ISSN: 2278 – 909X
International Journal of Advanced Research in Electronics and Communication Engineering (IJARECE)
Volume 1, Issue 4, October 2012

Text Extraction from Natural Scene Images

Shivananda V Seeri, Ranjana B Battur, Basavaraj S Sannakashappanavar

Abstract— **Text that appears in images contains important and useful information. The detection and extraction of text regions in an image is well known problem in the computer vision research and have been used in many applications. The present work shows two character extraction methods based on connected components. The performance of the different methods depends on character size. In the data, bigger characters are more prevalent and the most effective extraction method that is in the sequence: Sobel edge detection, Otsu binarization, connected component extraction and rule-based connected component filtering. It can be used in a large variety of application fields, such as mobile robot navigation, vehicle license detection and recognition, object identification, document retrieving, page segmentation, etc**

Keywords— **CoCos (connected components), images, sobel edge detection, binarization.**

I. Introduction

Text embedded in images contains large quantities of useful semantic information, which can be used to fully understand images. Text appears in images captured from natural scene through digital cameras or either in the form of documents such as scanned CD/book covers or video images. Video text can be broadly classified into two categories: overlay text and scene text. Overlay text refers to those characters generated by graphic titling machines and superimposed on video frames/images, such as video captions, while scene text occurs naturally as a part of scene, such as text in information boards/signs, nameplates, food containers, etc. 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. I found that the effectiveness of different methods strongly depends on character size. Since in natural scenes the observed characters may have widely different sizes, it is therefore difficult to extract all text areas from the image using only a single method. Also, current optical character recognition (OCR) techniques can only handle text against a plain monochrome background

[1] Xiaoqing Liu et al. proposed “Multiscale edge based text extraction from complex images”, method which automatically detects and extracts text present in the complex images using the multi scale edge information. This method is robust with respect to the font size, color, orientation and alignment and has good performance of character extraction. [2] D. Doermann et al. present a survey of application domains, technical challenges and solutions for recognizing documents captured by digital cameras.[3] T. Yamaguchi et al. The author presents a digits classification system to recognize telephone numbers written on signboards Candidate regions of digits are extracted from an image through edge extraction, enhancement and labeling. Since the digits in the images often have skew and slant, the digits are recognized after the skew and slant correction. To correct the skew, Hough transform is used, and the slant is corrected using the method of circumscribing digits with tilted rectangles.[4] J. Gllavata et al. The proposed approach is based on the application of a color reduction technique, a method for edge detection, and the localization of text regions using projection profile analyses and geometrical properties [5] Y. Liu The author introduces a new method to extract characters from scene images using mathematical morphology Kim et al. [6] implemented a hierarchical feature combination method to implement text extraction in natural scenes. However, authors admit that this method could not handle large text very well due to the use of local features that represents only local variations of image blocks. Yang[7], discusses problems of automatic sign recognition and translation. He presented a system capable of capturing images, detecting and recognizing signs, and translating them into a target language. He described methods for automatic sign extraction and translation. The sign translation, in conjunction with spoken language translation, can help international tourists to overcome language barriers. The technology can also help a visually handicapped person to increase environmental awareness.

Manuscript received Aug 15, 2012.

Shivananda V Seeri, MCA Dept, *BVBCET, Hubli, India,09844664391*

Ranjana B Battur*, CSE Dept, TCE, Gadag, India, 09886535592.*

Basavaraj S S*annakashappanavar, E&TC Dept. ,ADCET, Ashta , India , 09916319032.*

II. PROPOSED METHOD

2.1 TEXT EXTRACTION METHODS

The first step in developing text reading system is to address the problem of text extraction in natural scene images. Most studies are based on a single method for text detection. I found that the effectiveness of different methods strongly depends on character size. Since in natural scenes the observed characters may have widely different sizes, it is therefore difficult to extract all text areas from the image using only a single method. So in this work, I propose two text extraction methods based on connected components. The performance of the different methods depends on character size. In the data, bigger characters are more prevalent and the most effective extraction method that is in the sequence: Sobel edge detection, Otsu binarization, connected component extraction and rule-based connected component filtering

ISSN: 2278 – 909X
International Journal of Advanced Research in Electronics and Communication Engineering (IJARECE)
Volume 1, Issue 4, October 2012

2.1.1 CHARACTER EXTRACTION FROM THE EDGE IMAGE

In this method, Sobel edge detection is applied on each color channel of the RGB image. The three edge images are then combined into a single output image by taking the maximum of the three edge values corresponding to each pixel. The output image is binarized using Otsu's method [9] and finally CoCos (connected components) are extracted. This method will fail when the edges of several characters are lumped together into a single large CoCo that is eliminated by the selection rules. This often happens when the text characters are close to each other or when the background is not uniform.

