ISSN: 2278 – 909X
International Journal of Advanced Research in Electronics and Communication Engineering (IJARECE)
Volume 4, Issue 7, July 2015

Text Based Video Indexing and Retrieval by Using DLER Technique

Sneha S. Kapse, Prof. Pravin Kshirsagar

***Abstract*— This project extract the text in the video that means it extract the text in image. Because video comprises of number of images. In order to extract and search important information in huge video clip we are focusing on extracting the text from video. Here First we convert the video into frames or images then we choose area of interest in which we extract the text called region of interest and then continue with algorithm for localization and recognition.**

***Index Terms*—Text Extraction, Text Localization, Text Recognition, Text Segmentation, Video Text.**

I. INTRODUCTION

Due to increase of available network, many user use the videos from large video site like YouTube etc. For example, in YouTube, over one day, new videos are uploaded to the site in every minute or second. So it is difficult to manually index and retrieve the large video. It is also difficult to search a small portion or text in large video. So we extract the video in form of image and then in this image we can retrieve the text.

Text appearing in image and videos can be categorized into two main groups.

1.1 Artificial Text :

Artificial text can be laid over the image. It is also called as caption or superimposed text. It is added mechanically in text.

1.2 Scene Text:

Scene text is the video texts observe in real word object. Scene text exist naturally and appears accidentally which is captured by recording device e.g. Street sign, text on vehicle, logos and text on shirts of players etc.

The video contain the text, including scrolling text or caption text (superimposed text) and scene text embedded in background. In the first part we introduce the concept of text extraction from image or video as well as text base image or video retrieval. In the next part we discuss the text extraction from image and video by using process such as text detection, text localization, text extraction and optical character recognition (OCR).

Fig.1. Scene Text

1.3 Text Detection:

In this text detection stage, the text of input image need to be identified as a input image contain any text, the existence or non existence of text among the image. However in case of video, the amount of frame containing text is smaller than amount of frame while not in text. The text detection stage detects the text in image.

1.4 Text Localization:

Text localization stage included localization of text in image after detection. In other words, text present in the frame, was tracked by identifying boxes or region of smaller pixels intensity value and returning them to the next stage for further processing.

1.5 Text Extraction:

The text extraction stage discuss the text tracking and extraction.After the text was localised, text tracking step deals with seperation of text pixels from background pixels. The output of this step is a binary image where black text character appears on a white background. This stage included extraction of actual text region by diving pixels with similar properties into segment.

1.6 Text Recognition:

Text recognition performing OCR on the binaries text in image.

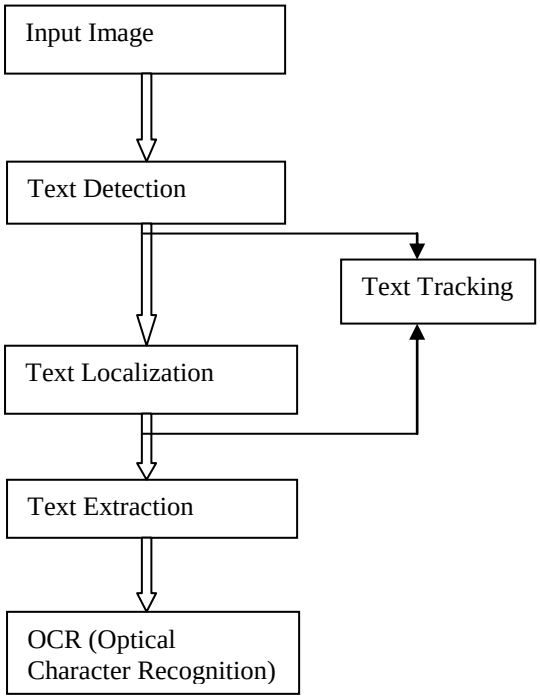

Fig.2. Text Extraction from image /video Architecture

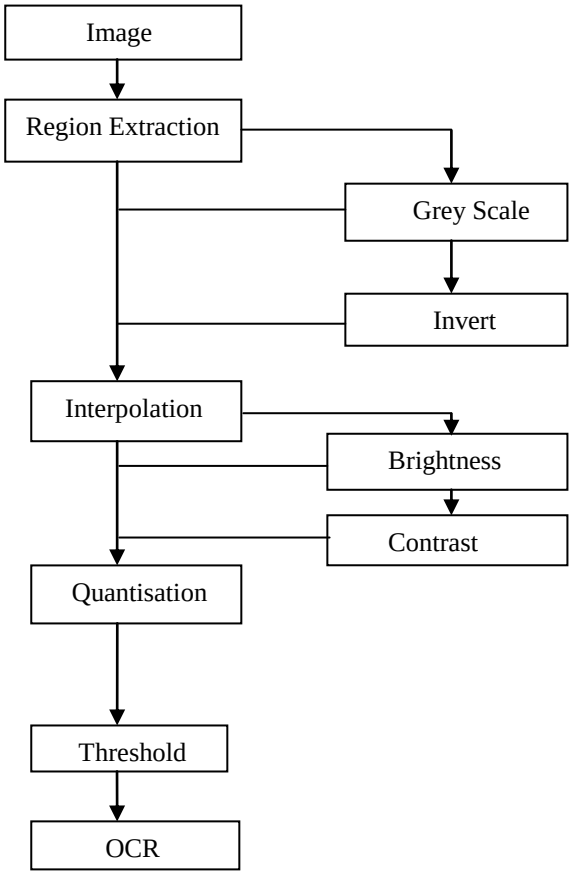

Fig.3 Image processing and OCR process.

