Text Extraction from Images: A Survey

A.J.Jadhav
Department of Information
Technology, University of Pune

Vaibhav Kolhe
Department of Information
Technology, University of Pune

Sagar Peshwe
Department of Information
Technology, University of Pune

Abstract— This is survey paper on text extraction from images. Images contain various types of useful information that should be extracted whenever required. A large number of algorithms and methods are proposed to extract text from the given image, and because of that a user will be able to access the text from any image. Extraction of this information involves text region detection, text localization, tracking, character extraction, enhancement, and recognition of the text from a given image. Variations in text may occur because of differences in size, style, orientation, alignment of text, and low image contrast, composite backgrounds make the problem during extraction of text. The purpose of this paper is to classify and review various text extraction algorithms, discuss working and performance evaluation, and finding a technique for getting maximum accuracy.

Keywords— Text extraction, Text detection, Text localization, Text retrieval, OCR.

I.   INTRODUCTION

Text in the image contains useful information which helps to acquire the overall idea behind the image. Character extraction from image is important in many applications. It is a difficult task due to variations in character fonts, sizes, styles and text directions, and presence of complex backgrounds and variable light conditions. Several methods for text (or character) extraction from natural scenes have been proposed. If we develop a method that extracts and recognizes those texts accurately in real time, then it can be applied to many important applications like document analysis, vehicle license plate extraction, text- based image indexing, etc and many applications have become realities in recent years [1]. Educational and training video and TV programs such as news contain mixed text-picture graphics regions. Classification of region is helpful in object-based compression, manipulation of text and accessibility. Text regions may contain very useful information regarding the image. Before we go for actual text extraction, first we should study the properties of text. Text contains various characteristics like size, motion, color, edge etc.

1) Size: differences in text size can make problem, but it can be minimized by assuming specific data during text region detection process.

2) Color: intensity of color also affects the quality of text extraction. If all characters of same color then extraction process become more simple and effective.

3) Motion: this property of text usually applied to the videos containing text, and this refers to the movement of text in vertical or horizontal direction.

4) Edge: edges are reliable feature of text as compared with the other features like color layout or motion.

In this paper, II section described various algorithms based on the text extraction. Architecture of proposed system is discussed in section III. Conclusion is given in section IV.

II.   RELATED WORK

Text extraction process mainly consists of five important phases: Text region detection, text localization, tracking, character extraction, text recognition. From which first two (text region detection, text localization) stages are more important and also they are more difficult to implement. The output of text extraction is mainly dependant on these two phases. There are various ways to complete these phases, some of the techniques are:

2.1 Edge based text extraction:

Edge based extraction is one of the more efficient method because edge is most reliable feature as compare to others like layout, color, or orientation etc , and this methods focus on the high contrast between background and actual text. The edges of the text boundary are identified and merged, and then several techniques are used to filter out the non-text regions. Some implementations based on edge based technique are as follows:

1) Algorithm by Xiaoqing Liu et al.: Xiaoqing's method [13] consists of three stages:

a) Candidate text region detection: In this stage a feature map is generated using three main characteristics of edge viz. strength, density and orientation. A feature map is binary image where pixel intensity gives possibility of text.

b) Text region localization: In this stage morphological dialation operator is used. There are two constraints utilized to find non text regions, first for finding very small isolated blocks and second for filter out the block whose width is very small than that's corresponding height.

c) Character extraction: Here existing OCR engines were used for character extraction.

This can only deal with printed characters against clean backgrounds and cannot handle characters embedded in shaded, textured or complex backgrounds. Results are shown in Fig.1

(Original)   (After extraction)

Fig. 1. Xiaoqing Liu's Algorithm

*2) Algorithm by Xin Zhang et al.:*Xin[7] proposed this two phase method:

a) Text background removal: For this first transition map model is utilized and to improve the accuracy of text extraction rate of first model second method edge based text detection is used. In this method, two methods are combined, and because it this method called as color-edge combined algorithm.

 b) Text extraction: in this phase the image is binarized and passed to OCR model for character extraction.

