Video Scene Text Deblurring Using Text-Specific Multiscale Dictionaries: A Survey

Puja Chavan, Abhijeet Cholke, Prabhudev S

Abstract—This paper is concerned with the study of scene text detection and recognition from blurry natural video scene i.e. blurred images extracted from video which is important for image related purposes. In digital imaging, it's a challenging task to restore a clear image from a single motion-blurred image due to camera shake. E.g. includes image retrieval, comparative studies of different images, as well as the tool that aid visually impaired individuals to access the pictorial information. Particularly when we use handheld cameras to capture natural scene images, a general problem, i.e., blur, frequently happens. There are several techniques available for text detection and recognition. For representation of the text and non-text fields, a string of text-specific multi-scale dictionaries (TMD) and a natural scene dictionary is used separately. The TMD-based text field reconstruction makes easier to deal with the different scales of strings in a blurry image effectively. A lot of work has been done for detecting text in images and a lot has to be done. In this survey, we extend an existing end-to-end solution for text detection and recognition in natural images to video.

***Index Terms*—Scene Text, Text detection, Image De-blurring, multi-scale dictionaries, Text localization.**

I. Introduction

In digital imaging, it's a challenging task to restore a clear image from a single motion-blurred image due to camera shake. Motion blur caused by camera shake has been one of the prime causes of poor image quality in digital imaging, especially when using telephoto lens or using long shuttle speed. In past, many researchers have been working on recovering clear images from motion-blurred images. Here we face the problem of recovering the clear scene text by exploiting the text field characteristics. A series of text-specific multi-scale dictionaries (TMD) and a natural scene dictionary is learned for separately modeling the priors on the text and non-text fields. [1]

Puja Chavan, *Department of computer engineering, VACOE, Ahmednagar, University of Pune, Maharashtra, India*,

Abhijeet Cholke, *Department of Computer Technology, P.Dr.V.V.Patil Institute of Tech. &Engg.(Polytechnic), Loni, MSBTE, Maharashtra, India*,

Prabhudev S *Department of computer engineering, VACOE, Ahmednagar, University of Pune, Maharashtra, India*.

Taking handheld photos in low-light conditions is challenging. Since less light is available, longer exposure times are needed – and without a tripod, camera shake is likely to happen and produce blurry pictures. Increasing the camera light sensitivity, i.e., using a higher ISO setting, can reduce the exposure time, which helps. But it comes at the cost of higher noise levels. Further, this is often not enough, and exposure time remains too long for handheld photography, and many photos end up being blurry and noisy [2].

Many deblurring techniques are used to improve the visual quality of images, which also play an important role in text recognition and image understanding. At the point when catching characteristic scene pictures, particularly by handheld cameras, a typical antiquity, i.e., obscure, To enhance the visual nature of such pictures, deblurring methods are sought, which likewise play a vital part in character acknowledgment and picture understanding oftentimes happens.

Text detection existing methods are mainly divided into 3 types. They are mainly sliding window based method [3], connected component based method [4], and hybrid method [5]. Sliding window based method also known as region based method here it uses a sliding window in order to obtain the region. In connected component based method it mainly uses connected component analysis to extract text. Combination of both sliding window and connected component method are used in hybrid method.

Text recognition may be possible by several methods OCR [6], can be used for recognition purposes. Optical character recognition is used to convert text in images into editable text. It recognizes the character by optical mechanisms. Several other methods are available for recognition purposes.

II. Literature Review

C. Yi, X. Yang, and Y. Tian [7] offers a whole implementation estimation of image based STC recognition, by using evaluating diverse sampling techniques, characteristic descriptors, dictionary sizes, coding and pooling schemes, and SVM kernels. They methodically examine the have an impact on of each option within the characteristic illustration and category. The estimation final results on two datasets CHARS74K and ICDAR2003 show that Histogram of Orientated Gradient (HOG) descriptor, soft- assignment coding, max pooling, and Chi-Square Support Vector Machines (SVM) gain the satisfactory overall performance among local sampling primarily based function representations. To higher STC recognition, they observe worldwide sampling function illustration. They produce worldwide HOG (GHOG) by means of computing HOG descriptor from global sampling. GHOG allows enhance character structure modeling and attain better performance than local sampling primarily based function representations. The GHOG outperforms available techniques within the two benchmark datasets.

