Available Online at www.ijcsmc.com

International Journal of Computer Science and Mobile Computing

A Monthly Journal of Computer Science and Information Technology

ISSN 2320–088X

IJCSMC, Vol. 4, Issue. 12, December 2015, pg.156 – 161

TEXT EXTRACTION FROM SKEWED IMAGES

KOMAL[1], PAWAN SINGH[2], SUBHA JAIN[3]

Dept. of CSE and IT, Kanpur Institute of Technology, GBTU, INDIA

[1] k.komal208@gmail.com; [2] pawan067.miet@gmail.com; [3] sj@kit.ac.in

ABSTRACT: The extraction of text in an image is a classical problem in the computer vision. Extraction involves detection, localization, tracking, extraction, enhancement and recognition of the text from the given image.

However variation of text due to difference in size, style, orientation, alignment, low image contrast and complex background make the problem of automatic text extraction extremely challenging. Text extraction requires binarization which leads to loss of significant information contained in gray scale images. The images may contain noise and have complex structure which makes the extraction more difficult. This paper proposes an algorithm which is insensitive to noise, skew and text orientation. It is free from artifacts that are usually introduced by thresholding using morphological operators. Examples are presented to illustrate the performance of proposed method. The text extraction system has been attempted over a corpus of three kinds of images and promising precision has been obtained.

Keywords: Mathematical Morphology, Morphological Operators, Edge Detection, Localization, Connected Component

1. INTRODUCTION

Text extraction from images and video sequences finds many useful applications in document processing [1], detection of analysis of technical papers with tables, maps, charts. Identification of parts in industrial automation [3], and content- based image/video retrieval from image/video databases [4], [5]. Educational and training video and TV programs such as news contain mixed text-picture-graphics regions.

Region classification is helpful in object-based compression, manipulation and accessibility. Also, text regions may carry useful information about the visual content. However due to the variety of fonts, sizes, styles, orientations, alignment effects of uncontrolled illuminations, reflections, shadows, the distortion due to perspective projection as well as the complexity of image background, automatic localizing and extracting text is a challenging problem. Characters in a text are of different

shapes and structures. Text extraction may employ binarization [7], [9]–[11] or directly process the original image [8], [12], [13]. In [5], a survey of existing techniques for page layout an analysis is presented. 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. Morphological feature extraction techniques have been efficiently applied to character recognition and document analysis, especially if dedicated hardware is used.
In this paper, we propose an algorithm for text extraction based on morphological operations.
The paper is organized as follows. In Section II, the proposed morphological text extraction technique is described. Examples and comparison with existing text extraction algorithms are presented in Section III. Conclusion is given in Section IV.

Why is skew and orientation determination important?

There are a variety of circumstances in which it is useful to determine the text skew and orientation:
Improves text recognition. Many systems will fail if presented with text oriented sideways or upside-down. Performance of recognition systems also degrades if the skew is more than a few degrees.
_ *Simplifies interpretation of page layout.* It is easier to identify textlines and text columns if the image skew is known or the image is deskewed.
_ *Improves baseline determination.* The textline baselines can be found more robustly if the skew angle is accurately known[7].
_ *Improves visual appearance.* Images can be displayed or printed after rotation to remove skew. Multiple page document images can also be oriented consistently, regardless of scan orientation or the design of duplex scanners.

2. METHODOLOGY

The method has considered the fact that edges are reliable features of text regardless of color or intensity, layout, orientation etc. The edge detection operation is performed using the basic operators of mathematical morphology. Using the edges the algorithm has tried to find out text candidate connected components. These components have been labelled to identify different components of the image. Once the components have been identified, the variance is found for each connected component considering the gray levels of those components.

