International Journal of Pure and Applied Mathematics
Volume 119 No. 14 2018, 103-110
ISSN: 1314-3395 (on-line version)
url: http://www.ijpam.eu
Special Issue

ijpam.eu

A Theoretical Approach on Detection and Recognition of Text in Complex Images and Video Frames

Ravikumar H C[1], Dr.P.Karthik[2]

[1]Research Scholar, [2]Professor, Dept. of ECE, KSSEM, Bangalore

Abstract

Detection and recognition of text in images and video frames plays an important role, to access images and videos from multimedia data base, containing huge amount of imagery and video data. This paper gives a brief review of methods and techniques used in the detection, recognition and analysis of text in images and video frames. This paper also highlights the various challenges facing in detecting and extracting text and the application of text in various fields. Text may suffer from various complex problems, for detection and extraction need to increase resolution and remove blur in images or video frames, so super resolution and deblurring techniques have been addressed. This paper gives information about data sets, compares result of existing techniques and gives brief idea to conduct research in future directions

1. Introduction

The explosive growth of smart phones and online social media have led to the increase in the size of multimedia databases at a tremendous speed, in particular, the massive and increasing collections of images and videos on the internet and social networks. The understanding of such data bases is more important for effective content-based indexing and retrieval. Video search function which is categorized by manual indexing requires excessive time and money. The analysis of imagery and video data is currently in wide demand because images and video are major source of sensory data in our lives. Text is a prominent and direct source of information in video, so it has received increased attention. Text such as caption in a video provides important information of the contents as well as description of the video scenes [1] [2]. Caption text usually annotates information concerning where and when and the events in video happened or who was involved [3]. Such text can be used as indices of the multimedia database. The indexing information, such as scene locations, speaker names, program titles, sports score, dates and time.

Text is one of the most expressive means of communications, and can be embedded into documents or into scenes as a means of communicating information. This is done in the way that is readable by others (4). Text plays a crucial role in our daily lives. Text has classified into caption text and scene text. Caption text is also called graphic text, it refers to machine print text overlaid graphically, found in captions, subtitles and annotations in video and born-digital images. Caption text has classified into two subcategories: layered caption text and embedded caption text. Layered caption text is always printed on a specifically designed background layer, while embedded caption text is overlaid and embedded on the frame. While scene text is part of the camera images and is naturally embedded within objects (e.g., trademarks, signboards, text on signs, packages and clothing in natural scenes, and is more likely to include handwritten material and buildings) in scenes [4][5][6], text in imagery and video has shown in fig.1.

Fig. 1 Text in images and videos (a) Video graphical text. (b) Point-and-shoot scene text. (c) Incidental scene text.

The increasing availability of high performance mobile devices with both imaging and computational capability creates an opportunity for image acquisition and processing to makes to recognise text in various environments anytime and anywhere[7][8]. Text recognition from any natural scene images and videos are application of image processing techniques. The problem of text detection and recognition in images and video has received increased attention in recent years [9][10]. The advance in computer vision and pattern recognition technologies makes it more reasonable to address challenging problems. 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 detection and information extraction, exciting event extraction from sports video, identifying athletes in marathon events, etc. [11].

The recognition of text gives rise to many applications, the fundamental goal is to determine whether or not there is text in a given image, and if there is, to detect, localize, and recognize it. In

the literature, various stages of these fundamental tasks are referred to by different names including text localization [12], which aims to determine the image positions of candidate text, text detection, which determines whether or not there is text using localization and verification procedures, and text information extraction [13], which focuses on both localization and binarization. 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 [14], [15] either focus on caption text in video or scene text in natural scene images but not focus and does not work well on both video and natural images.
In the problem of text detection and recognition, publicly available Optical Character Recognizers (OCR engines) works well for plane background and high contrast images and does not works for video and natural scene images which are suffering from non-uniform illumination, perspective distortion, low resolution, low contrast, blurred, varying font types and font sizes, multiple colors and arbitrary orientations,[13][16], [17][18].

Text detection and recognition distortion due to motion blur is a major issue among all other artifacts because blur interferes with the structure of the components which in turn changes the shape of the components and fail to give satisfactory results when blur exist in the video/image[19][20]. Low lighting or with motion between the camera and the scene resulting in blurred images. Therefore, there is a need for developing a method for deblurring the blurred image for improving the performance of the text detection and recognition methods in a real-time environment [61].

