ISSN 2395-695X (Print)
ISSN 2395-695X (Online)
International Journal of Advanced Research in Biology Engineering Science and Technology (IJARBEST)
Vol. 2, Issue 5 May 2016

Multi-Spectral Fusion for Laplacian -Wavelet based Scene Text Detection in Video Images

J.Anto Terancy [1] ,Mr.S.R.Surem Samuel [2] ,

[1] *PG Student, Applied Electronic, C.S.I Institute of Technology,*

[2]*Assistant Professor, Department of Electronics and Communication, C.S.I Institute of Technology,*

[1] *antoterancy1992@gmail.com*

Abstract: Scene text detection from video as well as natural scene images is challenging due to the variations in background, contrast, text type, font type, font size, and so on. Besides, arbitrary orientations of texts with multi-scripts add more complexity to the problem. The proposed approach introduces a new idea of convolving Laplacian with wavelet sub-bands at different levels in the frequency domain for enhancing low resolution text pixels. Then, the results obtained from different sub-bands (spectral) are fused for detecting candidate text pixels. We explore maxima stable extreme regions along with stroke width transform for detecting candidate text regions. Text alignment is done based on the distance between the nearest neighbor clusters of candidate text regions. In addition, the approach presents a new symmetry driven nearest neighbor for restoring full text lines. We conduct experiments on our collected video data as well as several benchmark data sets, such as ICDAR 2011, ICDAR 2013, and MSRA-TD500 to evaluate the proposed method. The proposed approach is compared with the state-of-the-art methods to show its superiority to the existing methods.

Index Terms: Laplacian wavelet, multi spectral fusion, maxima stable extreme regions, stroke width transform, arbitrarily oriented video text detection.

1. INTRODUCTION

The conventional approaches that use low level features may not be sufficient for handling such large databases due to the gaps between low level features and high level semantics. To alleviate this problem, text detection and recognition has become popular as it provides meaningful cues which are close to the content of video or image. So, it has been widely used in video summarization, content based image indexing and video sequence retrieval. On top of these applications, text detection and recognition has also been used for real time surveillance applications, such as assisting a blind person to walk freely on roads, assisting tourists to reach their destinations, enhancing safe driving, navigating vehicles based on license plate information, exciting event extraction from sports video, identifying athletes in marathon events

Video consists of two types of texts, namely, caption text and scene text. Caption text is manually edited, which has good clarity and visibility and hence is easy to process. Scene text exists naturally in video frames, the detection of which suffers from color bleeding, low contrasts, low quality due to distortion, different orientations, backgrounds, etc. Hence, scene text is hard to process compared to caption text.

ISSN 2395-695X (Print)
ISSN 2395-695X (Online)
International Journal of Advanced Research in Biology Engineering Science and Technology (IJARBEST)
Vol. 2, Issue 5 May 2016

Scene images captured through a high resolution camera usually contain only scene texts with high contrast and complex background, while video contains both caption and scene texts with low resolution and complex background. Achieving good accuracy for text detection from both video and natural scene images is still an open issue in the field of image processing and pattern recognition because most of the existing approaches either focus on caption text in video or scene text in natural scene images but not both video and natural images.

The problem of text detection and recognition from scanned document images is not new for the document analysis community because for different scripts we can find several Optical Character Recognizers (OCR engines) that are available publicly. However, the same methods may not be used for detection and recognition of the texts in video and natural scene images because the approaches work well for plane background and high contrast images but not for images like video and natural scene images.

To widen the scope of document analysis based approaches, there are methods proposed for text detection from natural scene images. These approaches directly or indirectly rely on the features of connected components and the shapes of characters to achieve a good accuracy. This is valid because text in natural scene images usually has high contrast as mentioned above and hence the shape of a character can be preserved in most of the situations. However, this is not necessarily true for video, where we can expect disconnections, loss of information, distorted shapes and so on due to low resolution and low contrast. Therefore, the approaches developed for text detection in natural scene images may not be used directly for text detection in video.

Text characters embedded in images and a video sequence represents a rich source of information for content-based indexing and retrieval applications. However, these text characters are difficult to be detected and recognized due to their various sizes, grayscale values and complex backgrounds. This thesis investigates methods for building an efficient application system for detecting and recognizing text of any grayscale values embedded in images and video sequences. Both empirical image processing methods and statistical machine learning and modeling approaches are studied in two sub-problems: text detection and text recognition. Applying machine learning methods for text detection encounters difficulties due to character size, grayscale variations and heavy computation cost.

