International Journal of Latest Trends in Engineering and Technology
Vol.(10)Issue(4), pp.001-004
DOI: http://dx.doi.org/10.21172/1.104.01
e-ISSN:2278-621X

TEXT DETECTION FROM CAMERA CAPTURED IMAGE AND VIDEO

Chowdhury Md Mizan[1], Tridib Chakroborty[2]

Abstract-In recent years, with increasing popularity of portable devices for capturing images,text extraction, etc.Extraction of text information from images or scene involves detection, localization, tracking, segmentation, enhancement and character recognition. But variations involved in text such as font style, size, orientation, alignment and illumination effect, as well as low image contrast and complex background details make text extraction from image more difficult job.Text in video and images is an extremely important feature to extract summary knowledge of video or image. But sometime it happen that few text components feel unnecessary, and so various methods have been proposed till a day to detect, and extract the text out of image and video. This paper performs coverimg survey a large area of various techniques available for text detection and extraction from images and video and images automatically to create visually plausible output like videos/images without any of the embedded text..

1. INTRODUCTION

Text from images or videos is an important applications in many research field like document processing [1, 2], image indexing, video content summary [3–5], video retrieval [6], video understanding [7], and since including.Texts embedded in an image or a frame originally capture from camera .The most important media contexts such as player's name, title,creation date, story introduction,scene text etc. including.

Now a days a large amount of videoes or lecture videoes are produce every day.Students can search a specific portion of video may get by video searching.for example we can say that NPTEL site has large no. of lecture videoes are available.Students can download it or search it directly.Suppose a student needs specially TCP/IP part.But he has to be browse all the video.Because he dosenot know in which section TCP/IP videoes are available or it may reside in any section. And each video size is 60 minute.Almost 40 lecture videoes are available.So,It is very tought job to locate the exact position of the video in a desired time.So,student has to individually browse through each of the choosen videoes to find the portion where TCP/Ip protocol is discussd.If the student is not familiar with the area or if the topic is very specific or particular the student may not be able to decide the videoes to be scrutinized.A very big amount of textual metadata will be created by using OCR and ASR(automatic speech recognition) which provide the contain of the lecture.

Digital video has become a popular storage and exchange medium due to rapid development in recording technology. Videotaping of lecture video is more common e-Learning. A number of universities and research institutes are taking the opportunity to record their lectures and publish them online for students. As a result, there has been a lot of multimedia data are available in on Web. Hence, for a user it is nearly impossible to find desired videos without a search function within a video archive.

Content-based video retrieval have a very big range of applications such as quick browsing of video folders, analysis of visual electronic commerce, remote instruction, digital museums, news event analysis, intelligent management of web videos (useful video search and harmful video tracing), and video surveillance also

2. RELATED WORK:

There are several technique to detection ,localization and extraction of text from image or videoes.Here we discuss some proposal to text detection .Wonjun Kim et. al. has proposed a novel frame for Overlay text detection and extraction from image or video. First the transition map is generated. Then candidate regions are extracted and overlay text regions are gets detected on the basis of occurrence of overlay text in each candidate region. At the last localization of overlay text regions is performed by projecting overlay text pixels in transition map and immediatsely a step of extraction is carried out [4]. X. Chen, J. Yang, J. Zhang, A. Waibel [7], combined 1) multiresolution and multiscale edge detection 2) adaptive searching, 3) color analysis, 4) affine rectification in a hierarchical framework for sign detection with different priority at each phase to handle the text in different orientations ,sizes, color distributions and backgrounds.here they have used affine rectification to improve deformation of the text regions caused by not proper circumstances camera view angle. They extracted features from an image directly rather using binary information for OCR.In this method binary conversion are not required. They proposed a local intensity normalization method to effectively handle lighting variations, a Gabor transform is used to find local

[1] Department of Information technology, GNIT, Kolkata, West Bengal, India
[2] Department of Information technology, GNIT, Kolkata, West Bengal, India

features and then for feature selection a linear discriminate analysis (LDA) method is applied. Best utilization of there is Chinese system,which can automatically take photo from camera and detect the text and translated into English automatically. The procedure can extensively improve text detection rate and optical character recognition (OCR) accuracy.
Y. Pan, X. Hou, and C. Liu [5], proposed a hybrid approach to localize scene texts by integrating region information into a robust CC-based method. Parameters of a conditional random field (CRF) model are jointly optimized by supervised Learning and the binary contextual component relationships with the unary component properties are incorporated in the CRF model. The proposed hybrid approach needs further improvements because this approach fails on some texts that are difficult to segment. The speed of the proposed hybrid approach need to be accelerated further. Text recognition should be included with text localization to complete the need of text information extraction as well.

