

An Adaptive Approach to Text Detection and Recognition in Natural Scene Images

Bramara K A
Dept. of CSE
GSSSIETW, Mysuru

Kavyashree S
Dept. of CSE
GSSSIETW, Mysuru

Madhurya T
Dept. of CSE
GSSSIETW, Mysuru

Madhushri S
Dept. of CSE
GSSSIETW, Mysuru

Rajath A N
Assistant Professor, Dept. of CSE
GSSSIETW, Mysuru

Abstract:- The present work is an attempt to develop a method for extracting the text from natural scene images. The images here are nothing but the camera captured images. Text messages in an image usually contain useful information related to the scene, such as location, name, direction or warning. Text detection in natural scenes is an important but challenging problem because of variations in the text fonts, size, line orientation, complex background in image and non-uniform illuminations. To overcome these problems, effective features for text image recognition are used. We can use an Optical Character Reader to recognize the extracted text. The basis of our scheme is to analyze the CCs. This is done to extract text from scenic images captured by camera. Our scheme makes use of mathematical morphological operations to extract square region that contains text. Also the binarization of scenic images was studied. Here the effectiveness of the adaptive thresholding approach was observed.

Keywords— Connected Components; Morphological Operations; Erosion and Dilation; Extracted text;

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. The 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. 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. Different approaches for the extraction of text regions from images have been proposed based on basic properties of text.

The text in natural images and video frames such as street signs, vehicle license plates, billboards, writing on shirts, sport scores, time and location stamps, is a powerful source of knowledge in building image and video indexing and retrieval systems. This kind of text also provides useful content information for video understanding and automatic navigation systems. Due to the wide range of applications,

methods for text extraction have been proposed. Some of them require binary input images; which restricts their application when the text is embedded in an image with a complex background, because binarization techniques[10] usually produce poor results for complicated images. On the other hand, some methods also use color information to detect text areas; color information can be helpful, but it is not available in all situations. Moreover, for a human observer, intensity information is enough to segment the text areas, so most methods perform text extraction on grey-scale images, i.e., even if a color input image is available, it is first converted to grey-scale.

Mainly, two types of methods exist for text detection and localization: one is region based and another one is connected components based (CC-based). In region based method of localization and detection, the texture of the text is analyzed. The text is recognized by the estimation of likelihood after feature vectors are extracted from each local region using a classifier. Then these neighboring text regions are merged to generate different text blocks. Since text regions have different textual properties from the non-text areas, so this method can detect and extract the text accurately even if some noise is present in the text. The connected components based method segments the text components, which are then detected by edge detection technique or by color clustering. The number of connected components which are already segmented is relatively very less, the connected components method lowers the time taken for computation and the located and extracted text components is directly used for the recognition and can be further processed.

The focus of this paper is on extracting the text from natural scene images captured by camera using connected components based method and then by applying morphological operations to detect and localize the text regions in images. Edge detection technique is applied to detect the edges in the text. Finally the text detected in the text regions is extracted.

II. REVIEW OF OTHER METHODS

This section provides a descriptive summary of some methods that have been implemented and tested for text

extraction. As far as detection of the text region is concerned, researchers have found many methods of locating text regions. For example, a method based on text recognition technique named as sliding windows (SW) was also proposed from oriented text to feed HMM for text string recognition [2], in sliding window foreground pixels of text lines are chosen and fed to a curved fitting algorithm, For feature extraction [2], a sliding window of fixed width to extract a sequence of frames from the curved text was also proposed, two frame-wise feature extraction algorithms namely: LGH (Local Gradient Histogram) and MB (Marti-Bunke) was also evaluated.

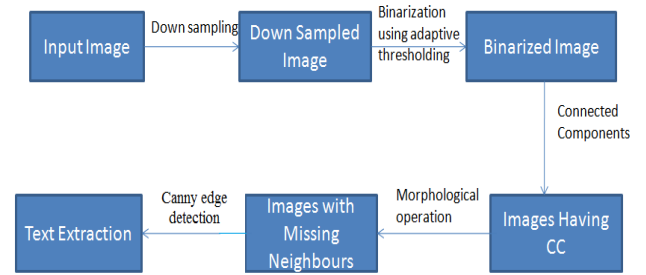
In addition, Real-time Scene Text Detection method was also proposed based on stroke model [3], which is a corresponding method to extract character strokes based on difference of Gaussian (DoG) filters, a novel method [3] based on the stroke model to extract character strokes was also used, also contains a aggregation method to group candidate characters into text lines and a text line aggregating method utilizing the inherent text layout. In an Integrated Approach for the Multilingual Scene Text Detection [4], an integrated scene text detection system to detect textual regions from scene images with no language restriction was proposed, to refine the character candidates information in images, canny detector and MSER [4], to retrieve the edges and local features of characters were used, if the text candidates actually contain a text string or not is determined using a SVM classifier.