II. **Proposed Method: A Practical Implementation of DLER Tool:**

The primary aim of this paper is to improve the techniques for video text detection, localization, extraction and recognition. The text extraction in video frames is difficult because of complex background, unknown text character color and various stroke widths. Although many methods have been propose for preprocessing, we propose a fully automatic method, a simple approach for preprocessing which integrate all the steps involved in detection, localization, extraction and recognition as a simple and single tool as shown in fig. The DLER tool is customer friendly and integrates all preliminary steps into a single tool. The frame work as shown in fig.

1. Image extraction from video object.
2. Identifying the candidate region.
3. Separation of alphabets or letters.
4. Applying brightness and contrast.
5. Quantized the image.

2.1 Invert:

The OCR system to recognize character efficiently the character of the video text must be in black and remaining pixels in white.

2.2 Interpolation:

After inversion or quantized, we have to used interpolation where T is threshold. This equation gives the 0 and 1 as output values giving a true binary image.

2.3 Threshold:

An expression for brightness and contrast modification of image is

$$G(x, y) = a+ (x, y) + b$$

Where, a is gain and b is a bias image quality can be improved using linear mapping where we map a particular range of gray levels [f1, f2] onto a new range [g1,g2], this is increases the gain factor until two adjacent levels f1 and f2 are mapped onto 0, whereas gray levels greater than f1 are mapped onto 255 where f1 acts as a threshold. And the mapping operation is termed as a thresholding.

Image thresholding is a segmentation technique because it classify the pixels into two catagories, those at which some property measures from image falls below the threshold and those at which that property equal to exceed the threshold. Because there are two possible output values, thresholding creats a binary image. The common form of image thresholding makes use of pixel gray level. Gray level thresholding applies to every pixel the rule is,

$$G(x, y) = 0, \quad f(x, y) < T$$
$$= 1, \quad f(x, y) > T$$

2.4 Experimental Result:

1. ORIGINAL IMAGE 2.GRAYSCALE IMAGE 3.INTERPOLATION 4. QUANTIZATION 5. INVERT 6. THRESHOLDING 7. OCR.

III. CONCLUSION:

In this paper we are provided the comprehensive literature of text extraction involved detection localization, tracking , extraction and recofnition from a given image. We can concluded that the text appearing in video can be extracted using this method.

ACKNOWLEDGMENT

A single person cannot work directly or indirectly, somebody needs to help you. Before we get into thick of things we would like to add few heartfelt words for all the Teachers of Electronics Engineering Department who took the pains to make our project successful.

We would like to express sincere thanks to Prof. Pravin Kshirsagar for their guidance, cooperation which helped us to successful completion of our project. We are also heartily thankful to our friends for their kind cooperation who have directly or indirectly helped us to complete the project successfully.

REFERENCES

[1] C.V. Jawahar, Balakrishna Chennupati, Balamanohar Paluri, Nataraj Jammalamadaka, "Video Retrieval Based on Textual Queries", IEEE Transaction on circuits and System For Video Technology, Vol. 9, No. 8, December 1999.

[2] Rainer Lienhart and Axel Wernicke, "Localizing and Segmenting Text in Images and Videos", IEEE Transaction on circuits and System For Video Technology, Vol. 12, No. 4, April 1999.

[3] Jayshree Ghorpade, Raviraj Palvankar, Ajinkya Patankar and Snehal Rathi, "Extracting Text from Video", Signal & Image Processing: An International Journal (SIPIJ) Vol.2, No.2, June 2011.

[4] Dr. Sunitha Abburu, "Multi Level Semantic Extraction for Cricket Video by Text Processing", International Journal of Engineering Science and Technology, Vol. 2(10), 2010.

[5] Michael R. Lyu, Jiqiang Song and Min Cai, "A Comprehensive Method for Multilingual Video Text Detection, Localization, and Extraction", IEEE Transaction on circuits and System For Video Technology, Vol. 15, No. 2, February 2005.

[6] Er. Navdeep Kaur and Dr. Mandeep Singh, "Content Based Video Retrieval Using Color, Texture and Time Based", International Journal of Applied Engineering and Technology , Vol. 4 (2), April-June 2014.

[7] H. Tran, A lux, H.L. Nguyen T. and A. Boucher, "A novel approach for text detection in images using structural features", The 3rd International Conference on Advances in Pattern Recognition, LNCS Vol. 3686, pp. 627-635, 2005.

[8] X. Liu, H. Fu and Y. Jia, "Gaussian Mixture Modeling and learning of Neighbor Characters for Multilingual Text Extraction in Images", Pattern Recognition, Vol. 41, pp. 484-493, 2008.

[9] P. Dubey, "Edge Based Text Detection for Multi-purpose Application", Proceedings of International Conference Signal Processing, IEEE, Vol. 4, 2006.

[10] D. Crandall, S. Antani, R. Kasturi, " Extraction of special effects caption text events from digital video", International Journal on Document Analysis and Recognition, Vol. 5, pp. 138-157, 2003.