Compared with documents and scene images, scene text in video is always blurred and video frames usually have a lower resolution (low contrast)[22].then it is difficult to detect and recognize text from images and video frames without increasing resolution and deblurring . To increase resolution, super resolution techniques can be used. Image super-resolution (SR) is the process of generating one or more high-resolution (HR) images from one or more low resolution (LR) observations. SR has found its usage in many image processing applications, varying from satellite and aerial imaging, medical image processing, sign and number plate recognition, facial image recognition and enhancement in surveillance videos, iris recognition, fingerprint recognition, enhancement of text documents, etc.[23,24].

2. Applications

Text related applications for both images and videos are classified as multimedia retrieval, Assisting Visually Impaired People, Industrial automation, Real-Time Translation and Number plate recognition.

Multimedia retrieval: Recognizing text and extracting keywords in multimedia resources enhances multimedia retrieval.
Industrial automation: Recognizing text on packages, containers, houses, and maps has broad applications related to industrial automation [28].
Assisting Visually Impaired People: Development of text recognition prototypes helpful for people who are visually impaired [29], in understanding scenes in their surrounding environment [30]. Developing personal text-to-speech devices assists them in understanding grocery signs, product and pharmaceutical labels, and currency and ATM instructions [8].
Real-Time Translation: Text extraction is also important for translation purposes. Development of automatic sign/text language translation system for foreign travelers used to translate detected text in a scene into a traveler's native language [31].
Number plate recognition: Extracting characters from the license plates of moving vehicles [32], which is helpful for the traffic department for automatic number plate recognition in order to control the traffic. Another application of video text recognition is to detect text on road signs from natural scene videos applicable for driving assistance systems. [33].

3. Related work

Text extraction in images and video frames divided into four tasks: text detection, text localization, text extraction, optical character recognition [34][5].text detection is used to determine whether an image or video frame contains text information, text localization is to find the actual location of the text present in the image or frame, mark a border area ,the purpose of text extraction is to extract and binarize the text for OCR. In this paper, we review the literature on detection and extraction of text on complex images and videos and methods of deblurring and super resolution techniques.

The literature on text detection in images and video can be categorized into connected component-based, texture-based and gradient and edge-based methods [4][14][35]. Connected component based methods study the shape of the components for extracting features. These methods are good for high contrast text such as caption in video, but not for scene text because scene text is unpredictable and if the image contains blur, the method fails to extract the shape of the components with the features. Zhao et al. proposed an approach for text detection using corners in the video. This method proposes to use dense corners for identifying text candidates. From the corners, the method forms regions of text using morphological operations. When the images or video contains blur, the method may fail to obtain the expected dense corners because of the loss of character component shapes due to blur artifacts [36].

The texture-based methods are proposed to overcome the limitation of the connected component-based methods. These methods define the appearance of the text pattern as a special texture, and they generally depend on a large number of features and classifiers for achieving good results, it is said to be computationally expensive. The methods are sensitive to font variation and distortion. Liu et al. proposed a method for text detection in video using a set of texture features and k-means clustering [37]. Shivakumara et al. proposed a method which combines the Fourier transform and color spaces for text detection in video [38]. These methods require more computations and not capable of handling blurred images because blur tempers the texture property defined for text detection.

Texture-based methods, the combination of edge and gradient feature-based methods are proposed to reduce the computational burden. Epshtein et al. proposed an image operator that seeks to find the value of stroke width for each component pixel, and demonstrates its use on the task of text detection in natural images. Since the method expects stroke width to be constant for every edge component, when blur exists in the video or images, the condition may not be satisfied [39]. Shivakumara et al. proposed multi-oriented scene text detection in video using a Bayesian classifier and boundary growing. This method uses the combination of Sobel and

Laplacian operations for enhancing the text information. This method is sensitive to blur because the method involves Sobel and Laplacian operations which may fail to produce fine edges for the blurred regions [40].

There are methods that utilize the temporal information for text detection in images or video, for enhancing low contrast text pixel. Huang proposed a method for detecting text in video based on the video's temporal redundancy. Video scene text detection is implemented in a single frame to retrieve candidate text regions. Finally, the synthesized motion image is used to filter out candidate text regions and only keep the candidate text regions which have motion occurrence as final scene text [41]. Mi et al. proposed a text extraction approach based on multiple frames. The edge features are explored with similarity measures for identifying text candidates. Since extracted features are sensitive to blur, the performance of the method degrades when the blur region exists in the video [42].

Sun et al. proposed a robust method for text detection in natural scene images based on color-enhanced contrasting extremal region and neural networks. Characteristics of text components extracted in this method to identify text candidates and false text candidates are eliminated. This method does not perform well for low contrast images and non-horizontal text images[43].Yin et al. proposed robust text detection in natural scene images based on maximally stable extremal regions (MSER).The method uses a pruning algorithm to select appropriate MSERs detecting character candidates.This method does not work well for blurred images, multi- lingual text images and it is not perfect for multioriented text images[44]. Yin et al. proposed a multi-oriented scene text detection approach with adaptive clustering. In their method, a unified distance metric learning frame work for adaptive hierarchical clustering was proposed [45] to overcome above [44] problems. This method detects text well for those images with high contrast and with no blurring effects and do not perform well for low contrast images with complex backgrounds.