3. SCENE TEXT DETECTION METHODS

3.1. Region-Based Methods

Region based method is use the properties of the color or gray scale in a text region or their differences with the corresponding properties of the background.In region based method those pixel having similar characterstics are grouped together.The methods in this category suffer from low computation speed but can simultaneously detect texts at any scale and are not limited to horizontal texts. These methods are divided into two sub-approaches: connected component (CC)-based and edge-based method.In connected comoponent based method first discurd the majority of the background pixel using low level filters and then construct component candidatesfrom remaining pixels.But in edge based method focus on high contrast between the text and the background.

3.1.1 CC-based Methods

In Connected Component based method we mainly use a bottom up approach.In bottom up we grouping the small components into larger components until or unless all regions are know in the image or videoes.A geometrical analysis is required to combine the text components using the spatial arrangement of the components so as to filter out non-text components and mark the boundaries of the text regions.The
main advantage of CC-based methods is that it have lower computational complexity.The performance of the CC-based methods is degraded while dealing with the texts in complex background.

3.1.2 Edge-Based Methods

Any other many textual properties in an image or videoes Edge based startegies specially in high contrast and in between the text and the background the text.The edges of the text boundary are unknow to programmer and and merged and then many heuristic are used to separate the non text regions. Generally, an edge filter is used for the edge detection and a morphological operator is used for the merging stage.

3.2. Texture-Based Methods

Texture based method is a feature based algorithm in which gray-level co-occurrence matrix which is used to calculate the features such as homogeneity, contrast and dissimilarity. Texture-based methods make use of the observation that text in images have distinct textural properties that distinguish them from the background. The techniques based on Wavelet, Gabor filters, FFT, spatial variance, etc. can be utilized to detect the textural properties of a text region in an image. The methods in this category also having certain limitations including big computational complexity because of the need of scanning the image at several scales, inability to detect sufficiently slanted text [7].

3.3. Stroke Width Transform (SWT)

The Stroke Width Transform (SWT), has been much more interesting due to its simplicity and efficiency. However, the SWT has difficulty in situations such as blur, low contrast, and illumination change images since it highly relies on the outcome from the edge detector.in order to overcome the limitations of the previous methods such as high computational complexity and the difficulty to select the best features for scene text detection.SWT converts the the each pixel value in widh of most suitable stroke and then neighbouring pixel with similar stroke and merge them into connected components so that the system can detect of its font.

4. STEPS IN IMAGE SEGMENTATION :

4.1. The general steps that are involved in Image segmentation systems are:

- Image acquisition
- Pre-processing
- Segmentation

4.1.1 Image Acquisition:

This is the stage where the image under consideration is taken. In the case of online recognition system, a specialized hardware is implemented ,for offline systems, the images are obtained either through a scanner or a camera. On any occasion an image is acquired, there will be some variations in the intensity levels along the image. Also noise gets added to the image. Hence preprocessing is required for adjusting the intensity levels and to denoise the image.

4.1.2 Preprocessing:

Pre processing is required for better performing recognition system. In this case, the acquired image is processed to remove any noise that may have incurred into the image during the time of acquisition or during the time of transmission. A colored image then it will be transformed to a gray image before proceeding with the noise removal procedure.The de-noised image is then converted to a binary image with suitable threshold.

Segmentation: This segmentation refers to a process of segregation an image into groups of pixels which are homogeneous with consideration to some benchmark. This distribution algorithms are area oriented instead of pixel oriented. At the output of this process is the splitting up of the image into connected areas. Thus, segmentation is concerned with dividing an image into meaningful regions. Image segmentation can be broadly classified into two types.

1. Local Segmentation: It deals with the segmenting sub images which are small windows on a whole image.

2.Global segmentation: It deals with the images subsist of comparably large number of pixels and makes estimated parameter values for global segments more robust.

For character segmentation, first the image must be segmented row-wise (line segmentation), then each rowmust be segmented column-wise (word segmentation).Certainly,aspect can be extracted using advisable algorithms such as edge detection technique; histogram based methods or connected component analysis

4.2. Text extraction Techniques

1) Compression Based Algorithm: This algorithm pre-suppose that the optimal segmentation is the one that minimizes the overall conceivable segmentation, coding length of the data. The connection between these two concepts is that segmentation tries to find patterns in an image and any consistency in the image can be used to compares it. The algorithm explain each segment by its texture and boundary shape. This algorithm was implemented by W.J Teahan, Yingying Wen, Rodger Mcnab and Lan H.

