

TEXT EXTRACTION FROM IMAGES: A SURVEY

Kumary R Soumya ¹, Ancy.T.V ², Athulya Chacko ³

¹ Asst. Prof.Dept. Computer Science and Engineering, Calicut University, Jyothi Engineering College, Kerala, India
soumyakr@jecc.ac.in

² Dept. Computer Science and Engineering, Calicut University, Jyothi Engineering College, Kerala, India
ancytv1992@gmail.com

³ Dept. Computer Science and Engineering, Calicut University, Jyothi Engineering College, Kerala, India
athulyachacko15@gmail.com



ABSTRACT

Text extraction is one of the key tasks in document image analysis. Automatic text extraction without characters recognition capabilities is to extract regions just contains text. The text extraction process includes detection, localization, segmentation and enhancement of the text from the given input image. In this paper we present a comparative study and performance evaluation of various text extraction techniques.

Key words: Text Detection, Text Enhancement, Text Extraction, Text Localization

1. INTRODUCTION

Text extraction from an image is a challenging problem because of image contains text due to different size, style, orientation, alignment, low contrast, noise and has complex background structure. This extracted text contains only black text in white background, i.e. it can be recognized by any recognition system. Extracting text from an image or video includes in different applications like document processing, image indexing, video content summary, video retrieval, video understanding etc [1].

The Figure1 shows the steps involved in the text extraction technique. Text detection refers to the determination of the presence of text in a given input image, done by exploiting the discriminate properties of text characters such as the vertical edge density, the texture or the edge orientation variance. Text regions should have high contrast than the background; otherwise it would not be easily readable. This is the basic idea behind text localization, which is referred as locating the text portion in the image. The located portion is extracted during the text extraction phase. The output from this phase is

given to OCR in order to eliminate as many falsely identified text regions as possible.

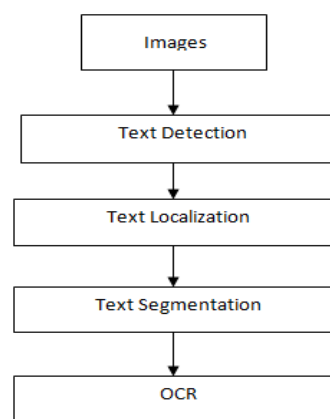


Figure 1: Text Extraction Block Diagram

This paper mainly focuses on different methods for text extraction and presents a survey of techniques which includes region-based technique, edge-based technique, texture-based technique and morphological-based technique. The rest of the paper is organized as follows. In the next Section 2, we briefly introduce the different text extraction techniques. Performance evaluations of described techniques are provided in Section 3 and conclusion is drawn in section 4.

2. TEXT EXTRACTION TECHNIQUES

Text extraction is one of the required stages prior to character recognition. The aim of text extraction is to separate each character so that it can be fed into the recognition stage. This paper discusses about different text extraction techniques such as region-based, edge-based, texture-based and morphological-based techniques.

2.1 Region-Based Technique

Region-based methods use the properties of the colour or gray-scale in a text region and their differences with the corresponding properties of the background. Regarding the image representation, region-based image representations provide a simplification of the image in terms of a reduced number of representative elements. In this representation, objects in the scene are obtained by the union of regions in an initial partition.

Debapratim [2] described the bottom-up approach of Line Segmentation from handwritten text. Line segmentation is a process in which the consecutive lines are extracted or separated from each other from a text. For a line segmentation of handwritten text, first the picture is divided into small squares with height and width 10 pixels each. If 50% of the square box is filled up with black pixels then the total square is filled with black pixels. In this way graphically smooth image is found. Then, the height of each of components in the graphically smooth image is computed. Next a rectangular template is created with a specified width and height as maximum portable height. Depending on the height and the position information, these smoothed blocks are joined to get the individual lines. Next the lines are extracted with the help of upper and lower boundaries. Then these are placed one after another in a link list, i.e. the nodes of the link list are the lines. Thus an unconstrained handwritten script is line segmented.

Karin[3] presented a method for identification of Text on colored book and journal covers. To reduce the amount of small variations in color, a clustering algorithm is applied in a preprocessing step. Two methods have been developed for extracting text hypotheses. One is based on a top-down analysis using successive splitting of image regions alternately in horizontal and vertical direction. Regions obtained under this top-down procedure are always of rectangular shapes and regions containing text include at least two colors. The other is a bottom-up region analysis detects homogeneous regions using growing algorithm. Beginning with the start region pixels within a 3x3 neighborhood are iteratively merged if they belong to the same cluster. The results of both methods are combined to robustly distinguish between text and non-text elements. Text elements are binarized using automatically extracted information about text colour. The binarized text regions can be used as input for a conventional OCR module. The proposed method is not restricted to cover pages, but can be applied to the extraction of text from other types of colour images as well.

2.2 Edge-Based Technique

Edge-based text extraction algorithm is a general-purpose method for text extraction. It quickly and effectively localizes and extracts the text from both document and images. Edges

are considered as a very important portion of the perceptual information content in a document image, which represents the significant intensity variations, discontinuities in depth, surface orientation, change in material properties etc. Vertical edges are detected by using smooth filter and it is connected into text clusters for the purpose of text extraction in edge-based technique.