2.1.2 CHARACTER EXTRACTION FROM THE REVERSE EDGE IMAGE

This method is complementary to the previous one; the binary image is reversed before connected component extraction. It will be effective only when characters are surrounded by connected edges and the inner ink area is not broken (as in the case of boldface characters).

2.2 METHODOLOGY

Edge based method

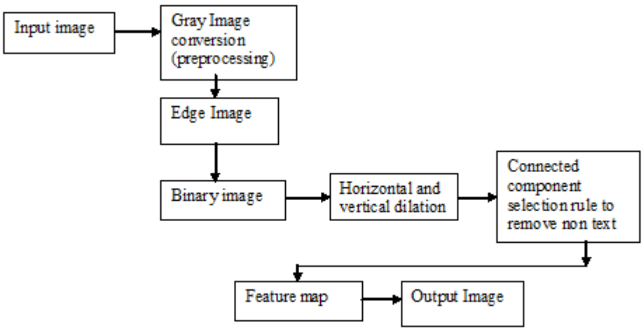


Fig 2.1 Block diagram for edge based method

Reverse edge based method

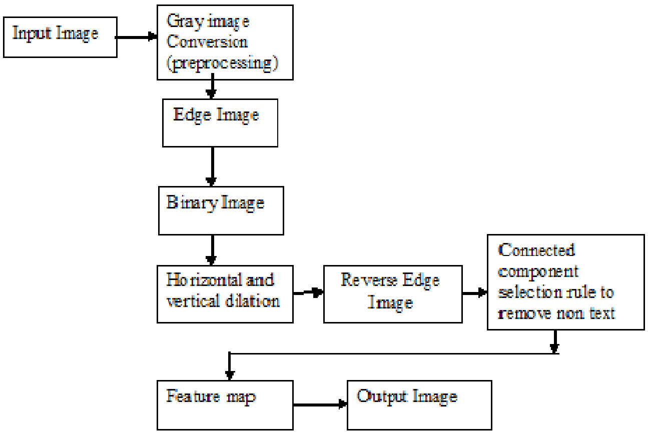


Fig 2.2 Block diagram for Reverse edge based method

Combined approach

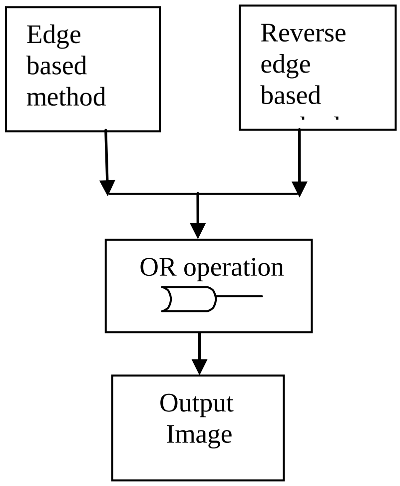


Fig 2.3 Block diagram for combined approach

2.2.1 PRE-PROCESSING

The input image is pre-processed to facilitate easier detection of text regions. The input image is RGB color image. The conversion is done using the MATLAB operation which takes the input RGB image and converts it into the corresponding gray image.

Fig 2.4:

1) Original image

2)Gray scale image.(a)x-axis

b) y-axis

c) z-axis

2.2.2 DETECTION OF EDGES

In this method, Sobel edge detection is applied on each color channel of the RGB image. The three edge images are then combined into a single output image by taking the maximum of the three edge values corresponding to each pixel. The output image is binarized using Otsu's method [9] and finally CoCos are extracted.

Fig 2.5:

ISSN: 2278 – 909X
International Journal of Advanced Research in Electronics and Communication Engineering (IJARECE)
Volume 1, Issue 4, October 2012

a)E1-image b)E2-image

c)E3 image d)combined edge image

The resultant edge image obtained is dilated in order to increase contrast between the detected edges and its background, making it easier to extract text regions. Both horizontal and vertical dilation is done. Figure 2.6(a) below shows the dilated edge image for the combined edge image from Figure 2.5(d), obtained.