Currently, some researchers prefer a hybrid detection algorithm, where a text is detected and located based on edge detection method. Novel approach method is used for Devanagari text extraction from natural scene images [1], the basis of their scheme is to analyze the connected components(CCs). This is done to extract Devanagari text from scenic images captured by camera. The presence of headline is unique to this script.

In this paper, an algorithm enforcing a text extraction using connected components, morphological operations and edge detection technique has been implemented.

III. PROPOSED ALGORITHM TOWARDS THE EXTRACTION OF TEXT

In the author’s previous work [1], a novel based approach was presented. We propose in this section an enhanced version of text extraction algorithm. As shown in figure(1) the proposed text extraction algorithm consists of five main stages: (1) Down-sampling the image. (2) Converting the down-sampled image into 8 bit gray scale then to binary image. (3) Detecting the connected components. (4) removing unwanted region by applying morphological operation. (5) Edge detection and text extraction from the morphologized image. General scheme for text extraction is shown in figure (1).

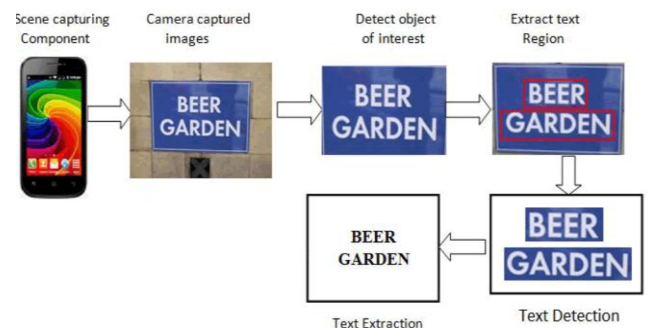


Figure(1) : Proposed model for text extraction

Algorithm for text extraction from natural scenic image:

- Step 1: The input image is down-sampled by an integral multiple, to diminish the dimension of the image to the closest multiple of 0.35 Megapixels.
- Step 2: The image is converted to 8-bit grayscale.
- Step 3: The CCs (C) are obtained from the binary image corresponding to the gray image.
- Step 4: All horizontal line segments are computed by applying morphological opening operations on every connected components CCs. We apply mathematical morphological operations on connected components, e.g. erosion and then dilation.
- Step 5: The unselected CCs are revisited. The neighbourhood connected components are estimated by using the canny edge detection technique over the selected connected component’s text.
- Step 6: A virtual circle of a finite radius is created which traverses through each pixel found on the canny edge.
- Step 7: All those connected components are selected whose pixels fall in that circle.
- Step 8: The intensity level of those components are compared so that false connected components are not selected.
- Step 9: The desired text is extracted from natural scenic image.

Figure below demonstrates how the flow chart of the proposed model works.



Figure(2) : Flow chart for the proposed model

In the above figure, image captured by camera is given as an input where it undergoes several steps of pre-processing.

PRE-PROCESSING:

Depending on the digital camera’s resolution there is variation in the input image’s size. Normally the resolution starts from 1 Megapixel. In the beginning, the input image is down-sampled by an integral multiple so that the dimension of the image is diminished to the close multiple of 0.35 Megapixels. This is done so that the proposed algorithm should work smoothly. Afterwards, the down-sampled image was changed to 8-bit grayscale.

Grayscale image is then converted binary image using adaptive thresholding method. Binarized image then undergoes connected component analysis[9] where the CC's of each pixel is calculated in the form matrix. Later morphological operations are applied i.e, erosion and dilation to remove unwanted pixel from the foreground of an image and to locate and extract the text regions from an image.

CC's which are unselected are revisited by selecting the neighbourhood connected components of text regions by using canny edge detection technique to detect the edges that are not visited while calculating the CC's. A virtual circle of finite radius is created which is used to traverse each pixel found in canny edge. All the connected components whose pixels fall in that circle are selected so as to avoid the selection of false pixels of connected components by comparing the intensity level of components in that circle. Finally the desired text is extracted from natural scene images.

