

A Survey: Text Detection, Extraction and Inpainting

Snehal A Natekar¹, Mrs.Rane C.V.²

¹(*Electronics &Telecommunication, RSCOE, PUNE/ University of Pune, India*)

²(*Electronics &Telecommunication, RSCOE, PUNE,/University of Pune,India*)

ABSTRACT : *Embedded text in the video or imaged is very much valuable which provides information of video or images. Though text present in video provides important information not all are of them are necessary as it hide important portion of the video. So, various techniques, methods have been invented to detect and remove these unwanted texts from video. This paper performs survey of various techniques which are present for text detection and removal from videos and images.*

Keywords - *Text detection, Bandlets, inpainting, edge detection Stroke Width Transform*

I. INTRODUCTION

Text embedded in the video present in the form of subtitles, captions, logs etc. provide valuable information. Example of such embedded text can be largely found in news and advertisement running at the bottom of TV shows. Even if this text plays an important role in providing additional information but it is also true that these texts occupy portion of the video. So, it is best have to remove undesired text from video. Video text removal schemes basically have two main stages i] automatic video text detection and ii] effective video restoration.

II. RELATED WORK

For text detection and removal many techniques are available. Few of the related important work are studied here.

W. Kimet. al. has proposed a novel frame for Overlay text detection and extraction from video. At First the transition map is generated. After that candidate regions are extracted. On the basis of overlay text, overlay text regions are detected in each candidate region. At the last localisation of overlay text regions is performed by projecting overlay text pixels in transition map and immediately a step of extraction is carried out[1].

M.R. Lyuet. al. proposed multilingual video text detection, localization and extraction using comprehensive method. Edge detection, local thresholding and hysteresis edge recovery are three methods jointly perform text detection. Then coarse -to-fine localisation scheme is applied. Then for text extraction adaptive thresholding, dam point labelling and inward filling methods are used [2].

Yen Lin Chen proposed automatic text extraction, removal and inpainting of complex document images. In this image is decomposed into distinct object planes such as textual regions, non textual regions, background textures etc. Then text detection is carried out using knowledge based text extraction and identification. Then an effective adaptive inpainting neighbourhood scheme is applied [3].

Leon [8] proposed caption text extraction for indexing purpose using a hierarchical region based image model. He uses geometric property like color, size, alignment, motion of text to detect caption text in image. Here we use description to detect various objects in image.

A.Moslehet. al. has proposed automatic text extraction, removal and inpainting scheme for video text detection. He proposed two state frameworks. For video text detection unsupervised clustering is performed on connected components produced by Stroke Width Transform. The restoration task is performed by applying spatio temporal geometric flow extracted by bandlets to reconstruct missing data. Then at last inpainting is performed [4].

N.Ezaki [9] proposed text detection from natural scene images. He uses character feature to detect the text in image. The text component region is detected from background region by using consistency of gradient of

pixels. To localize text character features like edge strength, edge density and horizontal distribution of intensity level of text pixels is used.

Mohammad Khodadadiet. al. has proposed text detection, extraction and inpainting algorithm in color image. Text is detected by using stroke filter, new segmentation and verification algorithm. Color histogram technique is used for background and text color estimation in candidate block [5].

III. METHODOLOGY

W. Kimet. al. [1].proposed method consists of combination of change of intensity and modified saturation. With help of this transaction map is generated. The basic advantage is that transaction map is generated very well even in complex background. Connected components which are smaller than threshold are removed and remaining all CC are reshaped in order to have smooth boundaries. For text region detection, in overlay text detection step the aspect ratio of overlay text region, density of transition pixels and texture based approach are the basic components. If probability of overlay (POT) is larger than predefined value, the corresponding region is considered as overlay text regions. For overlay text region refinement, horizontal projection of transitions of pixels on the transaction map is performed, then null point is removed. Then null point is found again by using vertical projection. An overlay text updation is used to detect text in next frame. In, the last step, adaptive thresholding and modified dam point labeling are performed and using inward filling of background of text region, overlay text is extracted.

Basically there are so many methods proposed for the detection of English language text from till a day. But if the text of any other language appears in the video due to certain reasons system fails many times while detecting, removing this text from video. There are so many characteristics are used in video text detection, localization and extraction. Color, contrast orientation, stationary location are the language independent characteristics are used. Michael R Lyuet. Al. [2] proposed comprehensive method for detection, localization and extraction of multilingual text from video.

M.R.Lyuet. al.[2] proposed sequential multiresolution paradigm technique to overcome drawback of parallel multiresolution paradigm. Initially video segment is sampled at two frames per second and each frame is converted into 256 level gray scale image. For further operation these images are considered as original images to detect, localize and extract text. The working of sequential multiresolution paradigm is divided into levels from 1 to n. The current edge map is first scaled down by factor $f(I)$ and text detection and localization operations are performed. At last image is scaled up by same factor $f(I)$ and the original image is generated. Text detection is carried out using local thresholding which separates text stream from background and hysteresis edge recovery. Using coarse-to-fine localization method [6] text localization is performed. Then using multiframe verification, the already detected text from previous frame or which are transitory is eliminated. The signature distance matrix is used for dealing with possible location offset and edge density changes of same text in different frames.

Following are the three basic tasks used for text extraction:

1) Adaptive thresholding: This is used to produce a better binary image for different background intensities.

