Available Online at www.ijcsmc.com

International Journal of Computer Science and Mobile Computing

A Monthly Journal of Computer Science and Information Technology

ISSN 2320–088X

IJCSMC, Vol. 3, Issue. 4, April 2014, pg.1047 – 1052

RESEARCH ARTICLE

A Dynamic Approach to Extract Texts and Captions from Videos

R.Bhavadharani[1], P.M. Sowmya[2], A.Thilagavathy[3]

[1,2]PG Scholar, Department or Computer Science, R. M. K Engineering College, Chennai, India
[3]Assistant Professor, Department of Computer Science, R. M. K Engineering College, Chennai, India
bhavadharaniravi@yahoo.com [1], sowmaha@gmail.com [2], thilsmailbox@gmail.com [3]

Abstract— Detecting text and caption from videos is of great importance and is of greater demand for video retrieval, annotation, indexing, and content analysis etc. In this paper, we have chosen 2 approaches to treat the text extraction problem, to detect text and caption from videos. Each of both uses the characteristics of artificial text. In connected regions approach we use split and merge algorithm in order to split non homogeneous regions and to merge adjacent homogeneous regions. Using edge detection, edges will be computed using certain threshold. Then each text area will be binarized. Then the region with a possibility of text in the video is detected. The most distinctive feature of the proposed operator over most other operators is that it offers excellent line, edge, detail, and texture preservation performance while, at the same time, effectively removing noise from the input image. Finally, the same process will be executed for the different frames, and the results will be temporally integrated in order to keep only the elements which are present in all the frames.

Keywords: binarized, content analysis, connected regions, artificial text, threshold

I. INTRODUCTION

Recent studies in the field of computer shows a great amount of interest in content retrieval from images and videos [1, 2, 3]. This content can be in the form of objects, colors, textures, shapes as well as relation between them. During the past decade, a lot of techniques have been put forward to extract the semantic and content information from image and videos. During the past decade, a lot of techniques are available to extract the text from an image. This includes *texture based methods, connected component based methods* and *edge based methods* [4, 5]. Among these techniques, text detection based approaches are of greater hike due to the precise information in the text. Texture based methods [6, 7, 8] are used to distinguish text regions from their background and or other regions within the image.

In Connected Component based methods, the image is divided into a set of smaller components known as connected components and then it [9, 10] recursively merge the small components to form larger ones. It usually scans an image and groups its pixels into components based on pixel connectivity *i.e.* all pixels that lies within a selected connected component share similar pixel intensity values and are in some way connected to each other. Edge based methods [11, 12, 13] are used to define the boundaries between regions in an image, which helps with segmentation and object recognition. Edge detection

usually refers to the process of identifying and locating shared discontinuities in an image. Other method for text region detection includes techniques such as support vector machines [14, 15], K-means clustering [16] and neural networks [17] etc

II. Related Search

Xu Zhao et al [18] proposed a technique called corner based approach to detect text and caption from videos. This method is based on the fact that there exists dense and orderly presences of corner points in a text Palaiahnakote Shivakumar et al [19] proposed a method based on Laplacian approach in order to detect text from video. Usually a text is horizontally oriented whereas this method is able to handle text from any orientation. Later K-means clustering is performed over the text

Xiaoqian Liu et al [20] proposed a technique to extract captions from videos. They have presented a novel stroke like edge detection method along with the highlights of temporal feature in extracting texts

III. Overall Approach

A video file is given as input. Basically an input is a set of frames extracted from a video file. First the image containing a set of frames is converted into a grayscale image.

Fig I (a) Original Image (b) Gray scaled image

Then a 3x3 median filter is applied to blur the background of the provided image. After applying this filter the characters are slightly blurred but the remaining areas are strongly blurred than the text regions. This approach utilizes the fact that usually the color data in text characters is different from the color data in the background.

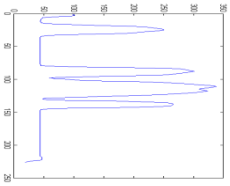
Fig II (a)Gray scaled image (b) After applying median filter

Now our next step is to compute the edges from the computed image. In order to detect its edges we have choosen sobel filter for this purpose. Using a Sobel vertical edge emphasizing filter, the edges of the image will be computed.

Fig III (a) Blurred image (b) Filtered image

Next is to detect the text area and compute its horizontal projection. For this, we have applied a 6-by-6 maximum filter and then evaluated its horizontal projection histogram.

