IJMTARC – VOLUME – IV – ISSUE - 16 - DEC 2016 ISSN: 2320-1363

# A STRAIGHT TECHNIQUE TOWARDS DEVELOPING LOW-LEVEL TEXT PIXELS

[1]BODDU PRASHANTHI, [2]Mr.P.V.Vara Prasad Rao

[1]PG Scholar, Department of ECE, SLC's Institute of Engineering and Technology, Piglipur Village, Hayathnagar Mandal, Near Ramoji Film City, Ranga Reddy District, Hyderabad, Telangana

[2]Assosciate Professor, Department of ECE, SLC's Institute of Engineering and Technology, Piglipur Village, Hayathnagar Mandal, Near Ramoji Film City, Ranga Reddy District, Hyderabad, Telangana

**ABSTRACT:**

The suggested approach is in contrast to the condition-of-the-art techniques to show its brilliance towards the existing methods. Besides, arbitrary orientations of texts with multi-scripts increase the complexity towards the problem. The suggested approach introduces a brand new concept of convolving Laplacian with wavelet sub-bands at different levels within the frequency domain for enhancing low resolution text pixels. Then, the outcomes acquired from various sub-bands (spectral) are fused for discovering candidate text pixels. Scene text recognition from video in addition to natural scene images is challenging because of the variations in background, contrast, text type, font type, font size, and so forth. We conduct experiments on the collected video data in addition to several benchmark data sets, for example ICDAR 2011, ICDAR 2013, and MSRA-TD500 to judge the suggested method. Text alignment is completed in line with the distance between your nearest neighbor clusters of candidate text regions. Additionally, the approach presents a brand new symmetry driven nearest neighbor for restoring full text lines. We explore maxima stable extreme regions together with stroke width transform for discovering candidate text regions.

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

## 1. INTRODUCTION:

The traditional approaches which use low-level features might not be sufficient to handle such large databases because of the

IJMTARC – VOLUME – IV – ISSUE - 16 - DEC 2016 ISSN: 2320-1363

gaps between low-level features and level semantics. Caption text is by hand edited, that has good clearness and visibility and therefore is simple to process. Scene text exists naturally in video frames, the recognition which is affected with color bleeding, low contrasts, poor because of distortion, different orientations, backgrounds, etc. To ease this issue, text recognition and recognition is becoming common as it offers significant cues that are near to the content of video or image. Video includes two kinds of texts, namely, caption text and scene text [1]. Hence, scene text is difficult to process when compared with caption text. Achieving good precision for text recognition from both video and natural scene images continues to be a wide open issue in the area of image processing and pattern recognition because the majority of the existing approaches either concentrate on caption text in video or scene text in natural scene images although not both video and natural images. The issue of text recognition and recognition from scanned document images isn't new for that document analysis community because for various scripts we are able to find several Optical Character Recognizers (OCR engines) that are offered openly. Presently, the approaches make use of a fixed quantity of frames because it is difficult to validate the creation of the enhancement process. Additionally, these approaches work just for video instead of a person image. However, exactly the same methods might not be employed for recognition and recognition from the texts in video and natural scene images since the approaches work nicely for plane background high contrast images although not for images like video and natural scene images [2]. To widen the scope of document analysis based approaches, strategies suggested for text recognition from natural scene images. You will find approaches suggested for text recognition by exploring temporal frames. Generally, these approaches use temporal frames to create specifics of texts and false positive elimination by integrating temporal frames into single enhanced frame.