2. EXISTING SYSTEM

Connected component based approaches are fast and good for images that have high contrast texts and plain background just like methods in the document analysis field. On the other hand, these approaches will not be suitable for text detection in video and natural scene images due to low video resolution and natural scene complexity. To improve the performance of text detection, edge and gradient based approaches are developed. These approaches are good at recall but poor at precision because the proposed features are sensitive to background complexity leading to more false positives.

Texture based approaches are developed to solve the problems of edge and gradient

ISSN 2395-695X (Print)
ISSN 2395-695X (Online)
International Journal of Advanced Research in Biology Engineering Science and Technology (IJARBEST)
Vol. 2, Issue 5 May 2016

based approaches because the texture property works well for complex background. Here, the approaches define the appearance of text pattern as a special texture.

The color and spatial relationship of the structure of text candidates are used for text candidates merging. SVM classifier is proposed for false positive elimination. The method uses language specific features for text detection.

The approach fuses color, gradient and log-Gabor images to enhance text information in frames. Then it proposes character segmentation based on vertical profiles before text extraction. However, this approach is limited to horizontal text but not arbitrarily oriented text of different scripts.

3. PROPOSED METHOD

Video text extraction using fusion of color gradient and log-Gabor filter. The approach fuses color, gradient and log-Gabor images to enhance text information in frames. Then it proposes character segmentation based on vertical profiles before text extraction. MSER properties, edge and gradient analysis like stroke width information, and texture analysis like the combination of Laplacian with wavelet sub-bands of text components, for text detection in video and natural scene images.

To detect low contrast and low resolution texts, the proposed approach sometimes fails to detect texts properly if an image contains too small fonts, low resolution and too complex background.

The proposed approach consists of four steps. In the first step, as we are inspired by the work presented for multi-oriented video text detection using the combination of Laplacian and Fourier, we propose a novel idea of convolving Laplacian with wavelet sub-bands at different levels in the frequency domain to enhance text pixels through a fusion concept, which results in multi-spectral fusion. The fused images are subjected to fuzzy k-means clustering to classify the Candidate Text Pixels (CTP). Since video and natural images are complex and text has large variations in font, font size, color, orientation, etc., conventional methods such as the non-fuzzy k-means clustering, the Max-Min clustering method and the adaptive thresholding technique do not work well. The reason is that the conventional k-means clustering produces inconsistent results because of random guess selection for clustering, the Max-Min clustering method is not accurate enough due to the lack of discriminations between text and non-text values, and the adaptive thresholding technique does not give good results because it is hard to decide threshold values dynamically for different situations.

Therefore, we prefer to use fuzzy K-means clustering, which classifies text and non-text pixels based on the probability of either text or non-text pixels with a membership function but not direct values. It is noticed that Laplacian helps in distinguishing text pixels as it gives high positive and negative values for text pixels and low values for non-text pixels. In the same way, Fourier in the frequency domain provides high coefficients for text pixels and low coefficients for non-text pixels. Therefore in this work, we propose to combine Laplacian with wavelet sub-bands for better enhancement because wavelet sub-bands have both low and high pass filters, while Fourier behaves either as a low pass or a high pass filter to eliminate noisy pixels in text detection. To tackle the problems of multi-size and multi-contrast texts;

ISSN 2395-695X (Print)
ISSN 2395-695X (Online)
International Journal of Advanced Research in Biology Engineering Science and Technology (IJARBEST)
Vol. 2, Issue 5 May 2016

we combine Laplacian with wavelet sub-bands at different levels through fusion.