2.2 Region Based text extraction:

In region based methods, we consider the properties of colour in text or the variance related to background. Here we first find small regions and then merge them successively in a large region. A geometrical analysis is done during merging process, to filter out text and non text regions in the image. Many methods are proposed for text extraction which is based on region based text extraction method. Here are some of the main studies are as follows:

1) Algorithm by bunke and Kronenberg : Kronenberg and Bunke [4], proposed a algorithm for "Identification of Text on Colored Book and Journal Covers", in this they minimize color variations by applying clustering methods in pre-processing step. They proposed two methods:

a) Top down analysis: in this phase the image is split in vertical and horizontal directions alternatively. The output is in rectangular shaped blocks and text containing at least two colors. Depending on this information we reject homogeneous regions means regions having no text.

b) Bottom up analysis: it detects homogeneous regions using a region growing method. Beginning with a starting pixel, pixels are merged if they are from to the identical cluster. We know that characters of printed text generally do not touch each other; several regions are detected for a text region.

After this the outputs of two methods are combined to distinguish between text and non text regions. After this phase region is binarized using previously gathered information. And this is given as an input to OCR. This method not only limited for book covers, we can use it for other types of images.

2) Algorithm by Otsu: Otsu is one of the most important methods used for binarizing the image. In Otsu algorithm we consider two aspects first is text and another is background, and calculates threshold value in such way that it should minimizes difference between the text and background [11]. Although the method is promising, it still depends on global threshold, which is not a good measure for differentiating between text and non-text regions for document images with very variable text and background [9].

Fig 2.Otsu algorithms output

3) Algorithm by debapritam sarkar and raghunath :

Debapritam [5] implemented an algorithm to detect the hand written text and it is divided in six steps:

a) Graphical Smoothing: In this they considered a square box of length n X n pixels. First they checked whether 50% or more square box is filled with black pixels, if yes then they fill complete square with black pixels to achieve smoothing of the image.

b) Smoothing by Neighbouring Black Squares: If 3 or more squares around a white square are black then the white block is also made black to have better smoothing.

c) Calculating the Height of the Component.

   d) Creation and Movement of Template:

A temporary block is created, and scanned for total no of black pixels in it. If this number is greater then the number of the previous position of the template, the previous information is discarded. Otherwise the previous position of the template is of current interest and the central line of the template in the horizontal direction is drawn and is called midline. This process continues till scanning of image is not completed.

e) Joining of Midlines: here we first calculate distance between two midlines and then joined. For joining two methods are implemented, in first they draw straight line between two midlines and in second method one midline extended to the other.

   f) Extraction and Linking.

2.3 Texture-Based text extraction :

   We all know that every text has its own texture, and texture based text extraction method we use that texture as distinguishing part from the background. Various methods are proposed to implement Texture-Based Technique[14].
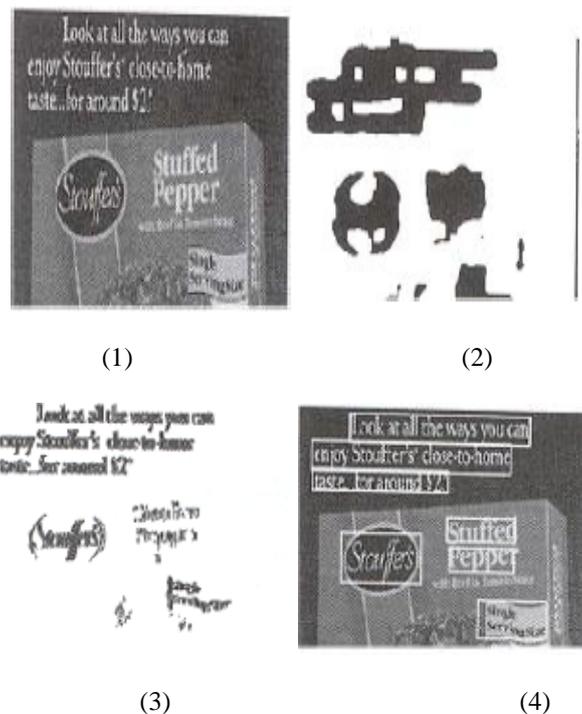

(1)   (2)

(3)   (4)

Fig3. Texture based technique.

