A Comparative Study of Segmentation in Mixed-Mode Images

Asha D
M.Tech Student, Dept of ECE
MCE Hassan, Karnataka,India

Dr.Shivaprakash Koliwad
Professor, Dept of ECE
MCE, Hassan
Karnataka, India

Jayanth.J
Lecturer, Dept of ECE
G.S.S.S.I.E.T.W, Mysore
Karnataka, India

ABSTRACT

The detection and extraction of text regions in an image is a well known problem in the computer vision research area. Text extraction is a critical and essential step as it sets up the quality of the final recognition result. It aims at segmenting text from background, i.e isolating text pixels from those of background. Since readymade mixed mode image data is not available, it is necessary to create our own database. The database plays an important role as segmentation is to be done in an image. In educational videos and in presentation of lectures, graphic play an important role. In television industry text and images are simultaneously transmitted. In such similar application compression of data and bandwidth play an important role. To achieve better compression and bandwidth utilization properly, an efficient segmentation technique is necessary. In this paper, we analyze mixed mode images by two methods.

General Terms

Image processing, Edge based text extraction, Connected Component based text extraction.

Keywords

Text Segmentation, Edge Detection, Recognition, Localization, Extraction, Enhancement, Connected Component, DCT**.**

1. INTRODUCTION

Multimedia documents contain texts, graphics and pictures. Therefore the scanned documents must often be segmented before other document processing tasks. Texts within an image play an important role in retrieval systems as they contain plenty of valuable information and can be easily extracted comparing with other semantic contents.

Segmentation is an important issue in document analysis. There are two types of segmentation algorithms namely, region based and pixel based algorithms. We compare two algorithms for block-based text segmentation. Our evaluation is based on two criteria: the accuracy of segmentation and the robustness across different types of images. We compare the different segmentation algorithms based on the number of text and non-text blocks wrongly classified.

Text data is particularly interesting, because text can be used easily and clearly to describe the contents of an image, it can be easily extracted compared to other semantic contents and it enables applications such as keyword-based image search, automatic video logging, and text-based image indexing. Various methods have been proposed in the past for detection and localization of text in images. We have chosen MATLAB for the simulation of the algorithms as it is a simple yet a powerful software tool.

The following three schemes are considered for segmentation in mixed mode images: i) Edge based text segmentation method ii) connected component based text segmentation method and iii) DCT based method.

2. ALGORITHMS FOR TEXT SEGMENTATION

In this section we briefly describe the two different segmentation algorithms.

2.1 Edge Based Text Segmentation Approach

The basic steps of the edge-based text extraction algorithm [2, 3] are as illustrated in Fig.1.

- Create a Gaussian pyramid by convolving the input image with a Gaussian kernel and successively down-sample each direction by half. (Levels: 4)
- Create directional kernels to detect edges at 0^0, 45^0, 90^0 and 135^0 orientations.
- Convolve each image in the Gaussian pyramid with each orientation filter.
- Combine the results of step 3 to create the Feature Map.
- Dilate the resultant image using a sufficiently large structuring element (7x7) to cluster candidate text regions together.
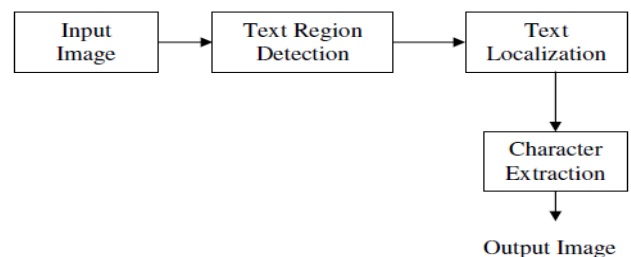- Create final output image with text in white pixels against a plain black background.

Fig 1: Basic Block diagram for Edge based text extraction.

2.2 Connected Component based Text Segmentation Approach

The basic step of the connected-component text extraction algorithm [2, 3] is as follows and is illustrated in Fig 2.

