Scientific Journal of Impact Factor (SJIF): 5.71
e-ISSN (O): 2348-4470
p-ISSN (P): 2348-6406

International Journal of Advance Engineering and Research Development

Volume 5, Issue 02, February -2018

Real-time Scene Text Detection via Connected Component Clustering and Non-text Filtering in videos

D.SONY

GEETHANJALI COLLEGE OF ENGINEERING AND TECHNOLOGY

Abstract: *In this paper, we present a new scene text detection algorithm which does the deblurring of the image in blurred images and then extracts the text components through connected component extraction techniques. The proposed image prior is motivated by observing distinct properties of text images. Based on this prior, we develop an efficient optimization method to generate reliable intermediate results for kernel estimation. The proposed method does not require any complex filtering strategies to select salient edges which are critical to the state-of-the-art deblurring algorithms. We discuss the relationship with other deblurring algorithms based on edge selection and provide insight on how to select salient edges in a more principle way. In the final latent image restoration step, we develop a simple method to remove artifacts and render better deblurred images. To be precise, we extract connected components (CCs) in images by using the maximally stable extremal region algorithm. These extracted CCs are partitioned into clusters so that we can generate candidate regions.*

Introduction

In recent years content based indexing of digital video has achieved research importance. In video detected text can offer a useful index with recognition [3].The text seen in video can be either scene text or artificial text. Scene text that occurs naturally in the recorded 3-D scene and is distorted by perspective projection, artificial text, the text overlaid on the video frame while editing.

There are certain unique problems in detection of text from color video images. The video frames are typically of low resolution and compressed which can lead to color bleeding between text and background. In a video frame the text can be low contrast and multi-colored and multi-font which means large-scale assumptions on the color of the text or background components cannot be made. This can be translucent and with an altering from frame to frame. The perspective deformation is performed in scene text.

Previous Work

Artificial text is identified from MPEG video by recognizing homogeneous regions in intensity images, segmentation by interactive threshold selection is performed and to remove non text regions heuristics is applied by Shim et al [13]. The character segmentation is accomplished by local thresholding trailed by pairing nearby regions on gray scale images with similar gray levels by Ohya etal [11].The text is extracted from video frames and images are compressed by background color separation and heuristic is applied that the text has high contrast with background and is of the foreground color by Jain and Yu [4].Automatic text recognition is technologically advanced by Lienhart and Stuber [8].Both spatial and temporal heuristics are used by the system to segment and track(possibly scrolling) caption text in video.

Proposed Approach

No algorithm is robust for finding of an unconstrained variety of text appearing in video from the study of published algorithms. We develop a system which uses a battery of diverse methods employing a variety of heuristics for detecting, localizing and segmenting both artificial and scene text and takes benefit of the temporal nature of video. In our system recognition of multi-font text is an entirely altered research problem is not addressed. The main task achieved

by the system is detection & localization, decision-fusion, and segmentation. To detect and segment scene text that contents certain constraints there is a specialized module.

