal. Mach. Intell, Volume-21, pp.1224-1229,1999.
4.      Chen, X., Yuille, A.L., "Detecting and reading text in natural scenes. Computer Vision and Pattern Recognition", IEEE Computer Society Conference, Volume- 2,  pp.366-373., 2004.
5.      Jain, A.K., Yu, B., "Automatic text location in images and video frames", Pattern recognition, Volume- 31,  pp. 2055-2076, 1998.
6.      Pan, Y.F., Hou, X., Liu, C.L., "A robust system to detect and localize texts in natural scene images",  Document Analysis Systems, IAPR International Workshop, Volume-1, pp. 35-42, 1998.
7.      Pan, Y.F., Hou, X., Liu, C.L., "Text localization in natural scene images based on conditional random eld."  ICDAR '09 Proc. of the 2009 10th International Conference on Document Analysis and Recognition, Volume-6 , pp.6-10, 2009.
8.      De Campos, T.E., Babu, B.R., Varma, M., "Character recognition in natural images", VISAPP, Volume-2, pp.1150–1157, 2009.
9.      Yokobayashi, Wakahara, T, "Segmentation and recognition of characters in scene images using selective binarization in color space and gat correlation", 8th International Conference on Document Analysis and Recognition, Volume-1, pp.167-171, 2005.
10.     Weinman, J.J., Learned-Miller, E., Hanson, A.R., "Scene text recognition using similarity and a lexicon with sparse belief propagation", IEEE Trans. Pattern Anal. Mach. Intell. Volume-22 , pp.1733-1746, 2009.
11.     Keechul Jung, Kwang In Kim and Anil K. Jain, "Text information localization in images and video: A Survey", Elsevier, Pattern Recognition, Volume-37, pp. 977–997, 2004.
12.     Mohieddin Moradi, Saeed Mozaffari, and Ali Asghar Orouji, "Farsi/Arabic Text Extraction from Video Images by Corner Detection", 6th, IEEE,Iranian conference on Machine Vision and image processing ,Isfahan, Iran,2010. IEEE.
13.     ICDAR. Icdar, "Robust reading and locating database.," http://algoval.essex.ac.uk/icdar/RobustReading.html, 2003.
14.     Matas, J., Chum, O., Urban, M., Pajdla, T., "Robust wide-baseline stereo from maximally stable extremal regions," Image and Vision Computing,Volume-22, pp.761-767, 2004.
15.     Nister, D., Stewenius, H, "Linear time maximally stable extremal regions," 10th European Conference on Computer Vision, Volume-15, pp.183-196, 2008.
16.      Cristianini, N., Shawe-Taylor, J., "An introduction to Support Vector Machines", Cambridge University Press (2000).

17.     Muller, K.R., Mika, S., Ratsch, G., Tsuda, K., Scholkopf, B., "An introduction to kernel-based learning algorithms", IEEE Trans. on Neural Networks,Volume-114, pp.181-201, 2001.

18.      Lukas Neumann, J. Matas, " Text Localization in scenic images", 25th Intl. Conf. on Computer Vision and Pattern Recognition,, 2012.

19.     Anjum Asma and  Gihan Nagib,'Energy Efficient Routing Algorithms for Mobile Ad Hoc  Networks–A Survey', International Journal of Emerging Trends & Technology in computer Science, Vol.3, Issue 1, pp. 218-223, 2012.

20.     Hong-ryeol Gil1, Joon Yoo1 and Jong-won Lee2 ,'An On-demand Energy-efficient Routing   Algorithm for  Wireless Ad hoc Networks', Proceedings of the 2nd International Conference  on Human. Society and Internet HSI'03, pp. 302-311, 2003.

21.     S.K. Dhurandher, S. Misra, M.S. Obaidat, V. Basal, P. Singh and  V. Punia,'An Energy-Efficient On Demand Routing algorithm for Mobile Ad-Hoc Networks', 15 th International conference on Electronics, Circuits and Systems, pp. 958-9618, 2008.

22.     DilipKumar S. M. and Vijaya Kumar B. P. ,'Energy-Aware Multicast Routing in MANETs: A Genetic Algorithm Approach', International Journal of Computer Science and Information Security *(IJCSIS*), Vol. 2, 2009.

23.     AlGabri Malek, Chunlin LI,  Z. Yang, Naji Hasan.A.H and X.Zhang ,' Improved the Energy of Ad hoc On- Demand Distance Vector Routing Protocol', International Conference on Future Computer Supported Education, Published by Elsevier, IERI, pp. 355-361, 2012.

24.     D.Shama and A.kush,'GPS Enabled E Energy Efficient Routing for Manet', International Journal of Computer Networks (IJCN), Vol.3, Issue 3, pp. 159-166, 2011.

25.     Shilpa jain and Sourabh jain ,'Energy Efficient Maximum Lifetime Ad-Hoc Routing (EEMLAR)', international Journal of Computer Networks and Wireless Communications, Vol.2, Issue 4, pp. 450-455, 2012.

26.     Vadivel, R and V. Murali Bhaskaran,'Energy Efficient with Secured Reliable Routing  Protocol  (EESRRP) for Mobile Ad-Hoc Networks', Procedia Technology 4,pp. 703- 707, 2012.