- Convert the input image to YUV color space. The luminance(Y) value is used for further processing. The output is a gray image.
- Convert the gray image to an edge image.
- Compute the horizontal and vertical projection profiles of candidate text regions using a histogram with an appropriate threshold value.
- Use geometric properties of text such as width to height ratio of characters to eliminate possible non-text regions.
- Binarize the edge image enhancing only the text regions against a plain black background.
- Create the Gap Image using the gap-filling process and use this as a reference to further eliminate non-text regions from the output.
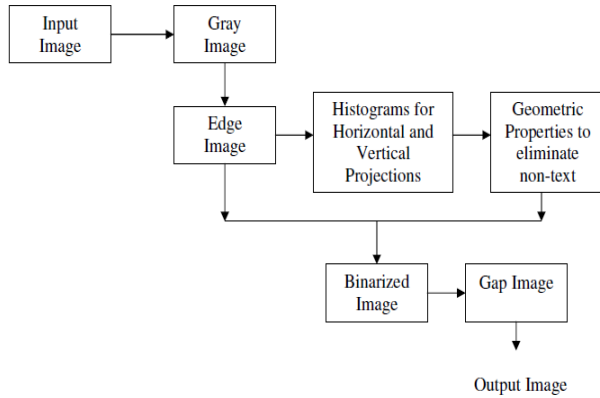
Fig 2: Basic Block diagram for Connected Component based text extraction.

2.3 DCT Based Approach

DCT based algorithms have found use in compression algorithms based on transform coding. The energy distribution of text blocks amongst some DCT coefficients is significantly higher compared to the non text blocks. Thus segmentation can be obtained by choosing an appropriate set of DCT coefficients which captures the difference between the text and the non text blocks and comparing their absolute sum to a threshold. Although DCT is much more expensive than the edge based and connected component based approach, we note that it is the most popular transform for block based compression algorithms and has been adopted as the algorithm of choice in JPEG, MPEG and H.261 standards. For such algorithms, the incremental computational cost of summing some of the DCT coefficients is negligible.

3. EVALUATION METHODOLOGY

This section is divided into two parts. The first section describes the test image database on which the segmentation algorithms are tested. The second section discusses the evaluation criterion for comparing different segmentation algorithms.

3.1 Test Image Database

There is no mixed mode image database that is available for test and comparison. Depending upon the application the mixed mode image data is created. The developed database contains several images corresponding to the following applications: Indoor name plate images, Type written text and photograph of a person. Scanned image data, Chalk on Black Board, Different objects with different orientation and illumination, A data set with different languages and Text displayed using an overhead projector and hands of a person. We have developed our own database containing several images corresponding to the above applications.

3.2 Evaluation Criteria

As stated earlier, we are interested in two main criteria: accuracy and robustness. We will measure accuracy in terms of false negatives, i.e., the number of actual text blocks segmented as non-text blocks, and false positives, i.e., the number of non-text blocks which are segmented as text. Our objective is to minimize the total error, i.e., the sum of the false negatives and false positives, subject to the constraint that false negatives are less than the false positives. We give greater emphasis to false negatives, because that way we ensure that text is compressed using the correct compression parameters/algorithm.

The Table 1 illustrate the description for different mixed mode images that we have used for analysis.

Table 1. Description of Mixed Mode images

Image type	Description
T1: Indoor N	Indoor name plate images with different font sizes, perspective distortion, colors and strong reflection
T2: TextCV	Type written text and photograph of a person
T3: VarFontP	Scanned image data of varying font size and pictures
T4:ChalkB	Written material on black board with different colored Chalk and different orientation.
T5:ObjL	Object label image with different font size, colors and orientation alignments
T6:TextDL	Different language text images with different font size and color
T7: OvP	Text displayed using an overhead projector with hands of a person
T8:ScbcN	Share certificates, Bank cheques and news paper

The experimentation of the proposed algorithm was carried out on the data set consisting of different images such as indoor, outdoor, posters etc. These test images vary with respect to scale, lighting and orientation of text in the image. The significance of testing the algorithms on variations of scale, lighting and orientation is to determine the robustness of each technique with respect to variance in these conditions, and also to determine where each technique is successful and where it fails.

The performance of each technique has been evaluated based on its precision and recall rates obtained. Precision and Recall rates are calculated as follows:

$$\text{Precision rate} = \frac{\text{Correctly detected words}}{\text{Correctly detected words + False positives}} \times 100\%$$

$$\text{Recall rate} = \frac{\text{Correctly detected words}}{\text{Correctly detected words + False negatives}} \times 100\%$$

4. EXPERIMENTAL RESULTS

Table 2. Statistical parameter calculations for the various set of image

| SL No	Image Type	Image Source No.	Parameters	
1	Indoor N	14	SD =55.4719 M=116.3715 E=389,222.797	R=0 255 H =7.0666 S =288.137
2	TextC V	50	SD =48.8395 M=231.6857	R =0 255 H =1.6392