J.-F. Cai, H. Ji, C. Liu, and Z. Shen [8] proposed an method to take away motion blurring from a single image with the aid of formulating the blind blurring as a brand new joint optimization hassle, which simultaneously maximizes the sparsity of the blur kernel and the sparsity of the clean photograph below certain appropriate redundant tight body systems (curvelet device for kernels and framelet device for images). Without requiring any previous records of the blur kernel as the input, their proposed approach is capable of get better extremely good snap shots from given blurred images. furthermore, the new sparsity constraints underneath tight frame structures allow the application of a fast set of rules referred to as linearized Bregman new release to successfully solve the proposed minimization trouble. The experiments on both simulated snap shots and actual images confirmed that their algorithms can effectively casting off complex motion blurring from nature images.

R. Fergus, B. Singh, A. Hertzmann, S. T. Roweis, and W. T. Freeman [9] brought a method to dispose of the consequences of digital camera shake from significantly blurred images. The approach assumes a uniform digicam blur over the picture and negligible in-aircraft digicam rotation. so that it will estimate the blur from the camera shake, the consumer need to specify an image region without saturation effects. They display consequences for a variety of virtual pictures taken from private photo collections.

Next method is via A. Levin, Y. Weiss, F. Durand, and W. T. Freeman [10] deals with analyze and compare recent blind deconvolution algorithms each theoretically and experimentally. They provide an explanation for the previously reported failure of the naive MAP method with the aid of demonstrating that it mostly favors no-blur factors. Also they demonstrate that since the kernel size is frequently smaller than the image size a MAP estimation of the kernel alone can be well constrained and correctly improve the true blur.

Q. Shan, J. Jia, and A. Agarwala [11] proposed a novel algorithm for removing motion blur from a single image. Their technique computes a deblurred image using a unified probabilistic model of both blur kernel estimation and unblurred image restoration. They present an analysis of the causes of common artifacts found in current deblurring methods, and then introduce several novel terms within this probabilistic model that are inspired by our analysis. These terms include a model of the spatial randomness of noise in the blurred image, as well a new local smoothness prior that reduces ringing artifacts by constraining contrast in the unblurred image wherever the blurred image exhibits low contrast. Finally, they describe an efficient optimization scheme that alternates between blur kernel estimation and unblurred image restoration until convergence. As a result of these steps, we are able to produce high quality deblurred results in low computation time. They are even able to produce results of comparable quality to techniques that require additional input images beyond a single blurry photograph, and to methods that require additional hardware.

Wexler method [12] mainly deals with detection of text with the help of stroke width transform. It is a novel image operator to find the value of stroke width. Stroke Width Transform computes width of stroke containing pixel. First compute edges in image using Canny edge detector. It considers gradient direction of pixel p, when p lies on stroke boundary and perpendicular to orientation of stroke.

Chucai Yi et al [13] propose a novel method to detect text structure by structure based partition and grouping. Framework to detect text strings with arbitrary orientations to locate text region in images based on image partitioning and connected component grouping. Steps are a) choose candidate text character from connected component by gradient and color feature b) character grouping is performed to combine the character into string. Text line grouping is able to extract the string with arbitrary orientation.

III. LIMITATIONS OF EXISTING SYSTEMS

There are numerous methods of image deblurring that should be investigated and numerous future research examinations stay to be led in this generally new area. Several limitations can be found in this technique. The previous systems cannot resolve the image caused by fast moving objects and unsuitable for text images. The method used is applicable to landscape images not for text images [8]. Scene Text Character (STC) recognition, which generally includes feature representation to model character structure, cannot recognize character to word & also less robust and less effective feature representations to improve STC recognition performance[7].

Locating text from a complex background with multiple colors is a challenging task & also developing learning based methods for text extraction from complex backgrounds and text normalization for OCR recognition is also much tedious task [13]. We also attempt to improve the efficiency and transplant the algorithms which are applicable for natural blurry scene images which further give robust and improved results [10, 11, 12].