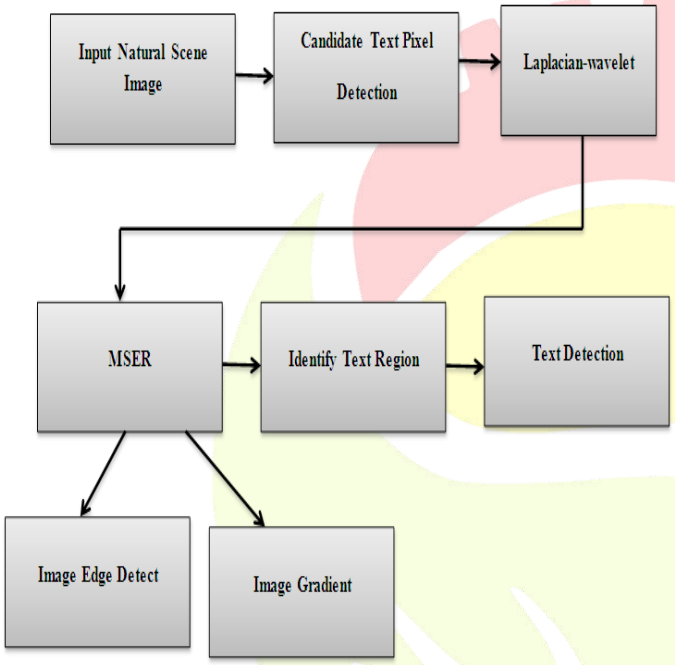

Figure: Block diagram of proposed method

Input Natural Scene Image

An observed scene from a continuous range of target viewpoints. Toward this end, that processes a set of input images to produce photorealistic scene re projections over a wide range of viewpoints.

Contrast Enhancement Image:

The equation that guides the sensitivity of the human eye to brightness differences at different intensities. Contrast detection has been studied in vision perception literature for decades. Threshold contrast sensitivity functions (CSF) de- fine the minimum contrast required to detect a sinusoidal grating of a particular mean and spatial frequency

Gaussian Filter

A Gaussian filter is a filter whose impulse response is a Gaussian function (or an approximation to it). Gaussian filters have the properties of having no overshoot to a step function input while minimizing the rise and fall time. This behavior is closely connected to the fact that the Gaussian filter has the minimum possible group delay.

Latent Image

The others MSER offers the most variety detecting about 2600 regions for a textured blur scene and 230 for a light changed. scene, and variety is generally considered to be good. Also MSER had a repeatability of 92% for this test.

Text Detection

Text detection by exploring temporal Image Scene. It combines the characters and links energies to compute text unit energy as a measure of the likelihood of the candidate being a text object. Grouping merges two text lines as one text line due to the variations in the spacing between two text lines.

4. SIMULATION RESULT:

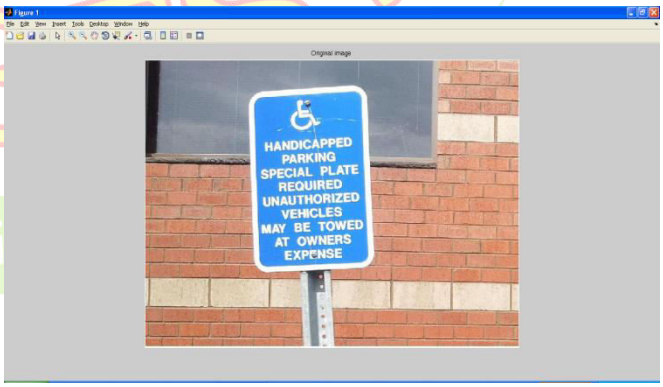

Fig: Input natural scene image

ISSN 2395-695X (Print)
ISSN 2395-695X (Online)
International Journal of Advanced Research in Biology Engineering Science and Technology (IJARBEST)
Vol. 2, Issue 5 May 2016

