IJIREEICE

ISSN (Online) 2321 – 2004
ISSN (Print) 2321 – 5526

International Journal of Innovative Research in
Electrical, Electronics, Instrumentation and Control Engineering

ISO 3297:2007 Certified

Vol. 5, Issue 5, May 2017

Text Extractions from Natural Scenes by Various Methods

M.Priyanka[1], Sheela S[2], Manjunath P[3]

M.Tech Student, Dept of ECE, Jawaharlal Nehru National College of Engineering, Shimoga, Karnataka, India[1]

Associate Professor, Dept of ECE, Jawaharlal Nehru National College of Engineering, Shimoga, Karnataka, India[2]

Head of the Department, Dept of ECE, Jawaharlal Nehru National College of Engineering, Shimoga, Karnataka, India[3]

Abstract: Text Extraction plays a major role in finding vital and valuable information. Text extraction involvesdetection, localization, tracking, binarization, extraction, enhancement and recognition of the text fromthe given image. These text characters are difficult to be detected and recognized due to their deviation ofsize, font, style, orientation, alignment, contrast, complex coloured, textured background. Due to rapidgrowth of available multimedia documents and growing requirement for information, identification,indexing and retrieval, many researches have been done on text extraction in images. Several techniqueshave been developed for extracting the text from an image. The proposed methods were based onmorphological operators, wavelet transform, artificial neural network, skeletonization operation, edgedetection algorithm, histogram technique etc. All these techniques have their benefits and restrictions. This article discusses various schemes proposed earlier for extracting the text from an image. This paperalso provides the performance comparison of several existing methods proposed by researchers inextracting the text from an image.

Keywords: Text Extraction, Document Text Images, Caption Text Images, Scene Text, Heterogeneous Images.

I. INTRODUCTION

Today, most of the useful information is available into the text which is present into the natural images. For eg., Name of the brand embedded into clothes, text written on the nameplates, signboards etc. Extracting the text from these images is still a difficult task. There should be some mechanism to extract the text from natural images. Images are the most convenient means of conveying ortransmitting information as they quickly convey information about positions, sizes and inter-relationships between the objects. Recent studies in the field of research on the content retrieval from images and videos identified a wide variety of applications that require automated systems for text extraction.

One such recently developed application is the mobile banking application provided by the banking institutions that facilitates the customers to carry out the transactions even on passing the image of the cheque to the server. All other such applications include tourist guide which facilitate the tourists to understand the display boards though they are unfamiliar with the local languageof that place and Image text translation systems to help the visually impaired people and also tourists. Every such application relies on a Textual Information Extraction (TIE) system which can efficiently detect, localize and extract the text information present in the natural images.

II. OPERATION

Input image is any image that contains text information. The first step detects the regionswhere the text is present. Text localization refers to locate and enhance text areas. Character extraction refers to eliminate non-text regions from image, such that output image contains only text. This image is sent to OCR system to obtain the characters that are extracted.

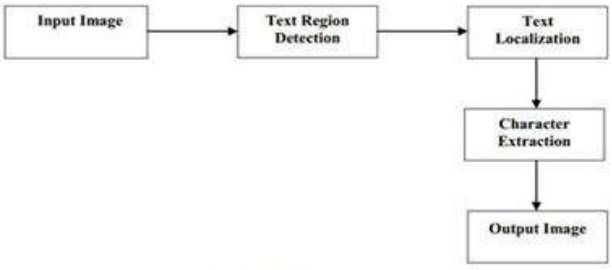

Fig.1: Procedure of Text Extraction from an image.

IJIREEICE

ISSN (Online) 2321 – 2004
ISSN (Print) 2321 – 5526

International Journal of Innovative Research in
Electrical, Electronics, Instrumentation and Control Engineering

ISO 3297:2007 Certified

Vol. 5, Issue 5, May 2017

Many methods for text detection in video have been proposed in the area of image processing and computer vision. These methods can be mainly classified as connected component based methods, texture based methods and edge based methods.

Image binarization successfully processes natural scene images, which is having shadows, non-uniform illumination, low contrast and largesignal dependent noise. Connected component analysis is used to define the lastly binary images that mainly consist of text regions. The proposed methodology results in increased success rates for commercial OCR engines. Experimental results based on public database of natural scene images prove the efficiency of the proposed approach.

III. ALGORITHM

In this paper, edge-detector based method is proposed andcompared with conventional Gaussian pyramidmethod.Comparison is done in terms of precision rate and recall rate.

A. Algorithm that uses Gaussian pyramid

1. Create a Gaussian pyramid by convolving the input image with a Gaussian kerneland 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) to cluster candidate text regions together.
6. Create final output image with text in white pixels against a plain black background.

B. Algorithm that uses Prewitt edge-detector

1. Convert the image into monochrome image bythresholding.
2. Filter the image for removing noise. Use Gaussian low-pass filter.
3. Apply Prewitt edge-detector to the filtered image.
4. Apply proper morphological operations, i.e. dilation to make clusters of text regions.
5. Multiply the resultant image with input black and white image to get text in contrast with plain background.