IV. DEBLURRING TECHNIQUES

Lucy Richardson algorithm techniques: The algorithm, also known as Richardson–Lucy de convolution, is an iterative procedure for recovering a latent image that has been blurred by a known PSF. Point Spread Function (PSF) is the degree to which an optical system blurs (spreads) a point of light. The PSF is the inverse Fourier transform of Optical Transfer Function (OTF). Inside the frequency area, the OTF describes the response of a linear, position-invariant system to an impulse. OTF is the Fourier transfer of the point (PSF).

Blind de convolution techniques: There are basically two types of de convolution methods. They are projection based blind de convolution and maximum likelihood restoration. In the first approach it simultaneously restores the true image and point spread function. This begins by making initial estimates of the true image and PSF. The technique is cylindrical in nature. Firstly we will find the PSF estimate and it is followed by image estimate. This cyclic process is

ISSN: 2278 – 7798

repeated until a predefined convergence criterion is met. The merit of this method is that it appears robust to inaccuracies of support size and also this approach is insensitive to noise.

Wiener de blurring techniques: A blurred image can be modeled by 2D convolution of the source image and a low pass filter. In order to restore the source image, the blurred image should convolute with a type of high pass filter. In ideal cases, the high pass filter can be determined by inverting all elements of discrete Fourier transform (DFT) of the low pass filter. In order to avoid near zero values to be inversed, a suitable threshold is necessary. Essentially an inverse filter is a high pass filter which is sensitive to noises, in this case, the Wiener filter is selected for image de blurring, which is a natural extension of the inverse filter when noises are present. [6]

V. PROPOSED SYSTEM

The overview of proposed method we learn a natural scene dictionary and a series of text-specific multi-scale dictionaries to model the priors on the background scene and text fields respectively. This step is dictionary learning highlighted by the red dashed rectangle. Second, we run the state-of-the-art text localization method to differentiate text fields from non-text ones. Third, based on TMD and the natural dictionary, we construct the dedicated priors for real world text and non-text fields. As a result, we optimize the cost function to estimate the blur kernel and the latent image. The last step will repeat until the blur kernel converges as shown in figure 1. In our proposed approach we are using video frames in case of blur images. Our method detects sharp regions in the video, and uses them to restore blurry regions of the same content in nearby frames. Proposed method will deblurs blurry videos in dynamic scenes, so get better solution for overcoming the limitations of existing systems.

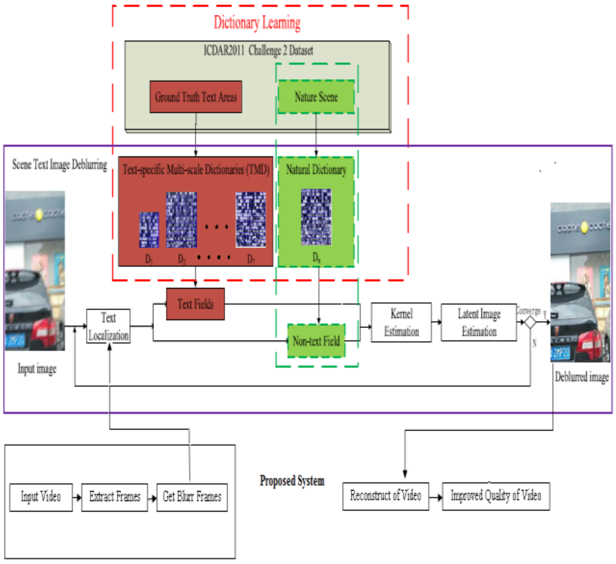

Figure 1: Proposed Method

VI. CONCLUSION

In this paper, we extend an end-to-end text detection and recognition solution to the video domain. By considering the pros and cons of different methods for Video Scene Text Deblurring we can choose the effective one. The existing system fails to deblur blurry videos in dynamic scenes, so our proposed method will successfully overcome this drawback.

REFERENCES

[1] Xiaochun Cao, Wenqi Ren, Wangmeng Zuo, Xiaojie Guo, Hassan Foroosh, "Scene Text Deblurring Using Text-Specific Multiscale Dictionaries", IEEE Transactions on Image Processing, Vol. 24, No. 4, April 2015, pp 1302-1314.