Then the text is extracted by selecting those connected components whose variance is less than some threshold value. The complete process of text extraction is given in the form of chart in figure 1

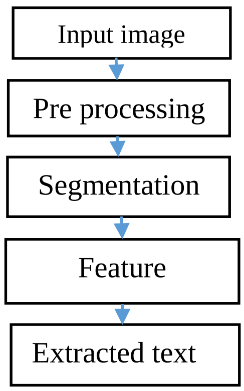

Figure 1 complete text extraction process

Text candidate region formation

From the threshold image the text candidate regions are obtained as follows. In text candidate region formation close operation is applied to connect all the edges.
Algorithm: Region Formation (e)
Input: threshold image
The basic steps of the edge-based text extraction algorithm are given below, and diagrammed in Figure 1. The details are explained in the following sections.

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, 10, 15 and 20 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 (. Create final output image with text in white pixels against a plain black 7x7 [1]) to cluster candidate text regions together.

Detection

This section corresponds to Steps 1 to 4 . Given an input image, the region with a possibility of text in the image is detected. A Gaussian pyramid is created by successively filtering the input image with a Gaussian kernel of size 3x3 and down-sampling the image in each direction by half. Down sampling refers to the process

In this process, the connected-component based approach is used to make possible text regions stand out as compared to non-text regions.

The algorithm for computing the edge image E, as proposed in [3] is as follows:

1. Assign left, upper, upperRight to 0.

2. For all the pixels in the gray image G(x,y) do

a. left = (G(x,y) – G(x-1,y))

b. upper = (G(x,y) – G(x,y-1))

c. upperRight = (G(x,y)-G(x+1,y-1))

d. E(x,y) = max(left, upper, upperRight)

W (x,y) = max(L,U,UR)

Performance Analysis
Metrics used to evaluate the performance of the system are Precision, Recall and F-Score. Precision and Recall rates have been computed based on the number of Correctly Detected Characters (CDC) in an image, in order to evaluate the efficiency and robustness of the algorithm. The metrics are as follows:
Definition 1: False Positives (FP) / False alarms are those regions in the image which are actually not characters of a text, but have been detected by the algorithm as text.
Definition 2: False Negatives (FN)/ Misses are those regions in the image which are actually text characters, but have not been detected by the algorithm.
Definition 3: Precision rate (P) is defined as the ratio of correctly detected characters to the sum of correctly detected characters plus false positives as represented in equation below.

$$P= \frac{CDC(\text{true positive})}{CDC+fp}*100$$

Definition 4: Recall rate (R) is defined as the ratio of the correctly detected characters to sum of correctly detected characters plus false negatives as represented in equation below.

3. Results and discussion:
These results are based on the horizontal and vertical projection analysis. The algorithms are tested and all the result have been texted through the vertical and horizontal view. In figure a simply icon are provided for selecting an options (scan and import). In scan an image all the letter match with its threshold value. Another match with letter with connected component methods used by edge detection algorithm by importing option easily tag the picture by gallery, photos and take picture.

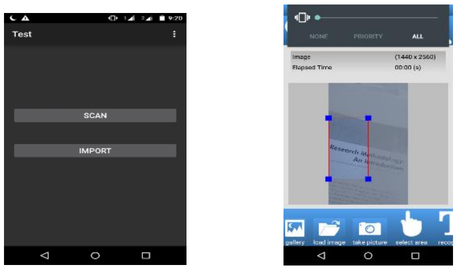

Figure 2.imges of text extraction using algorithm

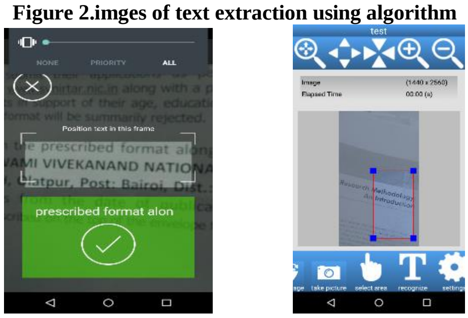

Figure 3.text extraction result images using feature extraction technique

Edge based results:

Image type	Distance from camera	PR%	RR%
indoor	15 cm	90%	80%
	18.5cm	75%	82.30%
	25.8cm	62.73%	72.3%
outdoor	16.5 cm	67%	80%
	15.5 cm	65%	81%
	26.5 cm	56.5%	76%

Orientation detection accuracies for book data set

Data set	Document resolution(in degree)	Connected component level accuracy (%)	Overall connected component accuracy (%)	Overall page level accuracy (%)
Printed page	0	100	89.5	100
	10	95		
	15	85		

4. Comparison with other text extraction techniques

To give an average estimate of the performance of the text extraction the results have been compared against two existing algorithms [14] and [15]. Both the methods have used the aspect ratio to identify the text and non text regions within an image. The first method has used the complicated procedure of finding inner, outer and inner-outer corners. The second procedure has identified edge at different orientation i.e. 0, 45, 90, and 135 degrees and grouping these strokes at different heights, text is extracted. This increases the complexity of algorithm to identify the edges at different orientations. The new Connected Component Variance (CCV) approach has solved the above problems. In order to group the isolated characters to a meaning full word, a dilation operation with varying structuring element is used. The algorithm identifies number of components and calculates the variance of each component if the variance of each component is high then it is believed to a kind of symbol rather than a text. This algorithm is in sensitive to skew and text orientation, the output of the text extraction algorithm is fed to an OCR system to recognize the contained information. The main objective of the text extraction algorithm is to reduce the number of false text candidate that may be fed to the OCR.

Further investigation on the threshold value to select the correct text candidate is being performed. Different images with different lighting and contrast is used for text extraction. These images are tested using the two algorithms Threshold with Average Gray values (TAG) and operators (TMO). To evaluate the performance of

TAG and TMO 25 text images with different font size, perspective, alignments any number of characters in a text string under different lighting conditions is considered. It has been observed that in many case the images gives good text extraction when TMO is used, when the image has uniform background. If various foregrounds and backgrounds are presented in the image, it has been observed that TMO is in sensitive to noise and introduce minimum noise on the removal of non text information. The TAG algorithm introduces noises while removing non text characters but has an advantage of extracting those text characters whose gray levels are close to the back ground shown in the Figure 4.

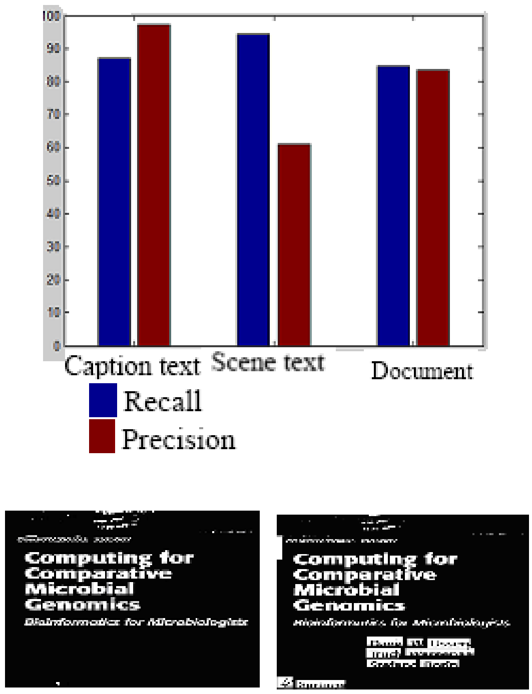

Figure 4: Text extraction using TMO and TAG algorithms

Conclusion and future scope

This paper proposes a new text extraction algorithm from a text/graphics mixed document images. This algorithm is in sensitive to skew and text orientation the output of the text extraction algorithm is fed to an OCR system to recognize the contained information. The results obtained on varied set of images are compared with respect to precision and recall rates. Promising results have been obtained on a number of images in which almost all text lines can be retrieved from the graphics and figure regions. The images have been tested using various threshold techniques. It has been observed that the threshold depends on a various parameters like the illumination condition reflections and the scan point spread function. This approach has used morphological clearing of images which would help to reduce the number of false positive obtained. This cleaning of the image could result in a higher precision rate.