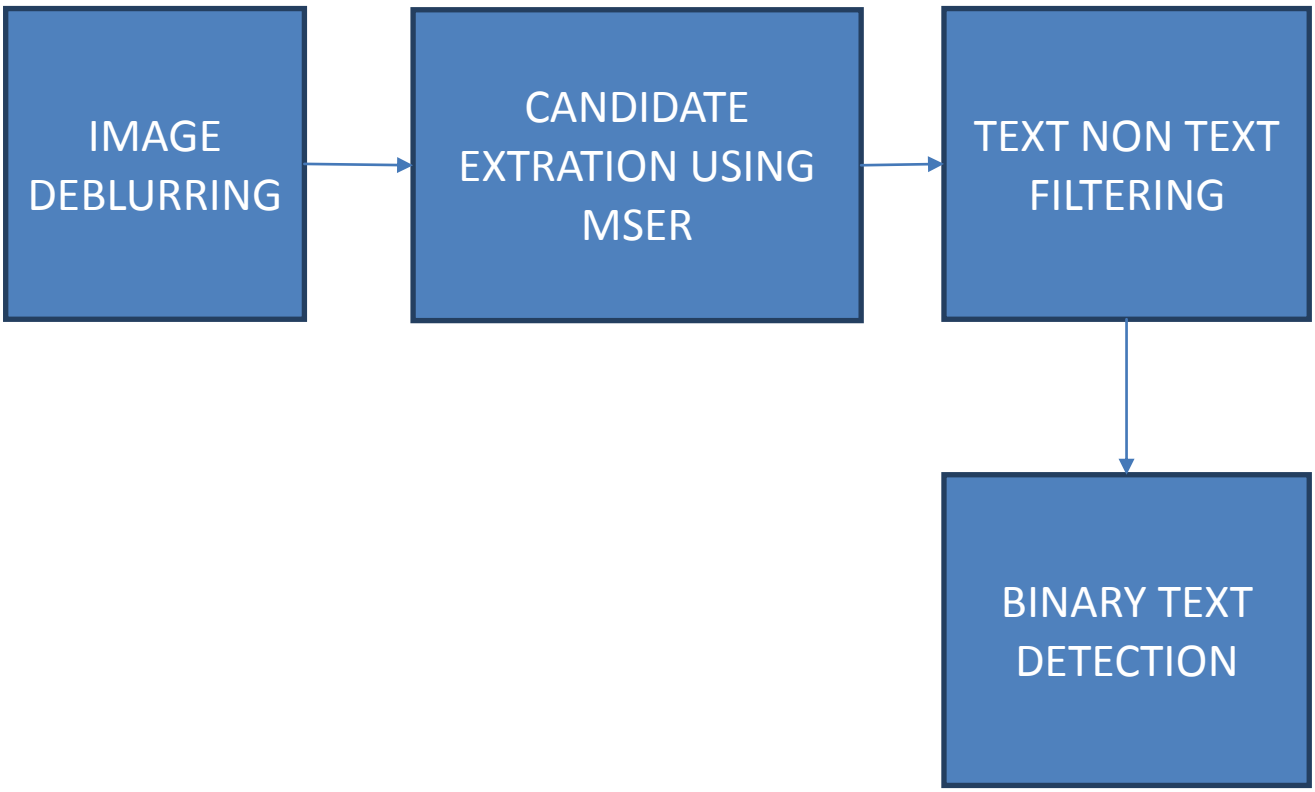

Fig 1. Block diagram

- The image is down sampled to $1/5^{th}$ of its original size.
- Then second order derivative algorithm is used to extract the edges at this level.
- Histogram based colour modification is done in order to adjust the colour around the edge and thus achieve deblurring.
- Then the image is up-sampled and the above two steps are repeated.
- The above 3 steps are done until the original image size is reached.

Scene Deblur

The recent years have seen significant advances in single image deblurring. Much success of the state-of- the-art algorithms can be credited to the use of learned prior from natural images and the assortment of salient edges for kernel estimation[6, 16, 3, 18, 12, 10, 20]. Although many methods have been proposed for motion deblurring, these priors are less operative for text images due to the subjects of concern being mainly two-toned (black and white) which do not follow the heavy-tailed gradient statistics of natural images. Text image deblurring has achieved considerable attention in recent years owing to its wide range of applications.