Liang et al. proposed a new idea of convolving Laplacian with wavelet sub-bands at different levels in the frequency domain for enhancing low resolution text pixels and maxima stable extreme regions along with stroke width transform for detecting candidate text regions are explored [46]. The accuracy is still lower than document analysis. Favorskaya et al. used the methodology based on the analysis of the gradient sharp profiles includes the automatic text detection in fully or partially blurred frames of a non-stationary video sequence and achieved better detection results for corrupted text fragments[54] .A cascaded method combines text line entropy with a Convolutional Neural Network model. It is used to verify text candidates, which reduces the number of non-text regions, leading to significantly enhanced performance on text detection, but some multi-orientation text lines are not detected correctly [55].

Text recognition in image or video frame is conventionally performed using OCR techniques; text regions are first segmented from images or video frames and then fed into a state-of-the art OCR engine [5]. However, the recognition performance relies heavily on text segmentation / binarization may suffer from noise and distortion in complex images videos. Hence, several methods have been specifically designed for text recognition.

A large number of features based on maximally stable extremal regions (MSER), scale invariant features(SIFT), histogram oriented features (HoG) and their own classifier with a large number of training samples can be used for recognition of text in images or video frames [47][48]. The recognition ability is restricted to individual languages and these methods are computationally expensive because of a large number of features.The extracted shape-based features are not sufficiently capable in extracting the structure of the components when blur exists in the video or images.

Authors proposed methods for recognizing text in degraded document images which include blurred handwritten document images containing images of ink-bleeding and double-sided writing[49][50]. Zhou et al. describes a new text segmentation method based on inverse rendering. The technique uses iterative optimization to solve the rendering parameters, including light source, material properties (e.g. Diffuse/specular reflectance and shininess) as well as blur kernel size [51].These methods require high contrast and homogenous backgrounds and these methods may not be suitable for video and natural scene images.

Shi et al. proposed using deformable part based models and sliding window classification to localize and recognize characters in scene images. The DPMs effectively recognize characters with distortion and with a variety of fonts [52].Yao et al. proposed a learned representation named Strokelets for character recognition. Strokelets captures the structural characteristics of characters at multiple scales, ranging from local primitives, like bar, arc and corner to whole characters. A histogram feature named Bag-of-Strokelets is formed by binning the Strokelets and is trained with Random Forest for recognition. This approach has robustness to distortion and generality to variant languages [53].

Phan et al. proposed a semiautomatic system for ground truth generation for video text detection and recognition. The precision of the proposed method is not as high, because many nonhorizontal texts are scene texts, they are much more challenging to detect than horizontal texts, and sometimes produces false positives [55].Roy et al. proposed a new method for binarizing text in video by keeping the limitations of the conventional binarization methods and its performance degrades when distortions are present in the images, such as disconnections, loss of information due to blur and illumination effects[56]. Tian et al. proposed a character shape restoration through medial axis points in video to overcome the problem of [56]. The method basically uses a ring radius transform concept for finding the medial axis and then a new restoration method using medial axis points. In the case of blur, the contours may lose geometrical coherency [57].

Pan et al. proposed a effective L0-regularized prior based on intensity and gradient for text image deblurring. L0-norms are an np-hard problem which makes it expensive in terms of time complexity, and hence restricts its use in video applications [58]. Cao et al. proposed the method which explores the combination of multi-scale dictionaries and an adaptive version of a non-uniform deblurring method. The performance of the method depends on the size of the dictionaries and kernel estimation for different situations [59].Khare et al. proposed a new model based on Gaussian Weighted L1 (GW-L1) with alternative minimization, which helps in enhancing edge strength to eliminate blur in the video/image. The proposed deblur model helps in improving the performance of both text detection and recognition methods for uniform bur region, it does not work well for non-uniform blur region [60]. From the literature review it is observed that less number of models developed for detecting and

extracting text in blurred image and video frames, most of the models proposed for general image deblurring , some deblurring techniques are discussed in this paper, by using this techniques we can able to solve text recognition problems in blurred images or video frames.

