A Hybrid Approach to detect a text in Digital Images

R. Bhavadharani, A. Thilagavathy

Abstract— In this paper we present an effective approach to detect text from an image. Detecting text from an image is of great demand. Image that contains text usually have a wide variety of applications. In this paper we present an edge based approach to detect text from an image. Given an input image it is first pre- processed to remove any noise; the image is grouped into different color layers and a gray component. The region with a possibility of text in the image is detected. This approach utilizes the fact that usually the color data in text characters is different from the color data in the background. The process of localization involves further enhancing the text regions by eliminating non- text regions. Dilation operation is then performed on those texts to group them, thereby eliminating pixels that are far-away from the candidate text regions. The robustness of this approach is concluded by conducting various experiments on images with complex style, type, background, scaling etc. The results from this proposed approach shows us that it is far superior to the existing approaches in terms of false positive and false negative. Experiments are conducted over a large volume of ICDAR datasets to demonstrate the effectiveness of our proposed method**.**

Index Terms— Dilation, Edge based approach, False negative, False positive, Localization, pixel, pre- process,

1 INTRODUCTION

RECENT studies in the field of computer shows a great amount of interest in content retrieval from images and videos [1], [2], [3]. This content can be in the form of objects, textures, colors, shapes as well as relation between them.

During the last decade, a lot of techniques are available to extract the text from an image. This includes *texture based methods, connected component based methods* and *edge based methods* [4], [5]. Among these techniques, text detection base approaches are of greater hike due to the precise information in the text.

Texture based methods [6], [7], [8] are used to distinguish text regions from their background and or other regions within the image. In Connected Component based methods, the image is divided into a set of smaller components known as connected components and then it [9], [10] recursively merge the smaller components to form larger ones. It usually scans an image and groups its pixels into components based on pixel connectivity *i.e*. all pixels that lies within a selected component share similar pixel intensity values and are in some way connected to each other. Edge based methods [11] are used to define the boundaries between regions in an image, which helps with segmentation and object recognition. Edge detection usually refers to the process of identifying and locating shared discontinuities in an image.

Other method for text detection includes techniques such as support vector machines[12], [13], k-means clustering [14] and neural networks[15] etc

- *R.Bhavadharani is currently pursuing masters degree program in the department of Computer Science and engineering affiliated to Anna University, India, PH-8807620272. E-mail: bhavadharaniravi@yahoo.com*
- *A.Thilagavathy is currently working as an Associate Professor in the department of Computer Science and engineering affiliated to Anna University, India, PH-9941354674. E-mail: thilsmailbox@gmail.com*
(*This information is optional; change it according to your need.)*

2 AN OVERALL APPROACH

Text embedded in an image has a wide variety of applications. Given an input image it is first preprocessed to remove any noise, and then the image is grouped into different color layers and a gray component. The region with a possibility of text in the image is detected. This approach utilizes the fact that usually the color data in text characters is different from the color data in the background. The process of localization involves further enhancing the text regions by eliminating non- text regions. Dilation operation is then performed on these texts to group those, thereby eliminating pixels that are far away from the candidate text regions. An edge based detection method is used in order to extract text from an image in an effective way. The experiments have also been conducted for images containing different font styles. Also the precision and recall rates(Equation (1) and (2)), have been calculated based on the number of correctly detected words in an image.

The *Precision rate* is defined as the ratio of correctly predicted letters to the sum of these correctly predicted words in addition to false positives. *False positives* are those regions in the image which are accurately not characters of text, but have been detected unknowingly.

$$\text{Precision rate} = \frac{\text{Correctly predicted letters}}{\text{Correctly predicted letters} + \text{False positives}} \times 100 \quad (1)$$

The *Recall rate* is defined as the ratio of correctly predicted letters to the sum of these correctly detected letters in addition to false negatives. *False negatives* are those regions in the image which are actually text characters, but have not been detected

$$\text{Recall rate} = \frac{\text{Correctly predicted letters}}{\text{Correctly predicted letters} + \text{False negatives}} \times 100 \quad (2)$$

3 Related Search

Xu Zhao et al [20] proposed a technique called corner based approach to detect text and caption from videos. This method is based on the fact that there exists dense and orderly presences of corner points in a text.

Palaiahnakote Shivakumar et al [21] proposed a method based on Laplician approach in order to detect text from video. Usually a text is horizontally oriented whereas this method is able to handle text from any orientation. Later K-means clustering is performed over the text.