*SD-Standard Deviation, M-Mean, E-Energy, R-Range, H-Entropy, S-Sum

Table 3. Results from Edge based algorithm

SL.No.	Image Type	Image Source No.	Precision Rate (%)	Recall Rate (%)
1	T1:IndoorN	14	74.60	77.94
2	T2:TextCV	50	82.85	91.78
3	T3:VarFontP	17	85.71	88.88
4	T4:ChalkB	15	60.00	99.00
5	T5:ObjL	18	85.11	76.88
6	T6:TextDL	20	88.28	99.00
7	T7:OvP	14	20.45	90.00
8	T8: ScbcN	20	68.56	93.70

Table 4. Results from Connected component based algorithm

SL.No.	Image Type	Image Source No.	Precision Rate (%)	Recall Rate (%)
1	T1:IndoorN	14	80.55	85.29
2	T2:TextCV	50	81.39	95.89
3	T3:VarFontP	17	93.75	55.55
4	T4:ChalkB	15	68.96	83.33
5	T5:ObjL	18	84.62	47.31
6	T6:TextDL	20	87.50	56.32
7	T7:OvP	14	24.24	80.00
8	T8: ScbcN	20	63.04	71.34

(a)

(b)
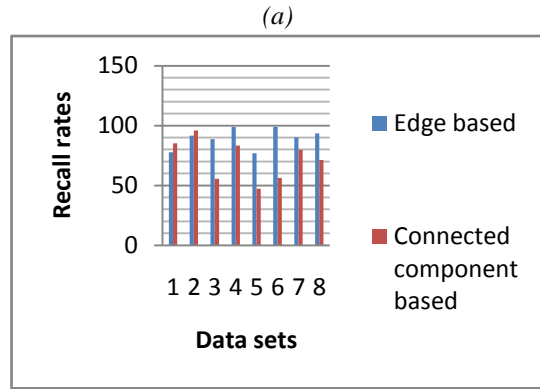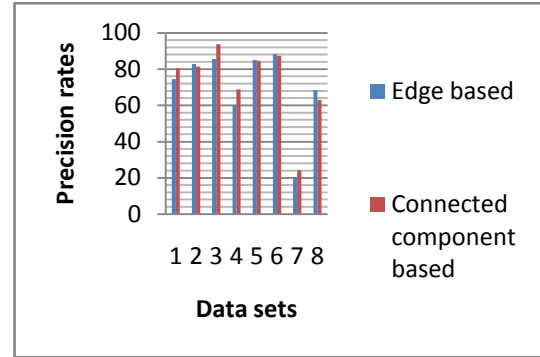
Fig 3: Bar graphs for (a) precision rates and (b) recall rates

The precision and recall rates calculated for the connected component algorithm takes into consideration each text line as one text region. The edge based algorithm takes into consideration each character of text to calculate precision and recall rates. In order to have a common method to evaluate and compare the results from each algorithm, in this paper, each text is considered in the calculation of precision and recall rate. False positives are the number of connected regions obtained by the algorithm, which are not text. False negatives are the total number of text in the test image minus the texts which were not detected by the algorithm. The overall average precision and recall rates shown in TABLES III and IV are over a varied data set of images. The average precision rate obtained by this project for the edge based algorithm is 70.69% and the average recall rate obtained is 89.65%. Thus, the recall rate as stated is more than the precision rate obtained by the edge based algorithm. The average precision rate obtained by experimentation for the connected component algorithm is 73.01% and the average recall rate obtained is 71.88%.

Table 5. Execution times for edge based and connected component based algorithms.

SL.No.	Image Type	Image Source No.	Edge Based (in seconds)	Connected Component Based (in seconds)
1	T1:IndoorN	14	12.4467	53.7523
2	T2:TextCV	50	29.3735	84.8674
3	T3:VarFontP	17	26.8638	74.3071
4	T4:ChalkB	15	8.4978	63.9093
5	T5:ObjL	18	26.6292	79.3897
6	T6:TextDL	20	15.9522	39.7047
7	T7:OvP	14	5.37313	30.8288
8	T8: ScbcN	20	34.25422	70.7136

From the Table 5, we can clearly see that the execution time for edge based algorithm is less when compared to connected component based algorithm.
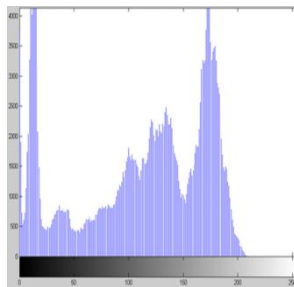