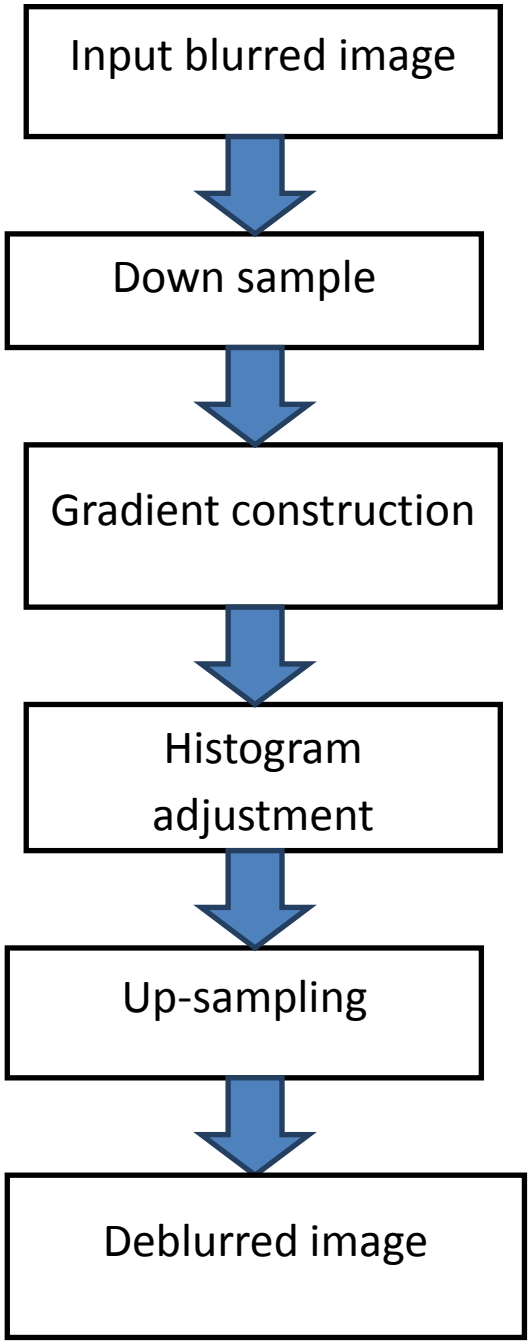

Fig 2: deblurring steps

Graphics/Text block classification

A simple algorithm to categorize video frame blocks in to graphics or video centered on the dynamic range and variation of gray levels within the block is proposed by Mitrea and de With [9].the image compression rate are enhanced by their use. To classify blocks as text and non text shadowed by morphological filtering of text regions this method is slightly altered.

Segmentation

A bounding box is passed in segmentation stage around a localized text region which is produced from multiple frames by decision fusion module. Tight constraints are applied due to minor extent. A few methods at segmentation are accomplished. LeBourgeois' method [7] for inter-character segmentation is unsuccessful for overlapping or italic characters. The background is the largest component of the histogram in numerous background/foreground detection methods proposed in [4] and [7].The assumption of this method is unacceptable in complex assumption. The logical level method of Kamel and Zhao [5] proposed for low contrast check images good results are attained by means of both original and inverted images and one of the result is chosen based on size and connected components distribution

Scene Text Detection using MSER

In computer vision, maximally stable extremal regions (MSER) are used as a technique of blob detection in images. This technique was suggested by Matasetal.[1]to find correspondences between image elements from two images with diverse viewpoints. This method of extracting a complete number of corresponding image elements contributes to the wide-baseline matching, and it has led to improved stereo matching and object recognition algorithms.

Image I is a mapping $I : D \subset \mathbb{Z}^2 \rightarrow S$. Extremal regions are well defined on images if:

1. S is totally ordered (total, anti-symmetric and transitive binary relations $\leq$ exist).
2. An adjacency relation $A \subset D \times D$ is defined.

Region Q is a contiguous subset of D. (For each $p, q \in Q$ there is a sequence $p, a_1, a_2, .., a_n, q$ and $pAa_1, a_iAa_{i+1}, a_nAq$.)

(Outer) Region Boundary $\partial Q = \{q \in D \setminus Q : \exists p \in Q : qAp\}$, which means the boundary ∂Q of Q is the set of pixels adjacent to at least one pixel of Q but not belonging to Q.

Extremal Region $Q \subset D$ is a region such that either for all $p \in Q, q \in \partial Q : I(p) > I(q)$ (maximum intensity region) or for all $p \in Q, q \in \partial Q : I(p) < I(q)$ (minimum intensity region).

Maximally Stable Extremal Region Let $Q_1, .., Q_{i-1}, Q_i, \cdots$ be a sequence of nested extremal regions ($Q_i \subset Q_{i+1}$). Extremal region Q_{i*} is maximally stable if and only if $q(i) = |Q_{i+\Delta} \setminus Q_{i-\Delta}|/|Q_i|$ has a local minimum at $i*$. (Here $|\cdot|$ denotes cardinality.) $\Delta \in S$ is a parameter of the method.

The equation checks for regions that remain stable over a definite number of thresholds. If a region $Q_{i+\Delta}$ is not significantly larger than a region $Q_{i-\Delta}$, region Q_i is taken as a maximally stable region.

The concept more simply can be described by thresholding. All the pixels lower than a given threshold are 'black' and all those above or equal are 'white'. If we are shown a sequence of threshold images I_t with frame t corresponding to threshold t, we would get first a white image, then 'black' spots corresponding to local intensity minima will appear then grow larger. These 'black' spots will finally merge, until the whole image is black. The set of all connected components in the arrangement is the set of all extremal regions. In that sense, the idea of MSER is linked to the one of component tree of the image. The component tree indeed offers an easy way for implementing MSER.