[2] Lin Zhong, Sunghyun Cho, Dimitris Metaxas, Sylvain Paris, Jue Wang, "Handling Noise in Single Image Deblurring using Directional Filters", IEEE Conference on Computer Vision and Pattern Recognition, 2013, pp 612-619.

[3] K. Kim, K. Jung, and J. Kim, "Texture-base approach for text detection in images using support vector machines and continuously adaptive mean shift algorithm," IEEE Trans. Pattern Anal. Mach. Intell., vol. 25, no. 12, pp. 1631–1639, Dec. 2003

[4] Y.-F. Pan, X. Hou, and C.-L. Liu, "A hybrid approach to detect and localize texts in natural scene images," IEEE Trans. Image Process, vol. 20, no. 3, pp. 800–813, Mar. 2011.

[5] C. Yi and Y. Tian, "Text string detection from natural scenes by structure-based partition and grouping," IEEE Trans. Image Process., vol. 20, no. 9, pp. 2594–2605, Sept. 2011.

[6] Sonia George, Noopa Jagdeesh, "A Survey on Text Detection and Recognition from Blurred Images" Conference on Innovative Engineering, Vol. II, Special Issue X, pp. 1180-1184, March 2015.

[7] C. Yi, X. Yang, and Y. Tian, "Feature representations for scene text character recognition: A comparative study," in Proc. 12th ICDAR, Aug. 2013, pp. 907 -911.

[8] J.-F. Cai, H. Ji, C. Liu, and Z. Shen, "Blind motion deblurring from a single image using sparse approximation," in Proc. IEEE Conf. CVPR, Jun. 2009, pp. 104–111.

[9] R. Fergus, B. Singh, A. Hertzmann, S. T. Roweis, and W. T. Freeman, "Removing camera shake from a single photograph," ACM Trans. Graph., vol. 25, no. 3, pp. 787 - 794, 2006.

[10] A. Levin, Y. Weiss, F. Durand, and W. T. Freeman, "Understanding and evaluating blind deconvolution algorithms," in Proc. IEEE Conf. CVPR, Jun. 2009, pp. 1964–1971.

[11] Q. Shan, J. Jia, and A. Agarwala, "High-quality motion deblurring from a single image," ACM Trans. Graph., vol. 27, no. 3, 2008, Art. ID 73.

[12] B. Epshtein, E. Ofek, and Y. Wexler, "Detecting text in natural scenes with stroke width transform," in Proc. IEEE Conf. CVPR, San Francisco, CA, USA, 2010, pp. 2963 - 2970.

[13] C. Yi and Y. Tian, "Text string detection from natural scenes by structure-based partition and grouping," IEEE Trans. Image Process., vol. 20, no. 9, pp. 2594 -2605, Sept. 2011.

[14] J.-F. Cai, H. Ji, C. Liu, and Z. Shen, "Framelet-based blind motion deblurring from a single image," IEEE Trans. Image Process., vol. 21, no. 2, pp. 562–572, Feb. 2012.

[15] H. Ji and K. Wang, "A two-stage approach to blind spatially-varying motion deblurring," in Proc. IEEE Conf. CVPR, Jun. 2012, pp. 73–80.

[16] L. Zhong, S. Cho, D. Metaxas, S. Paris, and J. Wang, "Handling noise in single image deblurring using directional filters," in Proc. IEEE Conf. CVPR, Jun. 2013, pp. 612–619.

[17] M. Elad and M. Aharon, "Image denoising via sparse and redundant representations over learned dictionaries," IEEE Trans. Image Process., vol. 15, no. 12, pp. 3736–3745, Dec. 2006.

[18] M. Elad, M. A. T. Figueiredo, and Y. Ma, "On the role of sparse and redundant representations in image processing," Proc. IEEE, vol. 98, no. 6, pp. 972–982, Jun. 2010.

[19] H. Zhang, J. Yang, Y. Zhang, N. M. Nasrabadi, and T. S. Huang, "Close the loop: Joint blind image restoration and recognition with sparse representation prior," in Proc. IEEE ICCV, Nov. 2011, pp. 770–777.

[20] H. Cho, J. Wang, and S. Lee, "Text image deblurring using text-specific properties," in Proc. 12th ECCV, 2012, pp. 524–537.