Li et al. proposed a blind image motion deblurring method which adopts L0-regularized priors both in kernel and latent image estimation. The proposed method is easy to implement since it does not require any complex filtering strategies to select salient edges which are critical to the explicit salient edges selection methods [61].The proposed method is superior because of the better performance when compared with other methods. kong et al. proposed a new blind deblurring method in the maximum a posterior (MAP) framework to remove the blur of the image. The hyper-Laplacian has chosen prior to be a regularization of the gradients of an image and adopted an operator called generalized soft thresholding (GST) to solve the non-convex problem during the whole deblurring process [62]. The approach of this method can only deal with the condition of spatially invariant blur kernel; it does not give solution to spatially variant blur kernel. kumar proposed an effective method of estimating the point spread function (PSF) parameters based on the concepts of Histogram of oriented gradients (HoG) and statistical properties of an image. The advantage of using moment domain are fast convergence and robust to the variations in the parameters of the PSF. This method effectively restores the original images [63]. Renzhen et al. proposed latent semantic concept regularized (LSCR) to reduce the blind deconvolution problem at a semantic level. The blind deconvolution problem can be regularized and the sharp version of the blurry image can be recovered at a new latent semantic level. Blind deconvolution problem at high semantic level is not solved in this paper [64]. Pan et al. explores Enforcing sparsity of the dark channel thus helps blind deblurring in various scenarios such as natural, face, text, and low-illumination images and does not work well for noisy images [65].

To detect and recognize text accurately in low contrast or low resolution images or video frames, it is necessary to increase the resolution, for that we can use some super-resolution techniques which have reviewed in this paper. Laghrib et al. proposed a method consists of a non-parametric image registration based on diffusion regularization and a nonlocal Laplace regularizer combined with a bilateral filter (BTV) in the reconstruction step to remove noise and motion outliers. This method is suffering from staircasing effect and computation of the TGV term is slower [66]. Jinsheng et al. proposed a novel image enhancement algorithm based on adaptive shock filter for image super-resolution. The proposed algorithm eliminates edge halos and jagged artifacts, whereas the fine image structures are reserved effectively [67]. Mourabit et al. proposed a new tensor based diffusion regularization that takes the benefit from the diffusion model of Perona–Malik in the flat regions and use a nonlinear tensor derived from the diffusion process of Weickert filter near boundaries. Thus, the proposed SR approach can preserve important image features (sharp edges and corners) much better while avoiding artifacts [68]. singh et al. proposed a multi-frame image super-resolution approach using OFMMs. The proposed approach referred as NLM-OFMMs-I. The proposed approaches generates high-quality HR images in the presence of factors like image noise, global motion, local motion, and rotation in between the image frames [69].

4. Challenges

Detecting and recognizing text in images and videos have some challenges, are categorized into groups based on the complexity of environments, image acquisition styles and variation of text contents[4][25][26]. Various challenges of text detection and recognition are summarized in table 1 presents the unique and shared challenges between scene text and embedded caption extraction in images and videos [2][27]

Table 1: Challenges of Detection and Recognition of Text in Images and Videos

Category	Sub-category	Description
Environment	Scene complexity Uneven lighting Complex background Background-foreground	Complex scenes with similar structures makes difficult to discriminate text from non-text. Due to illumination, introduces color distortion and deterioration of visual features. Non uniform and complex background in the scene. Similar visual distributions between background and foreground.
Image acquisition	Blurring Perspective distortion Low resolution Compressed degradation Moving objects Real-time processing	Blurring from defocusing , motion or low resolution. Due to optical axis of the camera is not perpendicular to the text plane. Images and video frames usually with a very low resolution. Due to encoding and decoding. Moving background objects and moving text in frames. Processing of captured images and video data in real time using computation efficient algorithms.
Text content	Variation of aspect ratio Skew curved text Unaligned text Variation of fonts and shapes Text size Multilingual environments	Search procedure with respect to location, scale and length increases computational complexity. Multi-orientation text captured by the camera. Naturally designed text in a curve shape. Text lines aligned in nonplanar or bent surfaces. Different fonts and shapes overlap each other and also leads to variation in aspect ratio. Various text sizes simultaneously found in frames. Various languages with different characteristics for text.

5. Datasets And Evaluation

5.1 Datasets

The commonly used and standard data sets for caption and scene text in complex images and videos are MoCA, TREC, Merino[71], Minetto[72], Merino-Gracia[73], YouTube Video Text[70], ICDAR 2003, ICDAR 2005, ICDAR 2011, ICDAR 2013, ICDAR 2015[74], SVT, MSRA-TD500, Char74k, MSR-1, MSR-II, VIDI, IIITSK, STD, OSTD, NECR, KIST[2][4].