2) Dam point labeling: The pixels of text strokes are called as dam points, which are used to prevent the text pixels to be filled by flooding.

3)Inward filling: The every pixel from Extended- Region are scanned and if a pixel is "white", all the connected white pixels are find using flood fill and set to "black". And all the non black pixels are set to "white" after inward filling.

Yen-Lin Chen has proposed an automatic approach for text extraction, removal and inpainting of complex document Images. In this document image is decomposed into distinct objects planes such as textual region, non-textual regions, background textures etc .Using knowledge based text extraction and identification the texts with different characteristics from each object planes are detected. Then for text removal and inpainting adaptive neighborhood adjustment scheme is applied [3].

Moslehet.al.[4] proposed new approach for automatic text detection and removal from video. In this method, video streams are splitted into frames. Each frame is considered as separate image. The text is detected in each frame by generating edge map. Connected components are generated by using stroke width transform (SWT) [7] and for detection of text regions unsupervised clustering is used. Then with the help of Lucas-Kanade optical flow computation algorithm video global motion patter is calculated. The text objects are tracked using CAMSHIFT algorithm [4] and filtering of non caption text takes place.

After extraction of text from video bandlet transform is used. Bandlet transform effectively represent image geometry and exploits spatio-temporal regularities to perform regularization to address the restoration. In this video inpainting method, each caption is considered as volume so smooth video without text is generated.

In Leon [8] hierarchical based image model is used as descriptor to classify text object. To detect caption from complex image property of caption text is exploited. We can find caption text from image using contrast value because caption text pixels have similar contrast value. Area is analyzed using discrete wavelet transform. Wavelet transform has efficient texture representation. For the classification text feature vectors are generated. Different geometric property like occupancy, aspect ratio, height of text string area occupied are calculated and fed to classification algorithm.

N.Ezaki [9] use character features to detect text in image. By applying local thresholding edge density is calculated. Then edge strength feature is used for binarization. Local thresholding technique is used to keep low contrast text. Local profiles are used as classifier for identification of text. Contrast levels of connected components are analyzed.

Bui [10] use image inpainting using wavelet based inter and intro scale dependency. Texture features are effectively represented by using wavelet transform. This will be used to approximate boundary data to be fill in the missing region. This method assumes global structure of image and use property like shape to interpolate data into missing regions. Wavelet based transformation algorithm decomposes image into four frequency composition. Wavelet coefficient is found for each segment considering scaling function inpainting is done.

Mohammad Khodadadi proposed method consists to get stroke filter image for all the RGB channels of image. Using histogram text localization and extraction is performed. For detecting border of text region accurately inpainting process using texture synthesis and matching is performed with little modification in algorithm by considering high intensity variation [5].

IV. CONCLUSION

There are many video text detection, extraction and inpainting techniques are present. All have different performance as per different type of data or background present in video. Using different attributes such as size, font, style, contrast, color, connected components the text regions are easily differentiated from their background or other regions within the image. This paper does survey of many previously proposed techniques for video text detection, extraction and inpainting.

REFERENCES

- [1]. W. Kim and C. Kim, "A new approach for overlay text detection and extraction from complex video scene," *IEEE Trans. Image Process.*, vol. 18, no. 2, pp. 401–411, Feb. 2009.
- [2]. M. R. Lyu, J. Song, and M. Cai, "A comprehensive method for multilingual video text detection, localization, and extraction," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 15, no. 2, pp. 243–255, Feb. 2005.
- [3]. Yen-Lin Chen, "Automatic Text Extraction, Removal and Inpainting of Complex Document Images," *International Journal of Innovative Computing, Information and Control*, Vol. 8, No.1(A), pp. 303-327, January 2012.
- [4]. A. Mosleh, N. Boouguila, and A. B. Hamza, "An Automatic Inpainting Scheme for Video Text Detection and Removal," *IEEE Transactions on Image processing*, vol. 22, no. 11, pp.4460–4472, Nov. 2013.
- [5]. Mohammad Khodadadi and AlirezaBehrad, "Text Localization, Extraction and Inpainting in color images", *ICEE2012*, Vol.12, IEEE, May 2012, pp. 1035-1040.
- [6]. M. Cai, J. Song, and M. Lyu, "A new approach for video text detection," *IEEE International Conference on Image Processing (ICIP)*, 2002, pp. 1–117–1–120.
- [7]. B. Epshtein, E. Ofek, and Y. Wexler, "Detecting text in natural scenes with stroke width transform," *IEEE conference on Computer Vision and Pattern Recognition(CVPR)*, 2010, pp. 2963–2970, June 2010.
- [8]. Leon, M., Vilaplana, V., Gasull, A. and Marques, F., "Caption text extraction for indexing purposes using hierarchical region-based image model," *IEEE ICIP2009*, El Cairo, Egypt, 2009.
- [9]. N. Ezaki, M. Bulacu, L. Schomaker, "Text Detection from Natural Scene Images: Towards a System for Visually Impaired Persons", *Int. Conf. on Pattern Recognition (ICPR 2004)*, vol. II, pp. 683-686.
- [10]. Dong wookcho and Tien D. Bui "Image In painting Using Wavelet-Based Inter and Intra-Scale Dependency" *IEEE Transactions on Image Processing*, 2008.