Xiaoqian Liu et al [22] proposed a technique to extract captions from videos. they have presented a novel stroke like edge detection method along with the highlights of temporar feature in extracting texts.

4 Text Detection

Text detection [16], [17] refers to the process of identifying and locating sharp discontinuities in an image. The discontinuities are abrupt changes in pixel intensity which characterize boundaries of objects in a scene. It significantly reduces the amount of data and filters out unwanted information while preserving the important structural properties.Text detection consists of the following phases

4.1 Text Region Extraction

An image is taken as an input. From that image, the region with a possibility of text is detected. A Gaussian pyramid is formed as a result of filtering the input image with the Gaussian kernel and it down samples the given image. For this process usually a Gaussian filter is used. Filtering is the process of suppressing the noise without edges of the text frame being blurred. These images are next convolved with directional filters at different orientation kernels for edge detection in various directions. Text characters of same string appearing closer to each other are detected as text. Fig1 sketches about the Gaussian pyramid.

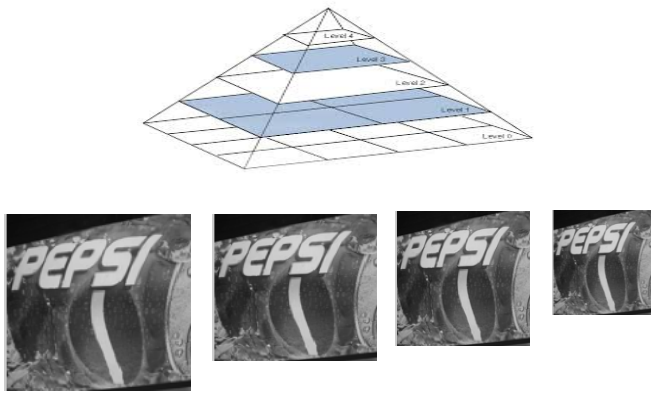


Fig1. Gaussian Pyramid

4.2 Localization

The process of *localization* consists of enhancing the text regions by eliminating the other regions that are not text [18], [19]. In text usually all characters appear closer to each other. Localizattion of the given image is shown in Fig2

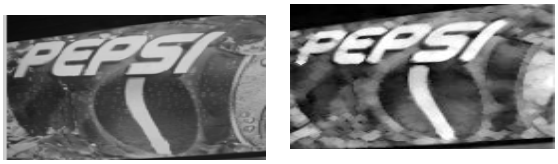


Fig2. (a) Before Localization (b) After Localization

Thus morphological dilation operation can be performed inorder to eliminate the pixels that are far away from the region. Dilation adds pixels to the boundaries of objects in an image. The number of pixels added to the object in an image depends on the size and shape of the structuring element used to process the image. In short, it is basically an operation which expands or enhances the region of interest using the structuring element. The outcome of this process consists of image with some non- text regions are noise which will be eliminated by the upcoming phases. Fig3 shows the result before and after dilation.

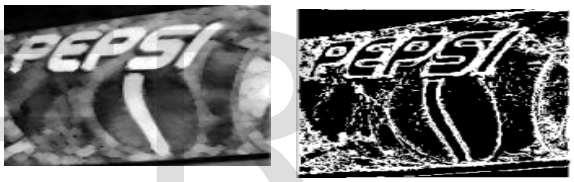


Fig3. (a) Before Dilation (b) After Dilation

4.3 Morphological Operations

Various morphological operations are performed during this process. *Thinning* is a morphological operation that is used to remove selected foreground pixels from binary images. It can be used for several applications, but is particularly useful for skeletonization. The result of this process is shown in fig 4. During this process it is commonly used to tidy up the output of edge detectors by reducing all lines to single thickness. The behavior of the thinning operation is determined by a structuring element. The thinning of an image I by a structuring element J is given by (Equation (3)):

$$\text{thin}(I,J) = I - \text{hit-and-miss}(I,J) \quad (3)$$

where the subtraction is a logical subtraction defined by the $X - Y = X \cap \text{NOT } Y$

(a) (b)

Fig4. (a)Before thinning (b) After thinning

Thresholding [18] becomes easy and it is an efficient tool which is used to separate objects from the background.