Extremal regions in this context have two significant properties that the set is closed under...

1. Continuous transformation of image coordinates. This means it is affine invariant and it doesn't matter if the image is warped or skewed.
2. Monotonic transformation of image intensities. The method is of course sensitive to natural lighting effects as variation of day light or moving shadows.

Results and Discussions

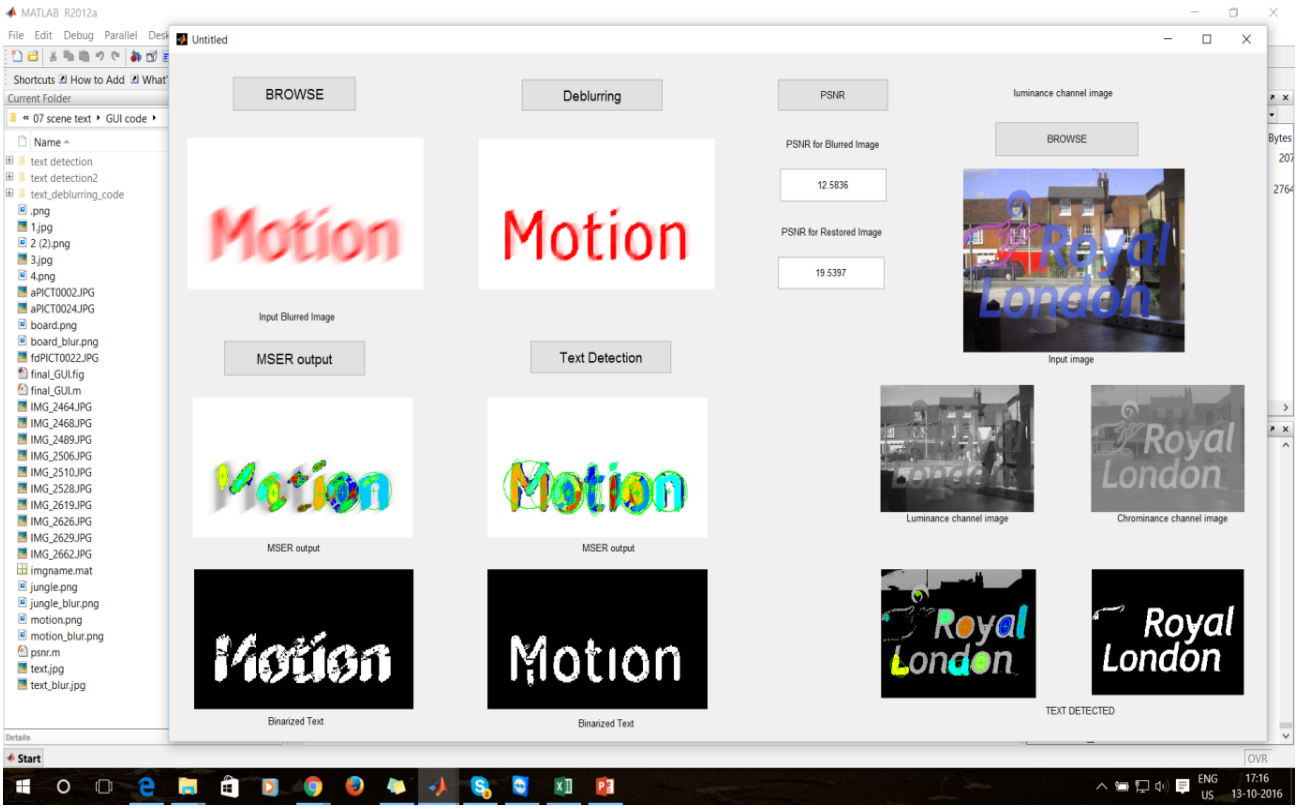

Fig 3: Text detection after deblurring

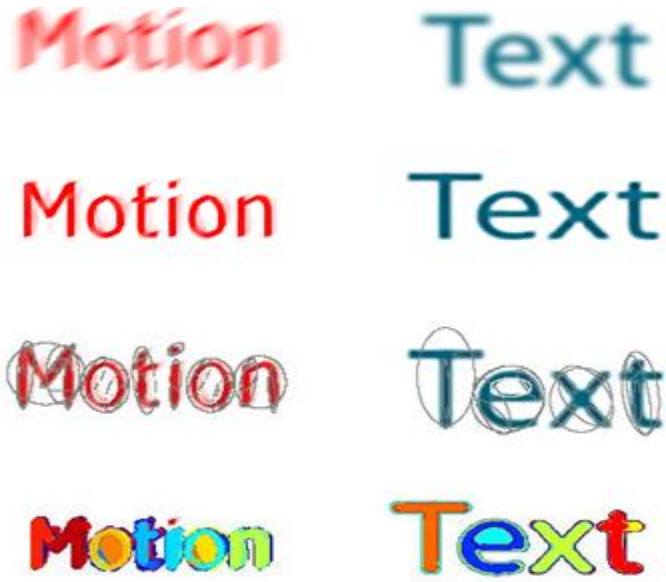

Figure 4 Results of Motion and Blur Images Text Detection

Conclusion

In this paper, we propose a simple yet effective prior for text image deblurring. While the proposed prior is based on the properties of two-tone text images, it can also be effectively applied to non-document text images and low-illumination scenes with saturated regions. With this prior, we present an effective optimization method based on a half-quadratic splitting strategy, which ensures that each sub-problem has a closed-form solution. The proposed method does not require any complex processing techniques. In addition, we develop a simple latent image restoration method which helps reduce artifacts effectively. Our future work will focus on a better non-blind deconvolution method and extend the proposed algorithm to non-uniform text image de-blurring thus helping in scene text detection.

References

[1] N. Chaddha, R. Sharma, A. Agrawal, and A. Gupta. Text Segmentation in Mixed–Mode Images. In *28thAsilomar Conference on Signals, Systems and Computers*, pages 1356–1361, October 1994.
[2] S. Devadiga. *Detection of Obstacles in Monocular Image Sequences*. PhD thesis, The Pennsylvania State University, Computer Science and Engineering Department, August 1997.