Fig 2.6:

a) Dilated edge image b) Reverse edge image

2.2.3 CONNECTED-COMPONENT SELECTION RULES

It can be noticed that, up to now, the proposed methods are very general in nature and not specific to text detection. As expected, many of the extracted CoCos do not actually contain text characters. At this point simple rules are used to filter out the false detections. We impose constraints on the aspect ratio and area size to decrease the number of non-character candidates. In Fig. 2.7, *Wi* and *Hi* are the width and height of an extracted area; Δx and Δy are the distances between the centers of gravity of each area. Aspect ratio is computed as width / height. An important observation is that, generally, text characters do not appear alone, but together with other characters of similar dimensions and usually regularly placed in a horizontal string. We use the following rules to further eliminate from all the detected CoCos those that do not actually correspond to text characters (Fig.2.5):

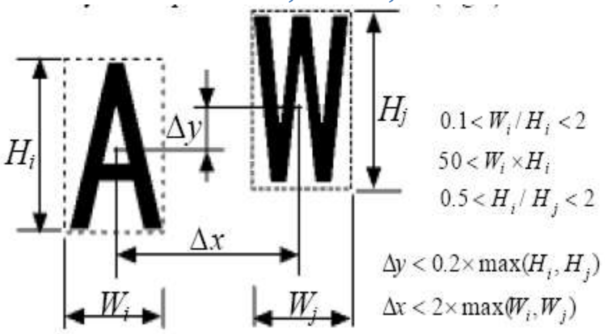


Fig 2.7 Character strings and rules

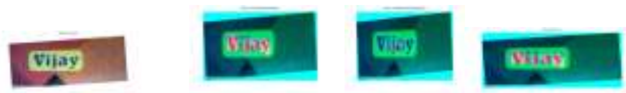
The system goes through all combinations of two CoCos and only those complying with all the selection rules succeed in becoming a number of the final proposed text region.

Fig 2.8 Final result

III EXPERIMENTAL RESULTS AND DISCUSSION

In order to evaluate the performance of the proposed method. We use test images of four types including book covers, object labels, nameplates and outdoor information signs. Fig. 1-5 show some of the results.

Test1



(a) (b) (c) (d)

Test 2

(a) (b) (c)

(d)

Test 3


(a) (b) (c)

(d)

Test4


(a) (b) (c)

(d)

Test 5


(a) (b) (c)

(d)

Test 1-5(a) original image (b) Result of edge based method (c) Result of reverse edge based method (d) combined result

OVERALL PERFORMANCE

Tests	Edge based method(E)		Reverse edge based method(R)		Combined approach(E+R)	
	Precision rate (%)	Recall rate (%)	Precision rate (%)	Recall rate (%)	Precision rate (%)	Recall rate (%)
Test1	100	100	71.42	50	100	100
Test2	100	66.66	100	96.66	100	100
Test3	100	100	100	86.66	100	100
Test4	95.00	61.29	100	84.3	100	96.87
Test5	90.00	69.23	100	84.3	92.30	92.30
Total	97.00	79.43	94.28	80.38	98.46	97.83

Table 3.1

Table 3.1 show the results obtained by each algorithm for given test images. 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.

Proposed method is compared with existing text extraction

Method	Precision rate (%)	Recall rate (%)
Proposed method	98.46	97.83
Samarabandu et al.[1]	91.8	96.6
J. Gllavata et al[4]	83.9	88.7
Wang et al [1] [2]	89.8	92.1
K.C. Kim et al [1][6]	63.7	82.8
J. Yang et al[7]	84.90	90.00

Table 3.2 Comparisons

Table 3.2 shows the performance comparison of our proposed method with several existing methods, where our proposed method shows a clear improvement over existing methods. In this table, the performance statistics of other methods are cited from published work.

IV. APPLICATIONS

There are numerous applications of a text information extraction system, including document analysis, vehicle license plate extraction, technical paper analysis, and object-oriented data compression. In the following, I briefly describe some of these applications.

Wearable or portable computers: With the rapid development of computer hardware technology, wearable computers are now a reality. A TIE system involving a hand-held device and camera was presented as an application of a wearable vision system. Translation camera can detect text in a scene image and translate Japanese text into English after performing character recognition similarly it can be implemented for translating Indian languages.

License/container plate recognition: There has already been a lot of work done on vehicle license plate and container plate recognition. Although container and vehicle license plates share many characteristics with scene text, many assumptions have been made regarding the image acquisition process (camera and vehicle position and direction, illumination, character types, and color) and geometric attributes of the text.

Text-based image indexing: This involves automatic text-based video structuring methods using caption data.

Texts in WWW images: The extraction of text from WWW images can provide relevant information on the Internet.

Industrial automation: Part identification can be accomplished by using the text information on each part.