Yingzi Du[4] propose an edge-based technique consists of four modules: multistage pulse code modulation(MPCM),text region detection (TRD) module, text box finding (TBF) module and optical character recognition (OCR).In the first module ,MPCM ,is used to locate potential text region in colour image. It convert image to coded image. In coded image each pixel encoded by a priority code ranging from 7 down to 0 in accordance with its priority and further produces a binary threshold image. The TRD module uses spatial filters to remove noisy regions and it also eliminate regions that are unlikely to contain text. Five filtering steps are included in this module: thresholding, elimination of isolated blocks, elimination of long vertical blocks, elimination of diagonally connected blocks and elimination of weakly connected vertical blocks. TBF module uses merge text regions and produces boxes that are likely to contain text. That is this module rectangularizes the text regions detected by TRD module and produce text boxes. The final OCR module eliminates the text boxes that produce no OCR output. The output of OCR module is a simple binary decision to determine whether a text box contains text.

Xiaoqing Liu [5] proposes a multi-scale edge-based text extraction algorithm which can quickly and effectively localize and extract text from both documents and images. The proposed method consists of three stages: candidate text region detection, text region localization and character extraction. The first stage aim to build a feature map by using three important properties of edges: edges strength, density and variance of orientations. The feature map is a gray-scale image with same size of the input image. Normally text embedded in an image appears in clusters that is, it's arranged compactly. Thus characteristics of clusters can be used to localize text regions. The purpose of character extraction stage is to extract accurate binary characters from the localize text regions so that we can use existing OCR directly for recognition.

2.3 Texture-Based Technique

Texture-based methods use the observation that texts in images have distinct textural properties that distinguish them from the background, to decide whether a pixel or block of pixels belong to text or not. Text feature extraction lies essentially on image pre-processing techniques, which is usually performed by linearly transforming or filtering the textured image followed by some energy measure or non-linear operator.

Wenge Mao [6] used wavelet transform and local energy variation analysis to discriminate between text-like regions, boundary and interior of other objects, and backgrounds. The pixels within text-like objects and near the boundary of other objects have large local energy variations, but the pixels within the background and the interior of non text-like objects have relatively smaller local energy variations. It can effectively detect text regions within images whether they are aligned horizontally and vertically. Furthermore, the method can simultaneously detect characters of various font sizes in a single image. In the first step, to characterize the local energy variations (LEV) of pixels in the successive scale levels of image, wavelet transform of image is done. Wavelet transform is more powerful to do this than conventional differential filters. In this paper it is done on the basis of Harr wavelet. In each scale level, the corresponding local energy variations are computed and then are thresholded. The threshold value is set at a certain percentage of the largest local energy variation in an image. The ratio of the threshold value to the largest local energy variation was chosen to be 0.45. The result of thresholding an LEV analyzed image in each scale level is a binary map image, in which pixels with value 1 correspond to large local energy variations and pixels with value 0 denote low local energy variations. The resulting binary map image in each scale level is subsequently analyzed by connected component analysis (CCA) technique to label different objects and background. Text regions are located from other connected object regions by the predefined geometric filtering. Geometric filtering follows the CCA process. In the final step, all text regions in the consecutive scale levels are fused into the original image and text regions are detected.

Bassem Bouaziz [7] method has four steps: Sweeping the image, detecting segments, storing information about segments and detecting regularity, based on local application of Hough transform combined with use of transformation matrix. It is an improved algorithm for feature extraction that mentioned in [7]. Let S a set of collinear pixels forming a line segment within an image. Then the extremities of a given segment can be identified by sweeping sequentially the image from top to bottom and from left to right. When a line segment is detected, it is stored and removed from the image. Then sequential search continues until the whole image is swept. When a line segment extremity is reached, sweeping can be done in all directions to find direction where most connected pixels exists. In order to improve performances and avoid call of trigonometric functions, two transformation matrixes can be computed in the initialization step. So each element can be represented as a pixel coordinates in image space by using maximal length of line segment that detected in a direction between 0 and 180°, and by the coordinates of line segment extremity identified when sweeping the image. The obtained matrix represents neighbourhood's information of a detected

extremity concerning connected pixels. This consideration of neighbourhood will help to detect imperfect segment as the case of an edge image. The last step of algorithm consists of removing segment's pixels that are having length exceeding a threshold value, which represents the minimal length of segment that should detect. Regularity can be detected if distance between parallel segments is similar for a specified value. This texture based text extraction can be used to build a robust car license plate localization system.

2.4 Morphological-Based Technique

Mathematical morphology is a topological and geometrical based approach for image analysis. It provides powerful tools for extracting geometrical structures and representing shapes in many applications.