[3] U. Gargi, S. Antani, and R. Kasturi. Indexing Text Events in Digital Video. In *Proc. International Conferenceon Pattern Recognition*, volume 1, pages 916– 918, August 1998.
[4] A. K. Jain and B. Yu. Automatic Text Location in Images and Video Frames. *Pattern Recognition*, 31(12):2055–.126, 1998.
[5] M. Kamel and A. Zhao. Extraction of Binary Character/ Graphics Images from Grayscale Document Images.
[6] K. Kanatani. *Geometric Computation for Machine Vision*. Oxford Science Publication, Walton Street, Oxford, UK, 1993.
[7] F. LeBourgeois. Robust Multifont OCR System from Gray Level Images. In *International*
[8] R. Lienhart and F. Stuber. Automatic Text Recognition in Digital Videos. In *Proceedings of SPIE*, volume2666, pages 180–188, 1996.
[9] M.v.d.Schaar-Mitrea and P.H.N. de With. Compression of Mixed Video and Graphics Images for TV Systems. In *SPIE Visual Communications and Image Processing*, pages 213–221, 1998.
[10] Y. Nakajima, A. Yoneyama, H. Yanagihara, and M. Sugano. Moving Object Detection from MPEGCoded Data. In *Proceedings of SPIE*, volume 3.12, pages 988–996, 1998.
[11] J. Ohya, A. Shio, and S. Akamatsu. Recognizing Characters in Scene Images. *IEEE Transactions onPattern Analysis and Machine Intelligence*, 16:214– 224, 1994.
[12] M. Pilu. On Using Raw MPEG Motion Vectors to Determine Global Camera Motion. In *Proceedings ofSPIE*, volume 3.12, pages 448–459, 1998. [13] J.-C. Shim, C. Dorai, and R. Bolle. Automatic Text Extraction from Video for Content-Based Annotation and Retrieval. In *Proc. International Conference onPattern Recognition*, pages 618–620, 1998.
[14] L.L.Winger, M.E. Jernigan, and J.A. Robinson. Character Segmentation and Thresholding in Low-ContrastScene Images. In *Proceedings of SPIE*, volume 2660,pages 286–296, 1996.

