A SURVEY ON COMPARISION AND PERFORMANCE ANALYSIS OF TEXT EXTRACTION TECHNIQUES

Prachi R. Dussawar[1], Prof. Parul Bhanarkar Jha[2]

[1]Wireless Communication & Computing, TGPCET/RTM Nagpur University, (India)

[2]Head of Department, Information Technology Dept., and TGPCET/ RTM Nagpur University, (India)

ABSTRACT

Text data present in images and video contain useful information for automatic annotation, indexing, and structuring of images. Extraction of this information involves detection, localization, tracking, extraction, enhancement, and recognition of the text from a given image. However, variations of text due to differences in size, style, orientation, and alignment, as well as low image contrast and complex background make the problem of automatic text extraction extremely challenging. While comprehensive surveys of related problems such as face detection, document analysis, and image & video indexing can be found, the problem of text information extraction is not well surveyed. A large number of techniques have been proposed to address this problem, and the purpose of this paper is to classify and review these algorithms, discuss benchmark data and performance evaluation, and to point out promising directions for future research.

Keywords: OCR, Text Information Extraction, Text Detection, Text Localization, Text Tracking, Text Enhancement.

I INTRODUCTION

Recent studies in the field of computer vision and pattern recognition show a great amount of interest in content retrieval from images and videos. This content can be in the form of objects, color, texture, shape as well as the relationships between them. The semantic information provided by an image can be useful for content based image retrieval, as well as for indexing and classification purposes [4, 10]. As stated by Jung, Kim and Jain in [4], text data is particularly interesting, because text can be used to easily and clearly describe the contents of an image. Since the text data can be embedded in an image or video in different font styles, sizes, orientations, colors, and against a complex background, the problem of extracting the candidate text region becomes a challenging one [4]. Also, current Optical Character Recognition (OCR) techniques can only handle text against a plain monochrome background and cannot extract text from a complex or textured background [7].

Different approaches for the extraction of text regions from images have been proposed based on basic properties of text. As stated in [7], text has some common distinctive characteristics in terms of frequency and orientation information, and also spatial cohesion. Spatial cohesion refers to the fact that text characters of the same string appear close to each other and are of similar height, orientation and spacing [7]. Two of the main

methods commonly used to determine spatial cohesion are based on edge [1, 2] and connected component [3] features of text characters.

The fact that an image can be divided into categories depending on whether or not it contains any text data can also be used to classify candidate text regions. Thus other methods for text region detection, as described in more detail in the following section, utilize classification techniques such as support vector machines [9, 11], k-means clustering [7] and neural network based classifiers [10]. The algorithm proposed in [8] uses the focus of attention mechanism from visual perception to detect text regions.

II RELATED WORK

Various methods have been proposed in the past for detection and localization of text in images and videos. These approaches take into consideration different properties related to text in an image such as color, intensity, connected-components, edges etc. These properties are used to distinguish text regions from their background and/or other regions within the image. The algorithm proposed by Wang and Kangas in [5] is based on color clustering. The input image is first pre-processed to remove any noise if present. Then the image is grouped into different color layers and a gray component. This approach utilizes the fact that usually the color data in text characters is different from the color data in the background. The potential text regions are localized using connected component based heuristics from these layers. Also an aligning and merging analysis (AMA) method is used in which each row and column value is analyzed [5]. The experiments conducted show that the algorithm is robust in locating mostly Chinese and English characters in images; some false alarms occurred due to uneven lighting or reflection conditions in the test images.

The text detection algorithm in [6] is also based on color continuity. In addition it also uses multi-resolution wavelet transforms and combines low as well as high level image features for text region extraction. The text finder algorithm proposed in [7] is based on the frequency, orientation and spacing of text within an image. Texture based segmentation is used to distinguish text from its background. Further a bottom-up 'chip generation' process is carried out which uses the spatial cohesion property of text characters. The chips are collections of pixels in the image consisting of potential text strokes and edges. The results show that the algorithm is robust in most cases, except for very small text characters that are not properly detected. Also in the case of low contrast in the image, misclassifications occur in the texture segmentation.

A focus of attention based system for text region localization has been proposed by Liu and Samarabandu in [8]. The intensity profiles and spatial variance is used to detect text regions in images. A Gaussian pyramid is created with the original image at different resolutions or scales. The text regions are detected in the highest resolution image and then in each successive lower resolution image in the pyramid.

The approach used in [9, 11] utilizes a support vector machine (SVM) classifier to segment text from non-text in an image or video frame. Initially text is detected in multi-scale images using edge based techniques, morphological operations and projection profiles of the image [11]. These detected text regions are then verified

using wavelet features and SVM. The algorithm is robust with respect to variance in color and size of font as well as language.