Fig5 depicts the result before and after thresholding. To eliminate the non- text regions a ratio is set by experimenting on different kinds of images and an average value is choosen. Here Otsu global thresholding method [19] is used for binarization. For this process a threshold value is choosen and the pixels which have values greater than this threshold value appears as whiteand others as black. The binarization of an image is given by (Equation (4)):

$$Ibin(x,y) = \{ 1, \text{if } Igray(x,y) \le K \quad 0, \text{if } Igray(x,y) \le K \qquad (4)$$

(a) (b)

Fig5. (a) Before thresholding (b) After Thresholding

4.4 Text Detection

Texts are detected based on its properties. These properties are used to distinguish text regions from their background. Text characters of same string that appears closer to each other of similar height, size and shape are detected. thus this process generates our targeted output image. Fig6 shows the flow of detecting the text.

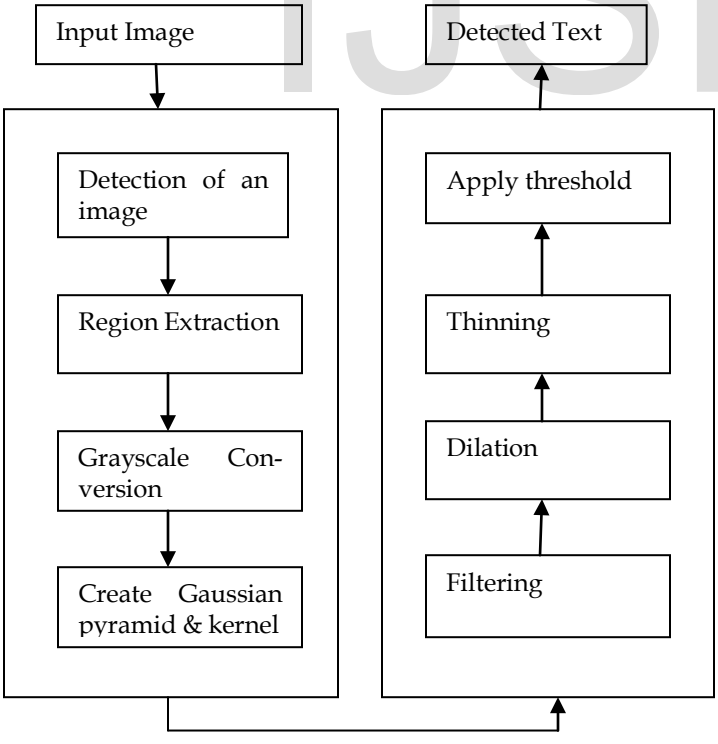


Fig6. Block diagram for detecting text

ALGORITHM FOR TEXT DETECTION

An algorithm for edge based text extraction consists of the following steps:

1. Select the input image tat you want to preprocess.
2. Resize images of different size to equal width frames.
3. Create a Gaussian pyramid and directional kernels to detect the edges from different orientations.
4. Convolve each image in the Gaussian pyramid with Gaussian filter.
5. The resultant image is then dilated which adds pixels to the boundaries of objects in an image.
6. Thinning is then done to remove selected foreground pixels from the binary image.
7. Apply threshold values to separate objects from background.
8. Output image is obtained with text in white pixels against a plain background.

1) (a)

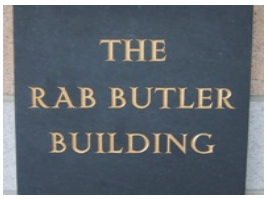
(b)

2) (a)

(b)

3) (a)

(b)



4) (a)

(b)


5) (a) (b)


6) (a) (b)


7) (a) (b)


8) (a) (b)


9) (a) (b)


10) (a) (b)


11) (a) (b)

Fig 7. Output images of Text detection

6 Experimental Results

The result obtained by this algorithm proves that this method is more robust.

Image no:	Image Name	Precision Rate (%)	Recall Rate (%)
1	Emergency stop	76.47	100
2	Bharti	100	100
3	Pepsi	26.75	100
4	The Rab Butler Building	100	100
5	Adidas	66.66	100
6	NBC	33.33	100
7	Tesco Value washing up liq-uid	69.44	92.00
8	Infosys Public services	95.45	100
9	Opera Software	93.33	100
10	Volvo	41.66	100
11	Success	77.77	100

The precision and recall rates obtained by this algorithm are higher and more precise than the simple edge based algorithm. The above result shows us thst it is more robust in terms of lightining variance, scaling and orientation. The overall precision and recall rate is 73.64% and 96.88% which is quite high when compared to any other methods used.

7 Future Enhancements

Our future aim is to test the input images on various factors such as scaling and lightning conditions such that it would be invariant to scale, lightning as well as orientation changes. Next is to implement an enhanced morphological cleaning of images which could result in a higher precision rate. Lastly, an interesting test would be to find out the text regions which are hidden behind other objects or the texts which are watermarked within an image.