1) Algorithm by Chu duc et al.: Chu duc[10] proposed this architecture for car license plate text detection. A scale and rotation invariant text descriptor is used which describes alignment, regularity, similarity and connectivity of segments. Method consists of following stages:

a) Improved connective Hough Transform (ICHT): it records all possible line segments passing through each extracted edge point.

b) Unconstrained Fast LP Detection: Reduce and sub-sample LS map for extracting dense region.

c) ICHT Texture Descriptor for LP Verification: candidates are filtered using aspect ratio constraint which is in between 2 and 5. After this false candidates are eliminated. Then, before the normalization in size remaining candidates are expanded by ten percent in both width and height.

2) Algorithm by Kwang et al.: Kwang[6] implemented a texture based algorithm to detecting text from image for this he used support vector machine (SVM) to detect the texture properties. Next, text regions are identified by Continuously Adaptive mean shift algorithm (CAMSHIFT) to the results of the texture analysis. The output of CAMSHIFT and SVMs gives robust and efficient text detection. For SVM we have to train it so that it produces good results. The accuracy of text extraction is totally depends on training given to the SVM.

2.4 Morphological Based Text Extraction :

Morphology is a geometrical based approach for image analysis. It is used to extract important text features from the processed images. The feature still can be maintained, despite the change in the lighting condition or text color.

1) Algorithm by Rama Mohan et al. The method considers that edge detection is more effective in text extraction. Basic operators of mathematical morphology are used to perform the edge detection. The algorithm is used to find out the connected component. By considering the gray levels of the components their variance is found out for each connected component, when components are found then labeling is done. After selecting the components whose variation is less than threshold value the text can be extracted [8]. This method consists of four steps:

a) Edge extraction
b) Text candidate region formation
c) Labeling of text candidate regions
d) Elimination of non text region

2) Algorithm by Jui-Chen Wu et al.: Jui-Chen Wu [9] presented a text line extraction algorithm for extracting text regions from jumbled images. The method defines a set of morphological operations for extracting important contrast regions. The main steps of this algorithm are :

a) Feature extraction: The relative contrast between texts and their background is an important feature for text line detection. A novel morphology-based scheme for extracting the high contrast feature for locating all possible text lines is used for feature extraction.

b) Text candidate selection: a labeling technique is used to select all possible text lines from the analyzed image.

c) Candidate verification: After candidate selection a verification process is carried out. The text verifications done on the basis of regularities of character size, the ratio between character width and height, and the period of characters.
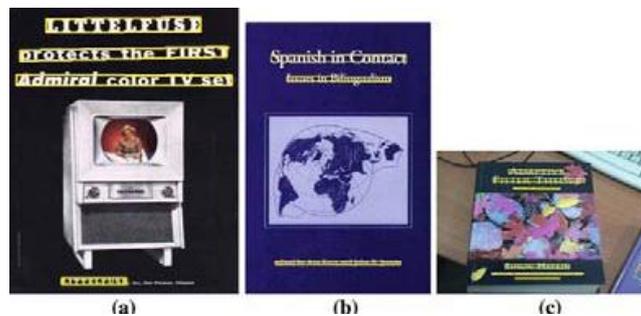

Fig.4 Jui-Chen Wu's Algorithm.

III. PROPOSED ARCHITECTURE

Up till now we had seen various algorithms for text extraction from images but they are proposed for specific applications like License Plate Localization [10], text on covers of books [4], page segmentation [6], etc. There is no general purpose system for text extraction from images and hence the proposed system is very useful. The three distinguishing characteristics of text embedded in images are edge strength, density and the orientation variance .These characteristics can be used as main components of detecting text. Our method consists of three stages: candidate text region detection, text localization and text extraction.

3.1 Image Pre-processing.

If the image data is not represented in specific color space (i.e. YUV), it is converted to this color space by means of an appropriate transformation.

3.2 Edge Detection.

This step focuses the attention to areas where text may occur. We use a simple method for converting the gray-level image into an edge image. Basically the character contours have high contrast to their local neighbours. As a result, all character pixels as well as some non-character pixels which also show high local color contrast are registered in the edge image. We decide a threshold value and compares it with pixel value, according to that the image is converted in binary image.