Fig:Detect Text Region Image

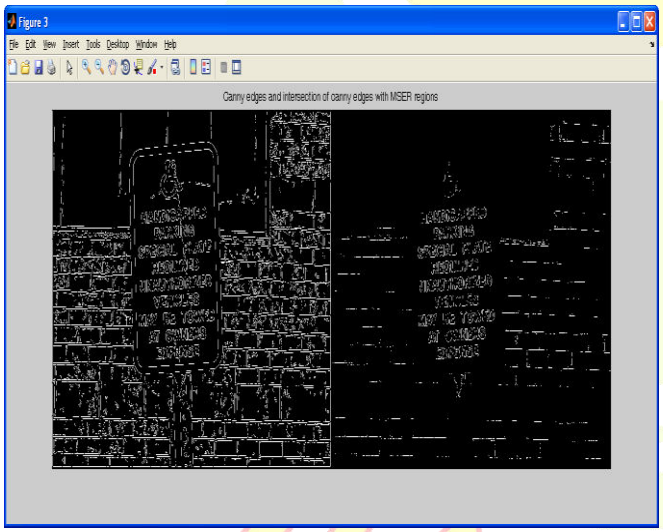

Fig: Edge Detect MSER Region Image

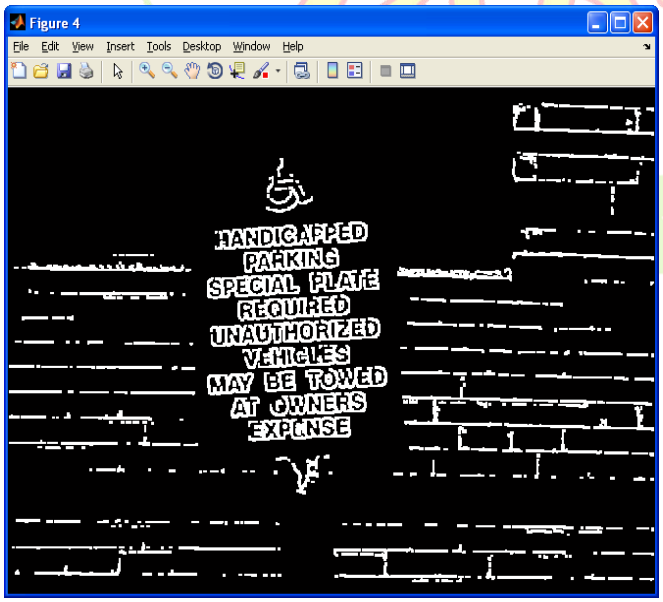

Fig: Text Region Detection

Fig: Original MSER Region and Segmented MSER Region

Fig: Text Candidate Before and AfterRegionFiltering

ISSN 2395-695X (Print)
ISSN 2395-695X (Online)
International Journal of Advanced Research in Biology Engineering Science and Technology (IJARBEST)
Vol. 2, Issue 5 May 2016