**Fig 4: Test Image (T5)**   **Fig 5: Result of Edge based algorithm**

Fig 6: Result of Connected Component based algorithm and Histogram of image (T5)
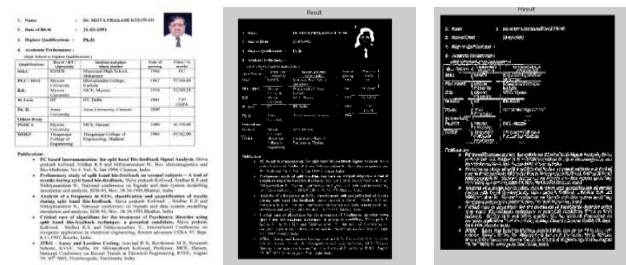
Table 6: Statistical parameter calculations for Test Image(T5)

Range(R)	0 255
Entropy (H)	7.0401
Std_Deviation (SD)	71.0895
Mean (M)	99.2654
Sum(S)	9.8573e+003
Energy (E)	7.4737e+008

The snap shots below shows the resultant images for the segmentation schemes considered for different data sets as listed in Table 1.
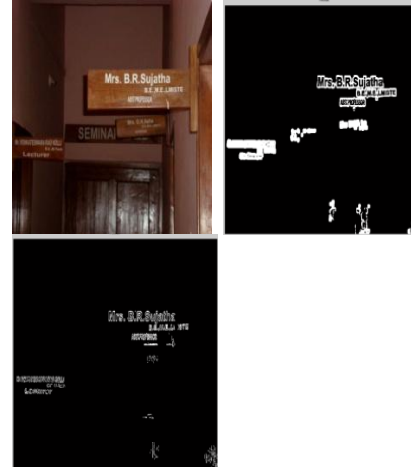
Fig 7:Test Image(T1)   Edge based   Connected Component

Fig 8:Test image(T2)   Edge based   Connected component
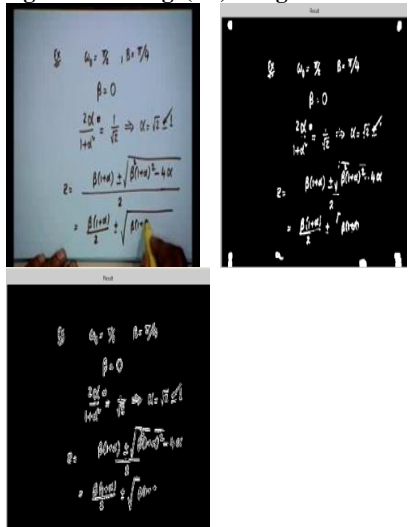

Fig 9:Test image(T3)   Edge based   Connected component

Fig 10:Test image(T4)   Edge based   Connected component

Fig 11:Test image(T6)   Edge based   Connected component

Fig 12:Test image(T7)   Edge based   Connected component

Fig 13:Test image(T8)   Edge based   Connected component

5. CONCLUSION

An edge-based text region extraction algorithm can localize and extract text from real scenes. Experimental results show that the two algorithms used are efficient in localizing and extracting text-based features. It is robust with respect to font sizes, styles, color/intensity, orientations, alignment / layout and perspective of text. In an indoor environment when images were captured, for this image data set it is also possible to detect text-based landmarks in real-time. For the entire data set containing wide variety of images, the overall performance was good. The proposed methods are: Not sensitive to image color/intensity. Unlike most of the connected component based methods, it does not need any assumptions for color/intensity within the same character. Distinguishes text regions from texture-like regions, such as window frames, wall patterns, etc., by using the variance of edge orientations.

6. ACKNOWLEDGMENTS

7. REFERENCES

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

[2] Xiaoqing Liu and Jagath Samarabandu, "An Edge-based text region extraction algorithm for Indoor mobile robot navigation", Proceedings of the IEEE, pp. 613-624, July 2005..

[3] Xiaoqing Liu and Jagath Samarabandu,"Multiscale edge based text extraction from complex images", IEEE,pp.245-252,2006.

[4] N. Otsu, "A threshold selection method from gray-level histograms", IEEETransactions on Systems, Man, and Cybernetics, vol. 9, no. 1, pp. 62–66, 1979.

[5] Jiang Gao and Jie 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, pp. II–84–II–89.

[6] Jain, A.K. and B. Yu, "Automatic text location in images and video frames",Pattern Recognition, vol. 31, no. 12, pp 2055-2076, 1998.

[7] Qixiang Ye, Wen Gao, Qingming Huang, "Automatic text segmentation from complex background", IEEE International Conference on Image Processing, Singapore, Vol.5, pp.2905-2908, Oct. 2004.

[8] Rafael. C. Gonzalez, Richard E. Woods, "Digital Image Processing", Prentice- Hall of India, Second edition.

[9] R. Lienhart and A. Wernicke, "Localizing and segmenting text in images and videos", IEEE Trans. Circuits Syst. Video Technology, Vol. 12, Nr. 4, 2002, pp. 256-268.