3.3 Detection of Text Regions.

The binary image analyzed in order to locate text areas. In processing, the local maxima are calculated; two thresholds are employed to the local maxima. Finally, the exact coordinates for each of the detected areas are used to create bounding boxes.

3.4 Enhancement and Segmentation of Text Regions

First, geometric properties of the text characters like the height, width, and ratio of width to height are used to discard those regions whose geometric features do not fall into the predefined ranges of values. All remaining text candidates undergo another treatment in order to generate the so called text image where detected text appears on a plain

background. After that the binary edge image is generated from the edge image, erasing all pixels outside the previously made text boxes and then binarizing it.

IV.   CONCLUSION

In this paper we provided various text extraction techniques. Though we have large no of algorithms and methods for text extraction from image but none of them provide a adequate output because of deviation in text. The proposed method may give a satisfactory output because it is based on edge detection and it supposed that edge is a more reliable feature of text as compared to others.

REFERENCES

[1]   F. Chassaing , C. Wolf, J. M. Jolion, and, "Text localization, enhancement and binarization in multimedia documents," in Pattern Recognition, Aug. 2002, vol. 2 of Proceedings. 16th International Conference on, pp. 1037–1040.

[2]   D. Crandall, R. Kasturi,and S. Antani, Robust Detection of Stylized Text Events in Digital Video, Proceedings of International Conference on Document Analysis and Recognition, 2001, pp. 865-869.

[3]   Yu Zhong, Hongjiang Zhang, and Anil K. Jain, Automatic Caption Localization in Compressed Video, IEEE Transactions on Pattern Analysis and Machine Intelligence, 22, (4) (2000) 385-392.

[4]   Karin Sobottka, Horst Bunke and Heino Kronenberg,"Identification of Text on Colored Book and Journal Covers", *Document Analysis and Recognition*, 20-22 ,1999.

[5]   Debapratim Sarkar, Raghunath Ghosh ,"A Bottom-Up Approach of Line Segmentation from Handwritten Text" ,2009.

[6]   Kwang In Kim, Keechul Jung, And Jin Hyung Kim ,"Texture-Based Approach For Text Detection In Images Using Support Vector Machines And Continuously Adaptive Mean Shift Algorithm" , *IEEE Transactions On Pattern Analysis And Machine Intelligence*, Vol. 25, No. 12, 2003

[7]   Xin Zhang, Fuchun Sun, Lei Gu, "A Combined Algorithm for Video Text Extraction",2010 Seventh International Conference on Fuzzy Systems and Knowledge Discovery, 2010.

[8]   Rama Mohan Babu, G., Srimaiyee, P. Srikrishna, A., "Text Extraction From Hetrogenous Images Using Mathematical Morphology" Journal Of Theoretical And Applied Information Technology © 2005 -2010 Jatit.

[9]   Jui-Chen Wu · Jun-Wei Hsieh · Yung-Sheng Chen,"Morphology-based text line extraction" Machine Vision and Applications 19:195–207 DOI 10.1007/s00138-007-0092-0,2008.

[10]   Chu Duc Nguyen, Mohsen Ardabilian and Liming Chen ,"Robust Car License Plate Localization using a Novel Texture Descriptor", Advanced Video and Signal Based Surveillance, 2009.

[11]   Reza Farrahi Moghaddam, Mohamed Cheriet "Adotsu: An adaptive and parameterless generalization of Otsu's method for document image binarization" PII: S0031-3203(11)00514-0 DOI:10.1016/j.patcog.2011.12.013

[12]   XiangyunYe, M.Cheriet, and C. Y. Suen, "Stroke-model-based character extraction from gray-level document images". IEEE Transactions on Image Processing, 10(8):1152{1161,2001.

[13]   X. Liu and J. Samarabandu, "An edge-based text re-gion extraction algorithm for indoor mobile robot navigation," inProc. of the IEEE International Conference on Mechatronics and Automation (ICMA 2005), Niagara Falls, Canada, July 2005, pp. 701–706.

[14]   V. Wu, R. Manmatha, and E. M. Riseman, TextFinder: An Automatic System to Detect and Recognize Text in Images, IEEE Transactions on Pattern Analysis and Machine Intelligence, 21 (11) (1999) 1224-1229.