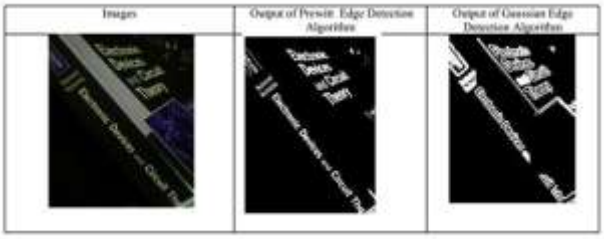

Fig 2: Images of Text Detection Methods

Table 1: Precision rate and recall rate of various natural scene images

	Precision Rate		Recall Rate	
Image	Prewitt (%)	Gaussian (%)	Prewitt (%)	Gaussian (%)
Cover.jpg	91.6	90	84.61	69.23
Exhibit.jpg	75	85.79	100	66.66
Station.jpg	71.42	83.3	100	100
uva.jpg	**69.23**	**0**	**90**	**0**
Book.jpg	**100**	**100**	**100**	**53.84**
Elly.jpg	93.1	95.5	90	76.66
Dept.jpg	28.5	0	15	0
Smoke.jpg	76.41	80	65	56.52
Raper.jpg	**89.47**	**66.66**	**54.83**	**6.66**

Many efforts have been done to address the problems of text area detection, text segmentation, and text recognition. Current text detection approaches can be classified into three categories. The firstcategory isconnected component-based method, which can locate textquickly but have difficulties when text is embedded incomplex background or touches other graphical objects.The second category is texture-based, which is hard to find accurate boundaries of text areas and usually yields many false alarms in “text-like” background texture areas. The third category is edge-based method. Generally, analysing the projection profiles of edge intensity maps can decompose text regions and can

IJIREEICE

ISSN (Online) 2321 – 2004
ISSN (Print) 2321 – 5526

International Journal of Innovative Research in
Electrical, Electronics, Instrumentation and Control Engineering

ISO 3297:2007 Certified

Vol. 5, Issue 5, May 2017

efficiently predict the textdata from a given video image clip. Text region usuallyhave a special texture because they consist of identical character components. These components contrast the background and have a periodic horizontal intensity variation due to the horizontal alignment of many characters. As a result, text regions can be segmented using texture feature. Document Image Segmentation is the act of partitioning a document image into separated regions. These regions should ideally correspond to the image entities such as text blocks and graphical images, which are present in the document image.

IV. APPROACH

A. Pre-processing:

Pre-processing steps are necessary to improve the performance and make the process efficient to the time. This includes gray-scaling and binarization of image and filtering to remove noise.

1. Gray-scaling: The given image is multicolor RGB image, in which text may not be separated from the background. In color image, each pixel is combination of R (Red), G (Green) and B (Blue) and values varying from 0 to 255. For gray-scaling, these values are added in a proportion of Red: 30%, Green: 59% and Blue: 11 % to get the gray scaled equivalent of that particular pixel.

2. Binarization: This converts gray-scale image into binary image i.e. containing only Black (0) and White (1) Pixels. Gray-scaling gives a threshold for binarization of image. To be specific, this is done by comparing each pixel value to a threshold value (that lies between black and white) and setting that pixel value to black or white as its consequence. This process noticeably separate (ordistinguish) text from imagebackground.

3. Filtering: Any image taken from camera contained noise such as blurred image, high frequency noise and white noise. To improve image quality and for further processing on image, Gaussian low pass filter isused. It has following properties:

1) Gaussian smoothing is very effective for removing Gaussian noise.
2) They are linear low pass filters.
3) Rotationally symmetric (perform the same in all directions).
4) The degree of smoothing is controlled by σ (larger σ or more intensive smoothing) In our case, it is important to remove white noise, while maintaining salient edges because text contains edges. This can be a contradictory task-white noise exists at all frequencies equally, while edges exist inthe high frequency range. (Sudden changes in spatial signals-text regions). Gaussian LPF is best-suited for this because it has a graceful and natural tail that becomes ever lower as the frequency increases. Thismeans that it will act as a low pass filter, but also allow in higher frequency components matching with how quickly its tail decays.

B. Edge-Detection

Edges are those places in an image that correspond to object boundaries. Edges are pixels where image brightness changes abruptly. Specifically in text data probably more edges are present than non-text areas. For example, Letters „E", „Z", „H", "A" etc. are having horizontal and/or vertical edges. If we detect these edges, there may be likelihood of other letters or words around (because words are usually grouped) Thus, the text region is detected. We chose Prewitt amongst several edge-detectors available like Sobel, canny and Roberts. Choice of Prewitt is quite empirical. Prewitt edge detector detects horizontal and vertical edges in an image and combines them to give resultant image.

C. Morphological Operations

After detecting text region(s), a cluster of it is created such that the all letters are covered. Morphological dilation is used for this purpose as dilation adds pixels to the boundaries of objects in an image there by thickening that object. Measure of thickness is defined by the type and size structuring element. Proper sized structuring element should be chosen such that least non-text area should be clustered within. Here, structuring element „disk" withsize 9 (a disk of radius 9) is used. To remove non-text objects significantly, morphological opening operation is used. Opening operation is erosion followed by dilation. It is performed to remove objects of specific size from image. This size is again determined by structuring element. After performing such operations, the resultant image holds clusters of text regions having pixel value 1 (white).