Further investigation on the threshold value to select the correct text candidate is being performed. The images have been tested using various threshold techniques. It has been found that TAG gives efficient extraction comparative to TMO when the images are taken in poor illumination. However the TAG gives noisy distorted binary images comparative to TMO. This approach has used morphological clearing of images which would help to reduce the number of false positive obtained. In corporately the OCR algorithm with proposed morphological text extraction method yields a useful system for text analysis in images.

REFERENCES

[1] K. Jain and B. Yu, "Document representation and its application to page decomposition," IEEE Trans. Pattern Anal. Machine Intell., vol. 20, pp. 294–308, Mar. 1998.

[2] K. Jain, and B. Yu, "Automatic Text Locationin Images and Video Frames", Pattern Recognition, 31 (12) (1998) 2055-2076.

[3] K. Jain, and Y. Zhong, "Page Segmentation using Texture Analysis", Pattern Recognition29 (5) (1996) 743-770.

[4] Chitrakala Gopalan and D. Manjula, "Contourlet Based Approach for TextIdentification and Extraction from Heterogeneous Textual Images", InternationalJournal of Computer Science and Engineering 2(4) (2008) pp.202-211.

[5] D. Crandall, S. Antani, and R. Kasturi"Robust Detection of Stylized Text Events inDigital Video", Proceedings of InternationaConference on Document Analysis and Recognition, 2001, pp. 865-869.

[6] J. Serra, Image Analysis and Mathematical Morphology. New York:Academic, 1982.

[7].Bouman K.L., G. Abdollahian, M. Boutin, and E.J. Delp, "A Low Complexity Sign Detection and Text Localization Method for Mobile Applications", *IEEE Transactions on Multimedia*, Vol. 13, No. 5, pp. 922-934, 2011.

[8]. Jafri S.A.R., M. Boutin, and E.J. Delp, "Automatic text area segmentation in natural images", *Proceedings of International Conference on Image Processing*, pp.3196- 3199, 2008

[9].Zhang J. and R. Kasturi, "Text Detection Using Edge Gradient and Graph Spectrum",*Proceedings of International Conference on Pattern Recognition*, 2010

[10].Shivakumara P., W. Huang, T. Phan and C.L. Tan, "Accurate Video Text Detection Through Classification of Low and High Contrast Images", *Pattern Recognition*, Vol.43, Issue 6, pp. 2165-2185, 2010.

[11]. Mancas-Thilou C., B. Gosselin, "Spatial and Color Spaces Combination for Natural Scene Text Extraction", *IEEE international Conference on Image Processing*, pp. 985-988, 2006.

[12]. Shivakumara P., T. Q. Phan, and C. L. Tan, "A Laplacian approach to multi-oriented text detection in video", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 33, No. 2, 2011.

[13]. Qian X. and G. Liu, "Text Detection, Localization, and Segmentation in Compressed Videos", *Proceedings of IEEE international conference on acoustics, speech and signal processing*, Vol. 2, pp. 385-388, 2006

[14].Qian X., G. liu, H. Wang, R. Su, "Text Detection, Localization, and Tracking in Compressed Video", *Signal Processing: Image Communication*, Vol. 22, pp 752-768, 2007.

[15].Anthimopoulos M., N. Vlissidis, and B. Gatos, "A Pixel-Based Evaluation Method for Text Detection in Color Images", *Proceedings of International Conference on Pattern Recognition*, pp. 3264-3267, 2010

[16]. Kasturi R., D. Goldgof, P. Soundararajan, V. Manohar, J. Garofolo, R. Bowers, M. Boonstra, V. Korzhova, and J. Zhang, "Framework for Performance Evaluation of Face, Text, and Vehicle Detection and Tracking in Video: Data, Metrics, and Protocol", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol 31(2), pp. 319-336, 2008.