5.2 Detection and Recognition Evaluation

Detection Evaluation: Text detection results in images or video frames can be measured by using speed, precision and recall. Speed indicates the average processing time per frame in text detection. Precision (P), which evaluates the percentage of text regions correctly detected compared to the text region and recall, which is defined as the ratio of the text regions correctly detected to the ground truth text regions, claimed as follows. [2],

$$P = \frac{\text{number of correctly detected text regions}}{\text{number of detected text regions}} \quad (1)$$

$$\mathrm{R} = \frac{\text{number of correctly detected text regions}}{\text{number of ground truth text regions}} \quad (2)$$

The harmonic measure f is adopted to combine the precision and recall figures:

$$\mathrm{f} = \frac{1.0}{\frac{\alpha}{Precision!} + \frac{1.0-\alpha}{Recall}} \quad (3)$$

Where the parameter α is usually set as 0.5 to give equal importance to precision and recall.

Recognition Evaluation: Text recognition performance is always measured by the accuracy of word recognition. The word recognition accuracy (WRA) is simply defined as the percentage of the recognized text is correct [2], i.e.,

$$\mathrm{WRA} = \frac{\text{number of words correctly recognized}}{\text{number of ground truth words}} \quad (4)$$

6. Result And Discussions

Table 2: Detection response on complex, blurred images and video frames.

Dataset	Recall (R)	Precision (P)	f
V.khare dataset [61]	89.70 [61][40]	83.40 [61][40]	86.43 [61][40]
ICDAR 2013 video frames	82.10 [38][61]	83.10 [38][61]	82.55 [38][61]
Youtube video text	88.68 [38][61]	86.38 [38][61]	87.51 [38][61]
ICDAR 2015 video frames	81.22 [38][61]	81.73 [38][61]	81.47 [38][61]
MSRA-TD500 frame	89.60 [38][61]	82.35 [38][61]	85.83 [38][61]
SVT frames	86.70 [38][61]	84.12 [38][61]	85.30 [38][61]
ICDAR 2013 scene images	89.50 [55]	77.63 [55]	83.14 [55]
ICDAR 2005 images	70.90[21]	73.60 [21]	72.20 [21]
ICDAR 2011 images	89.90 [55]	77.92 [55]	83.48 [55]
ICDAR 2015 images	61.68 [55]	39.53 [55]	48.18 [55]

According to literature review, most of the deblurring techniques applied for general images, not for text contains in images and video frames. To overcome blurring problems Khare et al. proposed a deblurring model [61], to detect and recognize text in complex blurred images and video frames. The deblurring model is applied on some existing techniques and various standard datasets. The performance of text detection and recognition on different datasets with different methods as shown in table 2 and table 3 respectively.

Table 3: Recognition accuracy for complex, blurred images and video frames.

Dataset	Recognition Accuracy
V.khare dataset [21]	68.91 [51] [61]
ICDAR 2013 video frames	72.61 [51] [61]
Youtube video text	78.40 [51] [61]
ICDAR 2015 video frames	75.22 [51] [61]
MSRA-TD500 frame	85.83 [38][61]
SVT frames	85.30 [38][61]
ICDAR 2013 scene images	83.20 [38][61]

7. Conclusion

This paper gives a brief review of methods and techniques used in the detection, recognition and analysis of text in images and video frames. Super resolution and deblurring techniques have been addressed. This paper gives information about data sets, compares result of existing techniques. From the literature survey it is evident that, there exist significant challenges in text detection and recognition from images and video frames and gives brief idea to conduct research in future directions.

References

[1] P. Shivakumara, S. Bhowmick, B. Su, C. L. Tan, and U. Pal, " A new gradient based character segmentation method for video text recognition," in Proc. IEEE Int. Conf. Doc. Anal. Recognit. 2011, pp. 126–130.

[2] X.C.Yin, Z.Y.Zuo, S.Tian, and C.L.Liu," Text Detection, Tracking and Recognition in Video: A Comprehensive Survey", IEEE Trans. Image Process., June 2016, vol 25, no.6, pp. 2752-2773.

[3] 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.

[4] Q. Ye and D. Doermann, "Text detection and recognition in imagery: A survey," *IEEE Trans.Pattern Anal. Mach. Intell.*, vol. 37, no. 7, pp. 1480–1500, Jul. 2015.

[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, May 2004.

[6] P. Shivakumara, A. Dutta, U. Pal, and C. L. Tan, "A new method for handwritten scene text detection in video," in Proc. IEEE Int. Conf. Front. Handwritten Recognit., 2010, pp. 387–392.

[7] X. Chen, J. Yang, J. Zhang, and A. Waibel, "Automatic detection and recognition of signs from natural scenes," IEEE Trans. Image Process., vol. 13, no. 1, pp. 87–99, Jan. 2004.

[8] X. Liu and D. Doermann, "A camera phone based currency reader for the visually impaired," in Proc. ACM SIGACCESS Conf. Comput. Accessibility, 2008, pp. 305–306.