IJMTARC – VOLUME – IV – ISSUE - 16 - DEC 2016 ISSN: 2320-1363

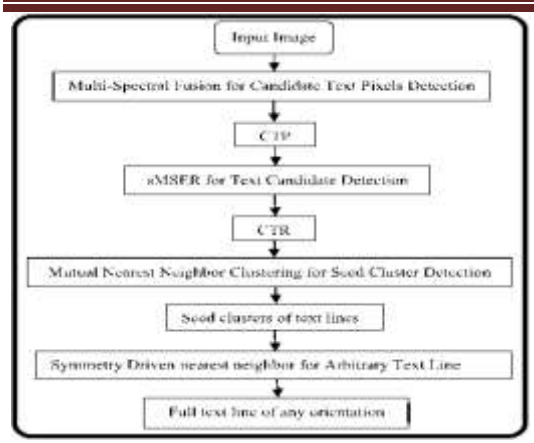


Fig.1.Block diagram of proposed system

## 2. PROPOSED SYSTEM:

The suggested approach includes four steps. In the initial step, once we are inspired through the work presented for multi-oriented video text recognition while using mixture of Laplacian and Fourier, we advise a manuscript concept of convolving Laplacian with wavelet sub-bands at different levels within the frequency domain to boost text pixels via a fusion concept, which leads to multi-spectral fusion. The fused images are exposed to fuzzy k-means clustering to classify the Candidate Text Pixels (CTP). Within the next step, for candidate text pixels, we explore Maximally Stable Extremely Regions (MSER) to group the candidate text pixels into text regions. MSER continues to be effectively employed for classifying text and non-text components previously. Within this work, we advise to change MSER as sMSER together with Stroke Width Transform (SWT) to cluster candidate text pixels as text regions, which we call Candidate Text Regions (CTR). Within the next step, we introduce Mutual Nearest Neighbor Clustering (MNN) for that CTR image to group candidate text regions owed towards the same text line [3]. The suggested approach blogs about the geometrical qualities of CTR before grouping them into just a single one. The creation of this task is stated to become seed clusters that represent a text line. Sometimes, this task may eliminate text components because of mismatching. So within the 4th step, we present a brand new symmetry driven nearest neighbor process for every seed cluster to revive missing text clusters, which lead to a complete text type of any orientation. The symmetry is understood to be the space between two nearest neighboring components. Laplacian-wavelet combination will work for finding specifics at edges within the image. We make use of a two-level wavelet tactic to identify candidate text pixels (CTP) for that input image. First, we apply wavelet. Decomposition to acquire sub-bands,

IJMTARC – VOLUME – IV – ISSUE - 16 - DEC 2016 ISSN: 2320-1363

namely, LH, HL and HH for level-1 and level-2. Next, the suggested approach convolves Laplacian of various directions using the particular three sub bands to boost text pixel information in numerous directions. Then text pixels are separated in the three enhanced images with the aid of fuzzy k-means clustering, which leads to three text clusters. Candidate text pixels are acquired by performing an intersection operation around the three text clusters. This method is known as intra fusion operation which outputs Candidate Text Pixels-1 (CT1). Exactly the same process can also be employed for wavelet decomposition level-2, which outputs Candidate Text Pixels-2 (CT2). The suggested approach correspondingly convolves Laplacian masks of various directions namely horizontal, vertical and diagonal with sub-band images within the frequency domain, leading to three enhanced spectral images. It's observed that scene texts in video and natural scene images have large variations in font size. Therefore, we advise to deploy the suggested approach at multi-level wavelet decomposition. We perform morphological operations to group the pixels which have close closeness. To be able to eliminate non-text pixels and retain text pixels, we advise to fuse the 3 images by performing an intersection operation, which provides one fused image that contains most critical text pixels [4]. The suggested approach repeats all of the steps in the input towards the fused image for that second level. To get rid of such non-text candidate text pixels, we advise to understand more about MSER combined with the stroke width property since it is correct that MSER is really an effective method for staring at the characteristics of connected components. Since CTP is really a binary image and MSER requires grey information, the suggested approach extracts grey values within the input image akin to the candidate text pixels within the CTP image. However, the suggested sMSER doesn't remove non-text candidate text pixels in the CTP image. Rather, it will help in preserving the shapes from the figures. For every text candidate, say an inside a seed cluster, the suggested approach finds the closest neighbor by utilizing Euclidean distance between your cancroids from the text candidates. Thus the primary benefit of this method is it works best for any direction of text line. Within this work, we use Stroke Width Transform

ISSN: 2320-1363

(SWT) like a supporting feature as it is a generally used baseline method for text recognition in natural scene images reported within the literature. The suggested concept of mixing Laplacian with wavelet decomposition and sMSER contributes more for locating accurate text candidates regardless of the look type [5].

## 3. CONCLUSION:

We've explored Maximally Stable Extremely Regions together with stroke width distances for preserving specifics of text candidates. Based on our understanding, this is actually the first make an effort to identify text both in video frames and natural scene images with higher accuracies. Within this paper, we've suggested a manuscript concept of mixing Laplacian with wavelet high frequency sub-bands through fusion at multi-level to recognize text candidates. The suggested approach introduces mutual nearest neighbor clustering according to geometrical qualities of text candidates to group text candidates of particular text lines into clusters. However, based on the results, the precision continues to be less than that in document analysis. The symmetry driven growing process is suggested to extract arbitrary text lines in line with the distance between text candidates in every cluster.

## REFERENCES:

[1] 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.

[2] Z. Zhang, W. Wang, and K. Lu, "Video text extraction using the fusion of color gradient and log-Gabor filter," in Proc. ICPR, Aug. 2014, pp. 2938–2943.

[3] X. Zhao, K.-H. Lin, Y. Fu, Y. Hu, Y. Liu, and T. S. Huang, "Text from corners: A novel approach to detect text and caption in videos," IEEE Trans. Image Process., vol. 20, no. 3, pp. 790–799, Mar. 2011.

[4] D. Crandall, S. Antani, and R. Kasturi, "Extraction of special effects caption text events from digital video," in Proc. IJDAR, 2003, pp. 138–157.

ISSN: 2320-1363

[5] Y. Liu, D. Zhang, Y. Zhang, and S. Lin, "Real-time scene text detection based on stroke model," in Proc. ICPR, Aug. 2014, pp. 3116–3120.