C. Character extraction

This step refers to identify the characters as they are in original image. This is done by multiplying resultant image with binary converted original image. In this operation, pixels having value 1 (i.e. text) are recovered as same in original image and pixels having value 0 are present as background. However, the final image may contain some non-text part,

IJIREEICE

ISSN (Online) 2321 – 2004
ISSN (Print) 2321 – 5526

International Journal of Innovative Research in
Electrical, Electronics, Instrumentation and Control Engineering

ISO 3297:2007 Certified

Vol. 5, Issue 5, May 2017

extent of which is measured by precision rate. Final result is the white text in black background or vice versa, dependent on the original image.

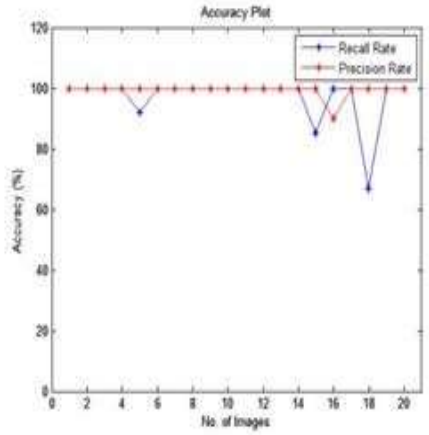

Fig 3: Performance analysis using line graph

Fig 4: Original text images and its detected text regions images.

Fig. 5: Sample results given by the existing method

V. CONCLUSION

In this work, I presented the design of scene text extraction which can be used for many applications as mentioned above. However the method fails when the edges of several characters are lumped together into a single large connected component that is eliminated by selection rule. The proposed algorithm is best for medium size text extraction. The results obtained by each algorithm on a varied set of images are compared with respect to precision and recall rates. The overall precision rate obtained by the combined approach algorithm (98.46 %) is higher than individual approach. Also, the overall recall rate obtained by the combined approach algorithm (97.83%) is higher than that obtained by individual approach. Refer Table hence proposed algorithm is best algorithm for text extraction from natural scene images.

IJIREEICE

ISSN (Online) 2321 – 2004
ISSN (Print) 2321 – 5526

International Journal of Innovative Research in
Electrical, Electronics, Instrumentation and Control Engineering

ISO 3297:2007 Certified

Vol. 5, Issue 5, May 2017

REFERENCES

[1] Y. Zhan, W. Wang, W. Gao (2006), "A Robust Split-And-Merge Text Segmentation Approach For Images", International Conference On Pattern Recognition,06(2):pp 1002-1005.
[2] Thai V. Hoang , S. Tabbone(2010),"Text Extraction From Graphical Document Images Using Sparse Representation"inProc. Das, pp 143–150. International Journal of Computer Science & Engineering Survey (IJCSES) Vol.3, No.4, August 201241
[3] AudithanR.M.Chandrasekaran (2009), "Document Text Extraction From Document Images Using Haar Discrete Wavelet Transform",European Journal Of Scientific Research, Vol.36 No.4 pp.502-512.
[4] Sachin, Grover,KushalArora,,Suman K. Mitra(2009),"Text Extraction From DocumentImages Using Edge Information",IEEE India Council Conference.
[5] P. Nagabhushan, S. Nirmala(2009) ,"Text Extraction In Complex Color Document Images For Enhanced Readabi
[6] lity",Intelligent Information Management, pp: 120-133.
[7] DavodZaravi, HabibRostami, AlirezaMalahzaheh, S.S Mortazavi(2011)," Journals Subheadlines Text Extraction Using Wavelet Thresholding And New Projection Profile", WorldAcademy Of Science, Engineering And Technology .Issue 73.
[8] Karin Sobottka, Horst Bunke and HeinoKronenberg(2009), "Identification Of Text On ColoredBook And Journal Covers", ICDAR.
[9] Zhixin Shi, SrirangarajSetlur And VenuGovindaraju(2005), "Text Extraction From Gray Scale Historical Document Image Using Adaptive Local Connectivity Map", Proceeding Of TheEighth International Conference On Document Analysis And Recognition, Vol. 2, pp: 794–798.
[10] Syed SaqibBukhari , Thomas M. Breuel,FaisalShafait(2009), "Textline Information Extraction From Grayscale Camera-Captured Document Images ", ICIP Proceedings Of The 16th IEEEInternational Conference On Image Processing, pp: 2013 – 2016.
[11] Boussellaa , AymenBougacha, AbderrazakZahour, Haikal El Abed, Adel Alimi(2009) ,"Enhanced Text Extraction From Arabic Degraded Document Images Using Em Algorithm",10th International Conference On Document Analysis And Recognition.
[12] S. A. Angadi , M. M. Kodabagi(2009) , "A Texture Based Methodology For Text Region Extraction From Low Resolution Natural Scene Images ", International Journal Of ImageProcessing (Ijip) Volume(3), Issue(5).