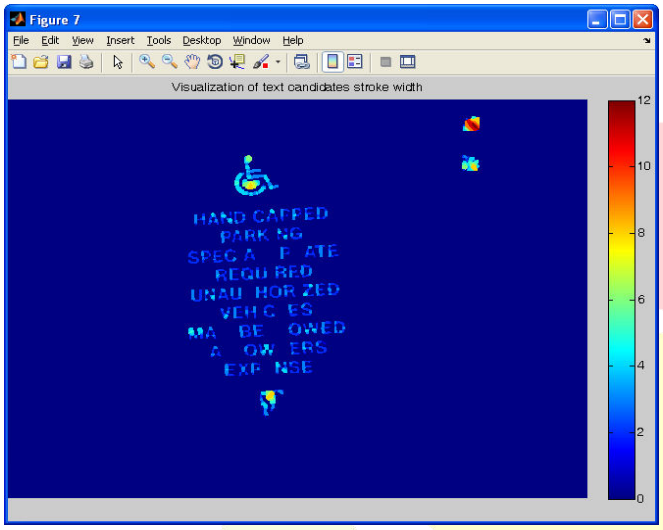

Fig:Visualization Of Text Candidate Stroke Width

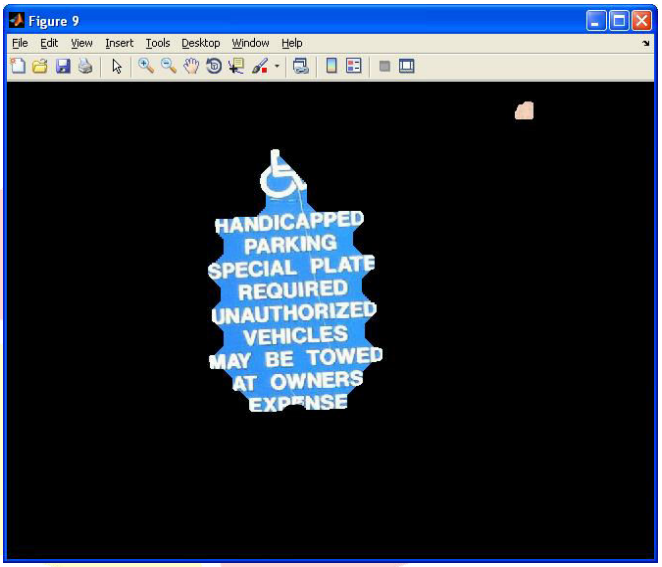

Fig: Text Region Detection

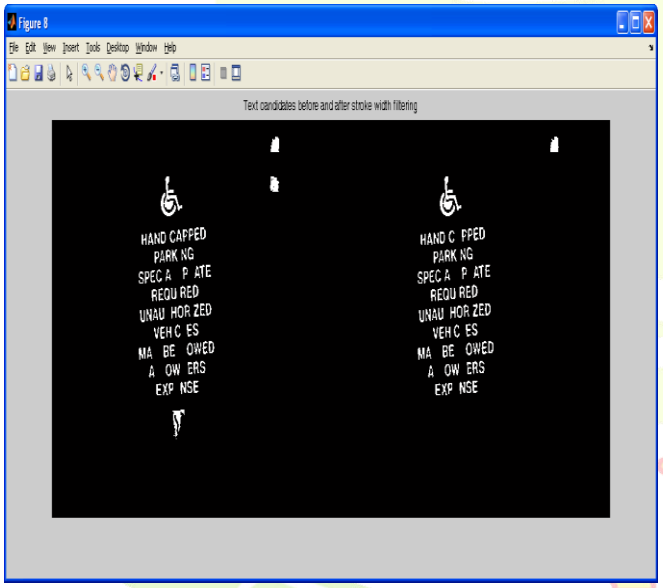

Fig: Text Candidate Filtering

Fig:Text Detection

5. CONCLUSION:

A novel idea of combining Laplacian with wavelet high frequency sub-bands through fusion at multi-level to identify text candidates. I have explored Maximally Stable Extremal Regions along with stroke width distances for preserving fine details of text candidates. The proposed approach introduces mutual nearest neighbor clustering based on geometrical properties of text candidates to group text candidates of respective

text lines into clusters. The symmetry driven growing process is proposed to extract arbitrary text lines based on the distance between text candidates in each cluster. According to my knowledge, this is the first attempt to detect text in both video frames and natural scene images with good accuracies. However, according to the results, the accuracy is still lower than that in document analysis, where usually the value would be more than 80%.

REFERENCES

1. D. Chen and J.-M. Odobez, "Video text recognition using sequential Monte and error voting methods," Pattern Recognit. Lett., vol. 26, no. 9, pp. Carlo 1386–1403, Jul. 2005
2. 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, May 2004.
3. J. Liang, D. Doermann, and H. Li, "Camera-based analysis of text and documents: A survey," in Proc. IJDAR, 2005, pp. 84–104.
4. P. Shivakumara, A. Dutta, C. L. Tan, and U. Pal, "Multi-oriented scene text detection in video based on wavelet and angle projection boundary growing," in Multimedia Tools and Applications. New York, NY, USA: Springer-Verlag, 2013.
5. P. Shivakumara, T. Q. Phan, and C. L. Tan, "A Laplacian approach to multi-oriented text detection in video," IEEE Trans. Pattern Anal. Mach. Intell., vol. 33, no. 2, pp. 412–419, Feb. 2011.
6. J. Yan and X. Gao, "Detection and recognition of text superimposed in images base on layered method," Neurocomputing, vol. 134, pp. 3–14, Jun. 2014.
7. C. Yao, X. Bai, and W. Liu, "A unified framework for multioriented text detection and recognition," IEEE Trans. Image. Process., vol. 23, no. 1, pp. 4737–4749, Nov. 2014
8. X.-C. Yin, Z. Yin, K. Huang, and H.-W. Hao, "Robust text detection in natural scene images," IEEE Trans. Pattern Anal. Mach. Intell., vol. 36, no. 5, pp. 970–983, May 2014.
9. J. Zhang and R. Kasturi, "A novel text detection system based on character and link energies," IEEE Trans. Image Process., vol. 23, no. 9, pp. 4187–4198, Sep. 2014.
10. J. Zhang and R. Kasturi, "Extraction of text objects in video documents: Recent progress,"in proc DAS,2008