8 Conclusion

In this paper, we present a new way to detect texts that are embedded in an image. This proposed method is more effective and efficient in calculating the precision and re-

call rates than the contemptory methods. The results obtained by this method are more precise and encouraging. Moreover, it can be applied to any image containing texts with different languages.

REFERENCES

[1] Y. Rui, T. S. Huang, and S. F. Chang, "Image retrieval: Current tech niques, promising directions, and open issues," *J. Vis. Commun. Image Represent*, vol. 10, no. 1, pp. 39–62, 1999.

[2] Y. A. Aslandogan and C. T. Yu, "Techniques and systems for image and video retrieval," *IEEE Trans. Knowl. Data Eng.*, vol. 11, no. 1, pp. 56–63, Jan./Feb. 1999.

[3] W. M. Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain, "Content-based image retrieval at the end of the early years," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 12, pp. 1349–1380, Dec. 2000

[4] X. Tang, X. Gao, J. Liu, and H. Zhang, "A spatial-temporal approach for video caption detection and recognition," *IEEE Trans. Neural Netw.*, vol. 13, no. 4, pp. 961–971, Jul. 2002

[5] K. Jung, K. I. Kim, and A. K. Jain, "Text information extraction in images and video: A survey," *Pattern Recognit.*, vol. 37, no. 5, pp.977–997, 2004

[6] K. Kim, K. Jung, and J. Kim, "Texture-based approach for text detection in images using support vector machines and continuously adaptive mean shift algorithm," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 12, pp. 1631–1639, Dec. 2003.

[7] Y. Zhong, H. Zhang, and A. K. Jain, "Automatic caption localization in compressed video," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 4, pp. 385–392, Apr. 2000

[8] W. Mao, F. Chung, K. K. M. Lam, and W. Sun, "Hybrid chinese/English text detection in images and video frames," in *Proc. 16th Int. Conf. Pattern Recognit.*, 2002, vol. 3, pp. 1015–1018.

[9] E.K. Wong and M. Chen, "A New Robust Algorithm for Video Text Extraction," Pattern Recognition, vol. 36, pp. 1397-1406, 2003.

[10] V.Y. Mariano and R. Kasturi, "Locating Uniform-Colored Text in Video Frames," Proc. Int'l Conf. Pattern Recognition, pp. 539-542, 2000

[11] M. R. Lyu, J. Song, and M. Cai, "A comprehensive method for multilingual video text detection, localization, and extraction," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 15, no. 2, pp. 243–255, Feb. 2005.

[12] 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.

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

[14] Victor Wu, Raghavan Manmatha, and Edward M. Riseman, "TextFinder: An Automatic System to Detect and Recognize Text in Images," IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 21, No. 11, November 1999.

[15] Rainer Lienhart and Axel Wernicke, "Localizing and Segmenting Text in Images and Videos," IEEE Transactions on Circuits and Systems for Video Technology, Vol.12, No.4, April 2002.

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

[17] Xiaoqing Liu and Jagath Samarabandu, "Multiscale edge-based Text extraction from Complex images," IEEE, 2006

[18] T. Romen Singh, Sudipta Roy, O. Imocha Singh, Tejmani Sinam, Kh. Manglem Singh, "A New Local Adaptive Thresholding technique in Binariztion," IJCSI International Journal of Computer Science Issues, Vol 8, Issue 6, No 2, November 2011, ISSN: 1694−0814

[19] Otsu, N., "A Threshold selection method from gray level histograms," IEEE *Trans. Syst. Man Cybern* 9, 62−66 (1979).

[20] Xu Zhao, Kai-Hsiang Lin, Yun Fu, Yuxiao Hu, Yuncai Liu and Thomas S. Huang, "Text From Corners: A Novel Approach to Detect Text and Caption in Videos" IEEE Transactions on Image Processing, VOL. 20, NO. 3, March 2011.

[21] Palaiahnakote Shivakumara, Trung Quy Phan, and Chew Lim Tan, "A Laplacian Approach to Multi-Oriented Text Detection in Video", IEEE Transactions on Pattern Analysis and Machine IntelligenceE, VOL. 33, NO. 2, February 2011

[22] Xiaoqian Liu and Weiqiang Wang, "Robustly Extracting Captions in Videos Based on Stroke-Like Edges and Spatio-Temporal Analysis", IEEE transactions on multimedia, vol. 14, no. 2, April 2012