[9] S. Lucas, A. Panaretos, L. Sosa, A. Tang, S. Wong, R. Young, K. Ashida, H. Nagai, M. Okamoto, H. Yamamoto, H. Miyao, J. Zhu, W. Ou, C. Wolf, J. Jolion, L. Todoran, M. Worring adn X. Lin, "ICDAR 2003 robust reading competitions: Entries, results, and future directions," Int. J.Doc .Anal. Recognit. vol. 7, pp. 105-122, 2005.

[10] A. Shahab, F. Shafait, and A. Dengel,"ICDAR 2011 robust reading competition challenge 2: Reading text in scene image," in Proc. IEEE Int. Conf. Doc. Anal. Recognit., 2011, pp. 1491-1496.

[11] 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.

[12] R. Lienhart and A. Wernicke, "Localizing and segmenting text in images and videos," IEEE Trans. Circuits Syst. Video Technol., vol. 12, no. 4, pp. 256–268, Apr. 2002.

[13] J. Zang and R. Kasturi, "Extraction of text objects in video documents: Recent progress," in Proc. IAPR Int. Workshop Doc. Anal. Syst., 2008, pp. 5–17.

[14] 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.

[15] C. Yi and Y. Tian, "Localizing text in scene images by boundary clustering, stroke segmentation, and string fragment classification," *IEEE Trans. Image Process.*, vol. 21, no. 9, pp. 4256–4268, Sep. 2012.

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

[17] D. Chen and J.-M. Odobez, "Video text recognition using sequential Monte Carlo and error voting methods," *Pattern Recognit. Lett.*, vol. 26, no. 9, pp. 1386–1403, Jul. 2005.

[18] L. Wu, P. Shivakumara, T. Lu, and C. L. Tan, "Text detection using delaunay triangulation in video sequence," in *Proc. DAS*, 2014, pp. 41–45.

[19] P.Shivakumara, Z.Yuan, D.Zhao, T.Lu, C.L.Tan, New gradient-spatial- structural features for video script identification, CVIU 130, 2015, pp. 35–53.

[20] X. Cao, W.Ren, W.Zuo, X.Guo, H.Foroosh, Scene text deblurring using text- specific multi scale dictionaries, IEEE Trans. IP 24, 2015, pp.1302–1314.

[21] J.Liu, H.su, Y. Yaohua, and H.wenbin, "Robust text detection via multi-degree of sharpening and blurring", Signal Processing 124, 2016, pp.259-265.

[22] C Yang, X-C Yin, W-Y Pei, S Tian, Z-Y Zuo, C Zhu, and J Yan,"Tracking Based Multi Orientation Scene Text Detection: A Unified Framework With Dynamic Programming", IEEE Trans. Image Process, vol. 26, no. 7, july 2017, pp. 3235-3248.

[23] D. Thapa, K. Raahemifar, W.R. Bobier, V. Lakshminarayanan, A performance comparison among different super-resolution techniques, Comput. Electr. Eng., vol. 54, 2016, pp. 313–329.

[24] K. Nasrollahi, T.B. Moeslund, Super-resolution: A comprehensive survey, Mach. Vision. Appl. Vol 25, 2014, pp.1423–1468.

[25] J. Liang, D. Doermann, and H. Li, "Camera-based analysis of text and documents: A survey," *Int. J. Document Anal. Recognit.*, vol. 7, nos. 2–3, 2005, pp. 84–104.

[26] X.-C. Yin, X. Yin, K. Huang, and H.-W. Hao, "Robust text detection in natural scene images," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 5, May 2014, pp. 970–983.

[27] R. Smith, D. Antonova, and D. Lee, "Adapting the Tesseract open source OCR engine for multilingual OCR," in Proc. Joint Workshop Multilingual OCR Anal. Noisy Unstruct. Text Data, 2011.

[28] P. Sermanet, S. Chintala, and Y. LeCun, "Convolutional neural networks applied to house numbers digit classification," in Proc. IEEE Int. Conf. Pattern Recognit., 2012, vol. 4, pp. 3288– 3291.

[29] C. Yi and Y. Tian, "Localizing text in scene images by boundary clustering, stroke segmentation and string fragment classification," IEEE Trans. Image Process., vol. 21, no. 9, Sep. 2012, pp. 4256-4268.

[30] J-P. Peters, C. Thillou, and S. Ferreira, "Embedded reading device for blind people: A user-centered design," in *Proc. Int. Symp. Inf. Theory (ISIT)*, Oct. 2004, pp. 217–222.