Visually impaired person: Every year, the number of visually impaired persons is increasing due to eye diseases diabetes, traffic accidents and other causes. Therefore computer applications that provide support to the visually impaired persons have become an important theme. When a visually impaired person is walking around, it is important to get text information which is present in the scene. For example, a 'stop' sign at a crossing without acoustic signal has an important meaning. In general, way finding into a man-made environment is helped considerably by the ability to read signs. As an example, if the signboard of a store can be read, the shopping wishes of the blind person can be satisfied easier.

V. CONCLUSION AND FUTURE SCOPE

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 3.1. Hence proposed algorithm is best algorithm for text extraction from natural scene images.

Future Scope

Future work will focus on new methods for extracting small text characters with higher accuracy. Future work will also focus on handling images under poor light condition, uneven illumination, reflection and shadow.

REFERENCES

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

[2] D. Doermann, J. Liang, and H. Li, "Progress in Camera-Based Document Images Analysis ", *Proc.of the ICDAR*, 2003, pp. 606-616.

[3] T. Yamaguchi, Y. Nakano, M. Maruyama, H. Miyao and T.Hananoi, "Digit Classification on Signboards for Telephone Number Recognition", *Proc.of the ICDAR*, 2003, pp.359-363.

[4] J. Gllavata, R. Ewerth, and B. Freisleben, "A robust algorithm for text detection in images," in Image and Signal Processing and Analysis, 2003. ISPA 2003, 2003, Proceedings of the 3rd International Symposium on, pp. 611–616.

[5] Y. Liu, T. Yamamura, N. Ohnishi and N. Sugie, "Extraction of Character String Regions from a Scene Image", *IEICE Japan*, D-II, Vol. J81, No.4, 1998, pp.641-650.

[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 17 International Conference on Pattern Recognition (ICPR '04), IEEE.

[7] J. Yang, J. Gao, Y. Zhang, X. Chen and A. Waibel, "An Automatic Sign Recognition and Translation System", Proceedings of the Workshop on Perceptive User Interfaces (PUI'01), 2001, pp. 1-8.

[8] Q. Yuan, C. L. Tan, "Text Extraction from Gray Scale Document Images Using Edge Information," icdar, pp.0302, Sixth International Conference on Document Analysis and Recognition (ICDAR'01), 2001

[9] N. Otsu, "A Threshold Selection Method from Gray-Level Histogram", *IEEE Trans. Systems, Man and Cybernetics*, Vol. 9, 1979, pp. 62-69.

[10] S.M. Lucas, A. Panaretos, L. Sosa, A. Tang, S. Wong, and R. Young, "ICDAR 2003 Robust Reading Competitions", *Proc.of the ICDAR*, 2003, pp. 682-687.

[11] R C.Gonzalez,"Digital Image Processing Using MATLAB".

[12] A.K Jain,"Fundamentals of Digital Image Processing", Englewood cliff, NJ: Prentice Hall, 1989, ch.9, pp, 356-357

Shivananda V.Seeri, Department of MCA,BVBCET,Hubli,Karnataka, India. Prof. Shivananda V Seeri received his B.E in Computer Science under Karanataka University Dharwad,in the year 1992 and M.Tech in Computer Science under VTU,Belgaum ,in the year 2001.He is currently working as Asst.Professor and HOD in MCA department in BVBCET,Hubli. He is also Persuing his Phd under VTU,Belgaum. His area of interest is image Processing.

Ranjana B. Battur, Department of Computer Science And Engineering, Tontadaraya College of Engineering, Gadag, Karnataka, India. Prof. Ranjana B Battur received her B.E. in Computer Science And Engineering from BVBCET, Hubli under VTU, Belgaum in the year 2007, and received her M.Tech in Computer Science from BVBCET,Hubli under VTU, Belgaum in the year 2010. She is currently working as a Assistant Professor in TCE, Gadag in the Department of Computer Science And Engineering. Her area of interest is Image Processing.

Basavaraj S. Sannakashappanavar, Department of Electronics & TeleCommunication Engineering, Annasaheb Dange College of Engineering and Technology, Ashta, Dist-Sangli, Maharastra, India, Mobile No: 09916319032.Basavaraj S Sannakashappanavar received his B.E. in Electronics and Communication Engineering from SKSVMACET, Laxmeshwar under VTU, Belgaum in the year 2010, and received his M.Tech in Digital Communication from BEC, Bagalkot under VTU,Belgaum in the year 2012.He is currently working as a Assistant Professor in ADCET,Ashta in the Department of Electronics and Telecommunication Engineering. His area of interest includes Speech Processing, Wireless Sensor Networks and Image Processing.