Corner Response Based Method: A novel text detection and localization method based on corner response consist of 3 stages: (1) Computing corner response in multi-scale space and thresholding it to get the candidate region of text; (2) Verifying the candidate region by combining color and size range features;(3) Locating the text line using bounding box. Corner is a special two-dimensional feature point which has high curvature in the region boundary. It can be located by finding the local maximum in corner response (CR). In corner points in video frame are used to generate connected component. But they use just the number of corner points, not CR, to classify text and non-text region.

Edge Detection Algorithm: This algorithm is well developed field on its own within image processing. The region boundaries and edge are closely related, since there is often a shape adjustment in intensity at the region boundaries. This detection techniques have therefore been used as the base of another segmentation technique. The edge identified by edge detection is often disconnected. To portion an object from an image however, one needs closed region boundaries. Salem Saleh Al-amril, Dr N.V kalyankar implemented image segmentation by using Edge detection .They did a comparative study using seven technique of the edge detection segment. They arerobert, canny, laplacian, and edge maximum technique on the Saturn original image and found that EMT and Perwitt techniques respectively are the best techniques for edge detection.

4)Nearest Neighbor Clustering Based Method (NNC): In this process, a novel approach for line and character segmentation in an epigraphically script based on closest neighbor clustering process is presented. The expected algorithm scans the given input image from the left corner. When it confrontation the first black pixel, it describe the complete character through connected component. This character is segmented and placed at different location. The centered of the character is computed. Similarly the second character is identified and the centered is computed. The Euclidean distance between the centroids is computed to know whether the character belongs to the same line or next line. This is determined based on the threshold which is based on the assumption that the space between the text lines is greater than that between the characters. In this way, the text lines and characters are segmented which could be used for the classification process

5. CONCLUSION

In this paper, we present a short survey on various methods used for scene text detection and video text detection. We have mainly discussed three methods namely region based methods, texture based methods and Stroke Width transform (SWT). Stroke Width Transform method is recently useful to solve the problems of scene text detection related to the large variations in character font, size, texture, color, etc. It is been concluded that text detection and text recognition are the techniques of patterns recognition. The technique of segmentations are required which will segment the text portion from the input image. The neural networks technique will recognize the text from the input image. In this paper , various techniques of image segmentation, edge detection and neural networks are reviewed and discussed.

6. REFERENCES

[1] K. Wang, B. Babenko, and S. Belongie, ―End-to-end scene text recognition,‖ 2011, in IEEE International Conference on Computer Vision (ICCV).

[2] T.Wang, David J. Wu, A.Coates, Andrew Y. Ng,‖ End-to-End Text Recognition with Convolutional Neural Networks‖, 2013, ICPR

[3] A.Bissacco, M.Cummins, Y.Netzer, H.Neven,‖ Photoocr: Reading Text in Uncontrolled Conditions‖, 2013, IEEE International Conference on Computer Vision (ICCV).

[4] M.Jaderberg, K.Simonyan, A.Vedaldi, A.Zisserman,‖ Reading Text in theWild with Convolutional Neural Networks‖, 2016, IEEE, Int J Comput Vis (2016)

[5] W.Huang, Y.Qiao, and X.Tang,‖ Robust Scene Text Detection with Convo lution Neural Network Induced MSER Trees‖, 2014, IEEE Part IV, LNCS 8692, pp. 497–511

[6] X.Cheng Yin, W.Yi Pei, J.Zhang and H.Wei Hao,‖ Multi-Orientation Scene Text Detection with Adaptive Clusteri.ng‖, 2015, IEEE Transcations on Psttern Anslysis and Machine Intelligence.

[7] Z.Zhang, W.Shen, C.Yao, X.Bai,‖ Symmetry-Based Text Line Detection in Natural Scenes‖, 2015, IEEE Computer Vision and Pattern Recognition (CVPR).

[8] B.Singh, and R.Maini,‖ Skew Detection and Correction of Gurmukhi Words from Natural Scene Images‖, International Journal of Signal Processing, Image Processing and Pattern Recognition., Vol.9, 2016.

[9] C. Yao, X. Bai, and W. Liu, ―A unified framework for multi oriented text detection and recognition,‖ Image Processing, IEEE Transactions on, vol. 23, no. 11, pp. 4737–4749, 2014.

[10] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama, and T. Darrell, ―Caffe: Convolutional architecture for fast feature embedding,'' 2014, in Proceedings of the ACM International Conference.

[11] L. Sun, Q. Huo, W. Jia, and K. Chen, ―A robust approach for text detection from natural scene images,‖ Pattern