[31] I. Haritaoglu, "Scene text extraction and translation for handheld devices," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, vol. 2. Dec.2001, pp. II-408–II 413.

[32] Y. T. Cui and Q. Huang, "Character extraction of license plates from video," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, Jun. 1997, pp. 502–507.

[33] W. Wu, X. Chen, and J. Yang, "Incremental detection of text on road signs from video with application to a driving assistant system," in *Proc. 12th Annu. ACM Int. Conf. Multimedia (ACM MM)*, 2004, pp. 852–859.

[34] S. Antani, D. Crandall, and R. Kasturi, "Robust extraction of text in video," in *Proc. Int. Conf. Pattern Recognit. (ICPR)*, Sep. 2000, pp. 831–834.

[35] D. Chen, J.M.Odobez, H.Bourlard, Text detection and recognition in images and video frames, Pattern Recognit. Vol.37, 2004, pp.595–608.

[36] X. Zhao, K.H.Lin, Y.Fu, Y.Hu, Y.Liu, T.S.Huang, Text from corners: a novel approach to detect text and caption in videos, IEEE Trans. IP 20, 2011, pp. 790–799.

[37] C.Liu, C.Wang, R.Dai, Text detection in images based on unsupervised classification of edge based features, In: Proceedings of ICDAR, 2005, pp.610–614.

[38] P.Shivakumara, T.Q.Phan, C.L.Tan, New fourier-statistical features in RGB space for video text detection, IEEE Trans. CSVT 20, 2010, pp.1520–1532.

[39] B.Epshtein, E.Ofek, Y.Wexler, Detecting text in natural scenes with stroke width transform, In: Proceedings of CVPR, 2010, pp.2963–2970.

[40] P.Shivakumara, R.P.Sreedhar, T.Q.Phan, S.Lu, C.L.Tan, Multioriented video scene text detection through Bayesian classification and boundary growing, IEEE Trans. CSVT 22, 2012, pp.1227–1235.

[41] X.Huang, A novel approach to detecting scene text in video, In: Proceedings of ICISP, 2011, pp.469–473.

[42] C.Mi, Y.Xu, H.Lu, X.Xue, A novel video text extraction approach based on multiple frames, In: Proceedings of ICICSP, 2005, pp.678–682.

[43] L. Sun, Q.Hou, W.Jia, K.Chen, A robust approach for text detection from natural scene images, PR 48, 2015, 2906–2920.

[44] X.C.Yin, X.Yin, K.Huang, and H.Hao, "Robust text detection in natural scene images," IEEE Trans. Pattern Anal. Mach. Intell., vol. 36, no. 5, May 2014, pp. 970–983.

[45] X.C. Yin, W.Y.Pei, J.Zhang, H.W.Hao, Multiorientation scene text detection with adaptive clustering, IEEE Trans. PAMI 37, 2015, pp.1930–1937.

[46] A.Mishra, K.Alahari, C.V.Jawahar, Top-down and bottom-up cues for scene text recognition, In: Proceedings of CVPR, 2012, pp.2687–2694.

[47] T.Q.Phan, P.Shivakumara, S.Tian, C.L.Tan, Recognizing text with perspective distortion in natural scene images, In: Proceedings of ICCV, 2013, pp.569–576.

[48] G.Liang, P.Shivakumara, T.Lu, and C-L.Tan, "Multi-Spectral Fusion Based Approach for Arbitrarily Oriented Scene Text Detection in Video Images", IEEE Trans. on Image Process., Vol. 24, no. 11, Nov 2015, pp.4488-4501.

[49] X.Chen, X.He, J.Yang, Q.Wu, An effective document image deblurring algorithm, In: Proceedings of CVPR, 2011, pp.369–376.

[50] R. Hedjam, M.Cheriet, Historical document image restoration using multi spectral imaging system, PR 46, 2013, pp.2297–2312.

[51] Y.Zhou, J.Feild, E.L.Miller, R.Wang, Scene text segmentation via inverse rendering, In: Proceedings of ICDAR, 2013, pp.457–461.

[52] C. Shi, C. Wang, B. Xiao, Y. Zhang, S. Gao, and Z. Zhang, "Scene text recognition using part-based tree-structured character detections," in Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit., 2013, pp. 2961–2968.