IV. EXPERIMENTAL RESULTS

We collected over 750 images from various environments (signboards, book covers and English and Kannada characters). Figure below shows the performance evaluation for the collected images (English and Kannada characters) where the proposed method is successfully implemented even on Kannada language (characters).



Figure(3) : Performance evaluation

Various results obtained by implementing the proposed algorithm of text extraction are shown below. Almost in every case, the text areas are detected very well as shown in the final binary images. Non-text areas are also eliminated effectively. The suggested method worked successfully even in the case of multilingual languages, and complex background. Overall, 87% of accuracy is obtained from this algorithm.



| Invariance to | Lightness change(s) &/or shift(s) | Color change(s) &/or shift(s) | Color & Lightness change(s) &/or shift(s) |
|-------------------|-----------------------------------|-------------------------------|---|
| Examples of image | | | |
| Results obtained | | | |
| Examples of image | | | |
| Results obtained | | | |

REFERENCES

1. Hrishav raj, Rajib Ghosh, "Devanagari Text Extraction from Natural Scene Images", 2014 International Conference on Advances in Computing, Communications and Informatics (ICACCI).
2. Sangheeta Roy, Partha Pratim Roy, Palaiahnakote Shivakumara, Georgios Louloudis, Chew Lim Tan, Umпада Pal, "HMM-based Multi Oriented Text Recognition in Natural Scene Image", Tata Consultancy Services, Kolkata, India. 2013 Second IAPR Asian Conference on Pattern Recognition.
3. Yi Liu, Dongming Zhang, Yongdong Zhang and Shouxun Lin, "Real-time Scene Text Detection Based on Stroke Model", Key Lab of Intelligent Information Processing of Chinese Academy of Sciences (CAS), Institute of Computing Technology, CAS, Beijing 100190, China, 2014 , 22nd International Conference on Pattern Recognition.
4. Wen-Hung Liao, Yi-Hsuan Liang, Yi-Chieh Wu, "An Integrated Approach for Multilingual Scene Text Detection", 2015 Seventh International Conference of Soft Computing and Pattern Recognition (SoCPaR 2015).
5. Vinod H C, Niranjan S K, Manjunath Aradhya V N, "An application of Fourier statistical features in scene text detection", 2014 International Conference on Contemporary Computing and Informatics (IC3I).
6. Ranjit Ghoshal, Anandarup Roy, Swapan Kr. Parui, "Recognition of Bangla text from Scene Images through Perspective Correction", 2011 International Conference on Image Information Processing (ICIIP 2011) .
7. U Bhattacharya, S K Parui and S Monda, "Devanagari and Bangla Text Extraction from Natural Scene Images", Computer Vision and Pattern Recognition Unit, Indian Statistical Institute, Kolkata – 108, India, 2009 10th International Conference on Document Analysis and Recognition.
8. Manoj Kumar , Young Chul Kim, Guee Sang Lee, "Text Detection using Multilayer Separation in Real Scene Images", Department of Computer Science, Chonnam National University, Gwangju, Korea, 2010 10th IEEE International Conference on Computer and Information Technology (CIT 2010).
9. Rajath A N, Parashiva Murthy B M, "An Adaptive Approach to Vehicle Number Plate Detection for Indian Style Based", Department of Computer Science and engineering, GSSS Institute of Engineering and Technology for Women, Mysuru, India, International Journal of Modern Computer Science (IJMCS 2012).
10. Rajath A N, "An Adaptive Approach: Text Line Extraction from Multi-Skewed Hand Written Documents", Department of Computer Science and engineering, GSSS Institute of Engineering and Technology for Women, Mysuru, India, IJCSSET(www.ijcsset.net) | June 2015 | Vol 5, Issue 6,158-161.



Figure(4) : Output images where the algorithm gave the perfect performance

V. CONCLUSION

Natural scene text extraction and understanding represents new challenges based on the explosion of digital still cameras or camera phones in the market. The proposed method resolves the difficulties of text extraction on natural scene images caused by non-uniform illumination, complex backgrounds and the existence of text like objects. Our approach can capable of handling multilingual texts in images and it can extract multi-line texts from the image. In future, work can be continued to implement on slanted images, curved images and night time condition images. Further it can also be converted to speech recognition in different languages.