III TEXT EXTRACTION SCHEMES

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.

3.1 Region-Based Technique

Region-based methods use the properties of the color 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 [12] 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[13] 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.

3.2 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[14] 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 [2] 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.

3.2.1 Algorithm for edge based text region extraction

The basic steps of the edge-based text extraction algorithm are given below:

1. Create a Gaussian pyramid by convolving the input image with a Gaussian kernel and successively down-sample each direction by half. (Levels: 4)
2. Create directional kernels to detect edges at 0, 45, 90 and 135 orientations.
3. Convolve each image in the Gaussian pyramid with each orientation filter.
4. Combine the results of step 3 to create the Feature Map.
5. Dilate the resultant image using a sufficiently large structuring element (7x7 [1]) to cluster candidate text regions together.

6. Create final output image with text in white pixels against a plain black background.

3.3 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 [16] 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 threshold. 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 analysed 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 analysed 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 [17] 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 [17]. 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.

3.4 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 [18] 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 [19] 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.

IV COMPARISION AND PERFORMANCE ANALISYS

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, whereas percentage of correctly detected text boxes that actually contain in text are measured by recall rate.

Precision rate= Number of correctly detected text boxes / Number of detected text boxes

Recall rate=Number of correctly detected text boxes / Number of text boxes

Performance evaluation and comparison of different text extraction techniques discussed in this paper are as follows.

4.1 Region Based Technique

Methods	Properties	Recall rate (%)	Precision rate (%)
Bottom-Up Approach of Line Segmentation[12]	Handwritten Text	87	
Clustering, Top-Down successive splitting ,bottom-up region growing [13]	Colored book and Journal covers	79.9	79.5

4.2 Edge Based Technique

Methods	Properties	Recall rate (%)	Precision rate (%)
Automated system for text extraction[14]	Scene text and superimposed text within images	87	81
Multi-scale Strategy ,Clustering[15]	Complex printed Document images	92.6	86.8

4.3 Texture Based Technique

Methods	Properties	Recall rate (%)	Precision rate (%)
Multi-scale Texture based Method using local energy analysis [16]	Hybrid Chinese/ English text detection in images and video frames	92.8	1.9
Hough Transform technique combined with an extremity segments neighbor-hood analysis [7]	Detect text in video images	95	2.56

4.4 Morphological Based Technique

Methods	Properties	Recall rate (%)	Precision rate (%)
A measure Of accumulated gradients And morphologic post processing [18]	Artificial text in images and video	92.5	85.4
Novel set of Morphologic operations and an x-projection Techniques [19]	Cluttered images	96.3	99.4

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

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

[3] Julinda Gllavata, Ralph Ewerth and Bernd Freisleben, A Robust algorithm for Text detection in images, Proceedings of the 3rd international symposium on Image and Signal Processing and Analysis, 2003.

[4] Keechul Jung, Kwang In Kim and Anil K. Jain, Text information extraction in images and video: a survey, The journal of the Pattern Recognition society, 2004.

[5] Kongqiao Wang and Jari A. Kangas, Character location in scene images from digital camera, The journal of the Pattern Recognition society, March 2003.

[6] K.C. Kim, H.R. Byun, Y.J. Song, Y.W. Choi, S.Y. Chi, K.K. Kim and Y.K Chung, Scene Text Extraction in Natural Scene Images using Hierarchical Feature Combining and verification, Proceedings of the 17th International Conference on Pattern Recognition (ICPR ’04), IEEE.

[7] 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.

[8] Xiaoqing Liu and Jagath Samarabandu, A Simple and Fast Text Localization Algorithm for Indoor Mobile Robot Navigation, Proceedings of SPIE-IS&T Electronic Imaging, SPIE Vol. 5672, 2005.

[9] 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.

[10] 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.

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

[12] Debapratim Sarkar, Raghunath Ghosh ,A bottom-up approach of line segmentation from handwritten text ,2009.

[13] Karin Sobottka, Horst Bunke and Heino Kronenberg, Identification of text on colored book and journal covers, Document Analysis and Recognition, 20-22, 1999.

[14] Yingzi Du Chein-I Change and Paul D.Thouin, Automated system for text detection in individual video images, Journal of Electronic Imaging ,pp410-422, July 2003.

[15] 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

[16] 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.

[17] 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.

[18] 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.

[19] Jui-Chen Wu · Jun-Wei Hsieh Yung-Sheng Chen, Morphology-based text line extraction, Machine