Christian Wolf [8] presented morphological post processing to detect the text. This paper describes the intermediate steps detection, tracking, image enhancement and binarization. A phase of mathematical morphology follows the binarization step for several reasons: To reduce the noise, to correct classification errors using information from the neighborhood of each pixel and to connect characters in order to form complete words. The morphological operations consist of the following steps: Close, Suppression of unwanted bridges between components, Conditional dilation followed by a conditional erosion and Horizontal opening. The effect of this morphological step is the connection of the all connected components which are horizontally aligned and whose heights are similar. After the morphological post processing, geometrical constraints are imposed on the rectangles in order to further decrease the number of false alarms. The goal of the tracking is the association of detected text rectangles in successive frames to create appearances of text. Before passing the images to the OCR software contents of images are enhanced and also increased their resolution, which does not add any additional information to it.

Jui-Chen Wu [9] presents a morphology-based text line extraction algorithm for extracting text regions from cluttered images. This method defines a set of morphological operations for extracting important contrast regions as possible text line candidates. The contrast feature is robust to lighting changes and invariant against different image transformations like image scaling, translation, and detects skewed text lines. A moment-based method is used for estimating their orientations. According to the orientation, an x -projection technique can be applied to extract various text geometries from the text-analogue segments for text verification. However, due to noise, a text line region is often fragmented to different pieces of segments. Therefore, after the projection, a novel recovery algorithm is then proposed for recovering a complete text line from its pieces of segments.

The proposed method improves the state-of-the-art work in terms of effectiveness and robustness for text line detection.

3. COMPARISON AND PERFORMANCE EVALUATIONS

The performance evaluation of information retrieval can be done using precision and recall rate. The precision rate measures the percentage of correctly detected text boxes with in each image as opposed to detected boxes, where as percentage of correctly detected text boxes that actually contain in text are measured by recall rate.

$$\text{Precision rate} = \frac{\text{Number of correctly detected text boxes}}{\text{Number of detected text boxes}}$$

$$\text{Recall rate} = \frac{\text{Number of correctly detected text boxes}}{\text{Number of text boxes}}$$

Performance evaluation and comparison of different text extraction techniques discussed in this paper are listed in Table 1.

Table 1: Comparison

Technique	Method	Properties	Recall rate (%)	Precision rate (%)
Region-based	Bottom-Up Approach of Line Segmentation[2]	Handwritten Text	92	
	Clustering, Top-Down successive splitting, bottom-up region growing[3]	Colored book and journal covers	83.3	83.5
Edge-based	Automated system for text extraction [4]	Scene text and superimposed text within video images	92	85
	Multiscale Strategy, Clustering [5]	Complex printed document images and scene text	96.6	91.8
Texture-based	Multiscale texture based method using local energy	Hybrid Chinese/English text detection in images and video	93.5	1.8

	analysis [6] Hough Transform technique combined with an extremity segments neighbourhood analysis [7]	frames Detect text in video images	94	2.48
Morphological-based	A measure of accumulated gradients and morphological post processing [8]	Artificial text in images and video	93.5	85.4
	Novel set of morphological operations and an x-projection techniques [9]	Cluttered images	96.3	99.4

4. CONCLUSION

Automatic text detection and extraction from an image is an important research branch of content-based information retrieval and text based image indexing. Some of the applications fields of text extraction are mobile robot navigation, vehicle license detection and recognition, object identification, document retrieving, page segmentation etc. Based on the information collected from various techniques it is found that morphological and edge based techniques can quickly and effectively localize and extract text from images. The remaining methods, region and texture based techniques, show the poor performance compared to morphological and edge based technique.

REFERENCES

1. C.P.Sumathi, T.Santhanam, N.Priya, **Techniques and challenges of automatic text extraction in complex images: A survey**, vol 35 No.2 ,Jan 2012.
2. Debapratim Sarkar, Raghunath Ghosh ,**A bottom-up approach of line segmentation from handwritten text** ,2009
3. Karin Sobottka, Horst Bunke and Heino Kronenberg, **Identification of text on colored book and journal covers**, Document Analysis and Recognition, 20-22, 1999.
4. Yingzi Du Chein-I Change and Paul D.Thouin, **Automated system for text detection in individual video images** , Journal of Electronic Imaging ,pp 410-422, July 2003.
5. Xiaoqing Liu and Jagath Samarabandu, **Multiscale edge-based text extraction from complex images**, Multimedia and Expo, 2006 IEEE International Conference, 2006.
6. W.Mao, F. Chung, K. Lanm, and W.Siu, **Hybrid chinese/english text detection in images and videos frames** , Proc.of International Conference on Pattern Recognition,2002,vol.3, pp. 1015-1018
7. Bassem Bouaziz, Walid Mahdi, Mohsen Ardabilain, Abdelmajid Ben Hamadou, **A New Approach For Texture Features Extraction: Application For Text Localization In Video Images** IEEE,ICME 2006.
8. C. Wolf, J. –M. Jolion F. Chassaing , **Text localization, enhancement and binarization in multimedia documents** , Patterns Recognition ,2002.proceedings. 16th International Conference on ,vol 2,11-15, pp.1037-1040, Aug. 2002.
9. Jui-Chen Wu · Jun-Wei Hsieh Yung-Sheng Chen, **Morphology-based text line extraction** , Machine Vision and Application 19:195-207 DOI,2008.