Fig IV (a) Filtered image (b)Detected text



(c) Horizontal Histogram

This image is then scanned. If it has a projection value less than the certain threshold then that line is excluded. Now we have to Detect edges on every remaining text area and exclude them correctly. For this, we have applied Sobel horizontal edge emphasizing filter for each possible text area. And then we have computed their histogram to exclude them properly. Finally we have to convert our segmented image to binary image by using a luminance threshold.

Fig V (a)Vertical Exclusion (b) Final image after segmentation

IV. TEXT EXTRACTION APPROACHES

A. CONNECTED REGIONS APPROACH

The main idea of this approach is that a letter can be considered as a homogeneous region (using our restrictions), and thus it could be very useful to divide the frame into homogeneous regions. To compute such a division, a split-and-merge algorithm seems to be very adequate. Its concept is: while there is a non-homogeneous region, then split it into four regions. And if two adjacent regions are homogeneous, then they can be merged. Then, using some size characterizations of the text (not too big and not too small), the inadequate regions will be deleted. The same process will be executed for the different frames, and the results will be temporally integrated in order to keep only the elements which are present in all the frames.

Fig VI Connected regions approach

B. EDGE DETECTION APPROACH

The main idea of this approach is that text contrasts a lot with background. Thus, using edge detection concepts seems to be a good idea. After the edges will be computed, the number of edges in x and y direction will be calculated and if it is higher than a certain threshold then it will be considered as a text area. Then each text area will be binarized using the luminance. So in a final result, the text will be in white and the background in black (or the inverse). Finally, the same process will be executed for the different frames, and the results will be temporally integrated in order to keep only the elements which are present in all the frames

(a)Text region detection

(b) Binary image by thresholding

(c) Integrated Frame

Fig VII Edge Detection approach

C. TEMPORAL INTEGRATION

In some consecutive video images, characters do not change and background does. In this case, apply text segmentation on consecutive video images, then use temporal integration to eliminate false positive (false positive is not a text area but resulted as a possible text area). For each of them, the temporal integration is actually a basic AND operator.

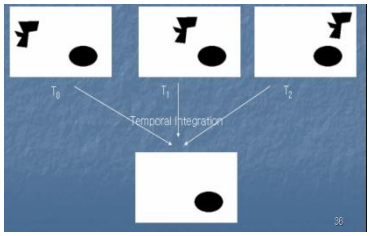


Fig VIII Temporal Integration

V. ALGORITHM FOR TEXT EXTRACTION

- An algorithm for edge based text extraction consists of the following basic steps:
- First, the image has to be converted in gray scale.
- Apply 3-by-3 median filter to blur background.
- Compute the edges using a Sobel filter.
- Detect the text area and compute its horizontal projection.
- Scan the horizontal projection histogram.
- Detect edges on every remaining text area and exclude them correctly.
- Binarize the image.

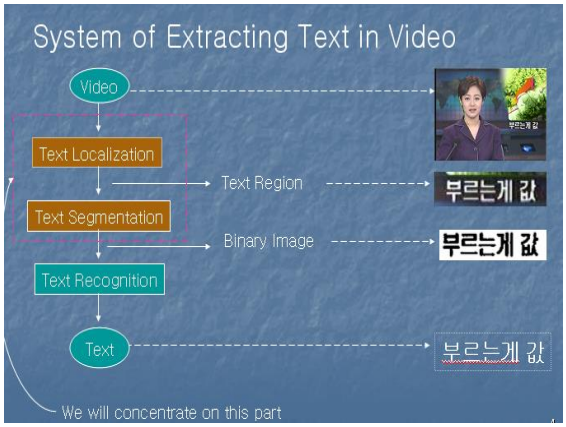


Fig IX Block Diagram For Text Extraction

VI. FUTURE ENHANCEMENTS

Our future aim is to test the input images on various factors such as scaling and lightning conditions such that it would be invariant to scale, lightning as well as orientation changes. Next is to implement an enhanced morphological cleaning of images which could result in a higher precision rate. Lastly, an interesting test would be to find out the text regions which are hidden behind other objects or the texts which are watermarked within an image.

VII. CONCLUSION

In this paper, we present a new way to detect texts that is embedded in a video. This proposed method is more effective and efficient in extracting the text than the contemporary methods. The results obtained by this method are more precise and encouraging. However, the output could be improved for some more complex video file. A future work could be to improve our algorithm such as it works for various types of videos. Then, the future algorithm could deal with color segmentation (instead of Gray segmentation), removing noise in text regions and deleting false regions.