[53] C. Yao, X. Bai, B. Shi, and W. Liu, "Strokelets: A learned multiscale representation for scene text recognition," in Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit., 2014, pp. 4042–4049.
[54] M.Favorskaya, V.Buryachenko, "Scene text deblurring in non-stationary video sequences", 20th International Conference on Knowledge Based and Intelligent Information and Engineering Systems, Vol. 96, Sep 2016, 744 – 753.
[55] Y.Zheng, Q.Li, J.Liu, H.Liu, G.Li, S.Zhang,"A cascaded method for text detection in natural scene images" Neurocomputing, Vol.238, 2017, pp. 307–315.
[56] T-Q.Phan, P.Shivakumara, S.Bhowmick, S.Li, C-L.Tan, U.Pal,"Semiautomatic Ground Truth Generation for Text Detection and Recognition in Video Images", IEEE Trans. on circuits and systems for video techn., vol. 24, no. 8, Aug 2014, pp. 1277-1287.
[57] S. Roy, P.Shivakumara, P.P.Roy, U.Pal, C.L.Tan, TongLu, "Bayesian classifier for multi-oriented video text system", ESWA 42, 2015, pp.5554–5566.
[58] S. Tian, P.Shivakumara, T.Q.Phan, T.Lu ,C.L.Tan, "Character shape restoration system through medial axis points", Neurocomputing, vol.161, 2015, pp.183–198.
[59] J.Pan, Z.Hu, Z.Su, M.H.Yang, "Deblurring text images via L0-regularized intensity and gradient prior", In: Proceedings of CVPR, 2014, pp.2901–2908.
[60] X. Cao, W.Ren, W.Zuo, X.Guo, H.Foroosh, "Scene text deblurring using text- specific multi scale dictionaries", IEEE Trans. IP 24, 2015, pp.1302–1314.
[61] V. Khare, P. Shivakumara, and P. Raveendran, "A blind deconvolution model for scene text detection and recognition in video," Elsevier., Pattern Recognition ,54, Jan. 2016, pp. 128–148,.
[62] J.Li, W.Lu,"Blind image motion deblurring with L0-regularized priors", J. Vis. Commun. Image R. Vol 40, 2016, 14–23.
[63] J.Kong, L.Kesai, M.Jiang,"A New Blind Deblurring Method via Hyper-Laplacian Prior", International Congress of Information and Communication Technology, Vol. 107, 2017, pp. 744-753.
[64] Ahlad Kumar, "Deblurring of motion blurred images using histogram of oriented gradients and geometric moments", Signal Processing: Image Communication, Vol. 55, 2017, pp. 55–65
[65] Y. Renzhen, Li. Xuelong,"Latent semantic concept regularized model for blind image deconvolution", Neurocomputing, Vol.257, 2017, pp. 206–213.
[66] J.Pan, D.Sun, H.Pfister, and M-H.Yang , "Deblurring Images via Dark Channel Prior", IEEE Trans. Pattern Anal. Mach Intell., 2017, 0162-8828 .
[67] A.Laghrib, A.Ghazdali , A.Hakim, S.Raghay,"A multi-frame super-resolution using diffusion registration and a nonlocal variational image restoration", Computers and Mathematics with Applications, Vol.72,2016, pp. 2535–2548.
[68] X.Jinsheng, P.Guanlin , Y.Zhang, Y.Kuang , Y.Yan, Y.Wang, "Adaptive shock filter for image super-resolution and enhancement", J. Vis. Commun. Image R. Vol.40, 2016, pp.168–177.
[69] I-E.Mourabit, M-E.Rhabi, A.Hakim, A.Laghrib,, E.Moreau,"A new denoising model for multi-frame super-resolution image reconstruction", Signal Processing, Vol.132,2017, pp.51–65.
[70] C.Singh, A.Aggarwal,"An efficient and robust multi-frame image super-resolution reconstruction using orthogonal Fourier-Mellin moments" Displays, Vol.49, 2017, pp. 101–115.
[71] Y. Chen and B. Wu, "A multi-plane approach for text segmentation of complex document images," Pattern Recognit., vol. 42, no. 7, 2009, pp. 1419–1444.
[72] C. Merino and M. Mirmehdi, "A framework towards realtime detection and tracking of text," in Proc. 2nd Int. Workshop Camera-Based Document Anal. Recognit. (CBDAR), 2007, pp. 10-17.
[73] R. Minetto, N. Thome, M. Cord, N. J. Leite, and J. Stolfi, "Snoopertrack: Text detection and tracking for outdoor videos," in *Proc. 18th IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2011, pp. 505–508.
[74] C. Merino-Gracia and M. Mirmehdi, "Real-time text tracking in natural scenes," *IET Comput. Vis.*, vol. 8, no. 6, Dec. 2014, pp. 670–681.
[75] D.Karatzas, L.Gomez-Bigorda, A.Nicolaou, S.Ghosh, A.Bagdanow, M.Iwa-mura, J.Matas, L.Neumann, V.R.Chandrsekhar, ICDAR 2015 competition on robust reading, In: Proceedings of ICDAR, 2015, pp.1156–1160.