REFERENCES

[1] Y. Rui, T. S. Huang, and S. F. Chang, "Image retrieval: Current techniques, promising directions, and open issues," *J. Vis. Commun. Image Represent.*, vol. 10, no. 1, pp. 39–62, 1999.

[2] Y. A. Aslandogan and C. T. Yu, "Techniques and systems for image and video retrieval," *IEEE Trans. Knowl. Data Eng.*, vol. 11, no. 1, pp. 56–63, Jan./Feb. 1999.

[3] W. M. Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain, "Content-based image retrieval at the end of the early years," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 12, pp. 1349–1380, Dec. 2000

[4] X. Tang, X. Gao, J. Liu, and H. Zhang, "A spatial-temporal approach for video caption detection and recognition," *IEEE Trans. Neural Netw.*, vol. 13, no. 4, pp. 961–971, Jul. 2002

[5] K. Jung, K. I. Kim, and A. K. Jain, "Text information extraction in images and video: A survey," *Pattern Recognit.*, vol. 37, no. 5, pp.977–997, 2004

[6] K. Kim, K. Jung, and J. Kim, "Texture-based approach for text detection in images using support vector machines and continuously adaptive mean shift algorithm," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 12, pp. 1631–1639, Dec. 2003.

[7] Y. Zhong, H. Zhang, and A. K. Jain, "Automatic caption localization in compressed video," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 4, pp. 385–392, Apr. 2000
[8] W. Mao, F. Chung, K. K. M. Lam, and W. Sun, "Hybrid chinese/English text detection in images and video frames," in *Proc. 16th Int. Conf. Pattern Recognit.*, 2002, vol. 3, pp. 1015–1018.

[9] E.K. Wong and M. Chen, "A New Robust Algorithm for Video Text Extraction," Pattern Recognition, vol. 36, pp. 1397-1406, 2003.

[10] V.Y. Mariano and R. Kasturi, "Locating Uniform-Colored Text in Video Frames," Proc. Int'l Conf. Pattern Recognition, pp. 539-542, 2000

[11] Abdullah A. Alshennawy and Ayman A. Aly. "Edge Detection in Digital Images Using Fuzzy Logic Technique," World Academy of Science, Engineering and Technology51 2009. *DOI* - 10.1.1.193.3230.

[12] Yasar Becerikli, Tayfun M. Karan and Ali Okatan. "A New Fuzzy Edge Detector For Noisy Images Using Modified WFM Filter," International Journal of Innovative Computing, Information and Control Volume 5, Number 6, June 2009 ISSN 1349-419.

[13] 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.

[14] Qixiang Ye, Qingming Huang, Wen Gao and Debin Zhao, "Fast and Robust text detection in images and video frames," Image and Vision Computing 23, 2005.

[15] Qixiang Ye, Wen Gao, Weiqiang Wang and Wei Zeng, "A Robust Text Detection Algorithm in Images and Video Frames," IEEE, 2003.

[16] Victor Wu, Raghavan Manmatha, and Edward M. Riseman, "TextFinder: An Automatic System to Detect and Recognize Text in Images," IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 21, No. 11, November 1999.

[17] Rainer Lienhart and Axel Wernicke, "Localizing and Segmenting Text in Images and Videos," IEEE Transactions on Circuits and Systems for Video Technology, Vol.12,
No.4, April 2002.

[18] Xu Zhao, Kai-Hsiang Lin, Yun Fu, Yuxiao Hu, Yuncai Liu and Thomas S. Huang, "Text From Corners: A Novel Approach to Detect Text and Caption in Videos" IEEE Transactions on Image Processing, VOL. 20, NO. 3, March 2011.

[19] Palaiahnakote Shivakumara, Trung Quy Phan, and Chew Lim Tan, "A Laplacian Approach to Multi-Oriented Text Detection in Video", IEEE Transactions on Pattern Analysis and Machine IntelligenceE, VOL. 33, NO. 2, February 2011

[20] Xiaoqian Liu and Weiqiang Wang, "Robustly Extracting Captions in Videos Based on
Stroke-Like Edges and Spatio-Temporal Analysis", IEEE TRANSACTIONS ON MULTIMEDIA, VOL. 14, NO. 2, APRIL 2012.

