Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 96 (2016) 744 – 753

Procedia Computer Science

20th International Conference on Knowledge Based and Intelligent Information and Engineering Systems, KES2016, 5-7 September 2016, York, United Kingdom

Scene text deblurring in non-stationary video sequences

Margarita Favorskaya*, Vladimir Buryachenko

Siberian State Aerospace University, 31 Krasnoyarsky Rabochy av., Krasnoyarsk, 660037 Russian Federation

Abstract

Text detection in natural scenes burdened by imperfect shooting conditions and blurring artifacts is the subject of the present paper. The text as a linguistic component provides a significant amount of information for scene understanding, scene categorization, image retrieval, and many other challenging problems. Usually real video sequences suffer from the superposition of the complicated impacts that are often analyzed separately. The main attention focuses on the text detection with geometric and blurring distortions under blurring and camera shooting artifacts. The original 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. Also, the blind technique of a blurred text restoration is discussed. Additionally some results of the text detection are mentioned. The detection results for corrupted text fragments from test dataset ICDAR 2015 achieve 76–83% and prevail the detection results of the non-processed by deblurring procedure text fragments upon 40–52%.

Peer-review under responsibility of KES International

Keywords: text deblurring; scene text detection; non-blind kernel; blind kernel; non-stationary video sequence

1. Introduction

Scene text detection in a still image or a sequence of images plays a significant role in image annotation, image indexing, traffic sign recognition, license plate recognition, signboards and information board recognition as the assistance to old people, among others. The scene text is a part of an image or frame unlike the imposed artificial text, such as subtitles, logotypes, information about sport competitions, and so on. The task of scene text detection and recognition meets with many challenging impacts, including text attributes (font sizes, alignment, colors), luminance in a scene (shadow, brightening, contrast), geometric location (orientation, perspective), scene complexity

* Corresponding author. Tel.: +7-391-291-9240; fax: +7-391-91-9147.
E-mail address: favorskaya@sibsau.ru

Peer-review under responsibility of KES International
doi:10.1016/j.procs.2016.08.259

(cluttered background, moving objects), and visual artifacts (noise, distortions due to a blur). The distortion due to a motion blur or defocusing is a major issue among other impacts because the blurring changes the shape significantly. Moreover, the most of methods, which are useful to detect and recognize the inherent and imposed text, fail to give satisfactory results under blur influence, e. g., conventional feature-based, texture-based, and edge-based methods.

It is well-known that some intrinsic or extrinsic factors may lead to a quality degradation of images, among which blurring is the main one. The situation is further complicated by superposition of the shooting conditions with the camera shakes and jitters during, for example, a hand-held shooting. Five reasons of a blurring appearance can be mentioned (the first four items are the extrinsic reasons, while the last one is the intrinsic reason):

- A blurring is caused by fast object motion like a moving car. This type is called a motion blur.
- The blurred images may appear from the required long exposure time in dark lighting conditions, when a hand-held camera cannot be stabilized well. This type is called a camera shake blur.
- A defocus blur emerges in the resultant images obtained from visual equipment having a single focus plane.
- Some atmospheric turbulence impacts, such as fog or smoke, produce a blurred image.
- The cause of an intrinsic physical blur lies in a lens system, if the lenses have different refractive indices for different wavelengths of light.

Even though an image blurring in photography may be motivated as an aesthetic representation, the blurred images in computer vision are often considered as the corrupted ones. Also hereinafter, it is reasonable to consider the camera shakes and jitters as a wider extrinsic reason, provoking 3D geometric distortions and, in particular case, a camera shake blur. In the current research, various cases, including types of blurring, degree of blurring, and degree of blur coverage, are studied under the non-stationary conditions of camera shooting.

In the remainder of this paper, Section 2 presents a literature survey of the related works. Text properties and limitations are analyzed in Section 3. The proposed methodology of the scene text detection under blurring and shooting artifacts is described in Section 4, while Section 5 provides a deblurring technique of the extracted letters. The comparative experimental results are drawn in Section 6. The conclusions in Section 7 complete the paper.

2. Related Work

The study of the image blurring caused by defocusing and/or diffraction began since 1990s. Some mathematical models were built based on the Gaussian kernel[1] in order to improve a human depth-from-blur perception. At the same time, the image deblurring as the difficult and useful for practice task attracts many researchers. A good analytical survey of the deblurring methods was presented by Wang and Tao[2], according to which the most deblurring methods are grouped into five categories, such as the Bayesian inference framework, variational methods, sparse representation-based methods, homography-based modelling, region-based methods, and other methods:

- A probability hypothesis in the Bayesian framework can be adapted to estimate the imposed uncertainty attributes on either the unknown sharp image or the unknown blur kernel, or both. The commonly-used estimators are the Maximum A Posteriori (MAP), minimum mean square error[3], and variational Bayesian methods[4]. Thus, in the MAP approach, a classic non-blind algorithm called as Richardson-Lucy (RL) deconvolution[5,6] can be mentioned. Levin et al.[4] show how to make the MAP estimation successfully recover the true blur kernel.
- Variational methods are typically used as approximation methods, incorporating the regularization techniques into a constraint space. In contrast to classical regularizers based on the first-order derivatives, nowadays the second-order regularization techniques are also developing in deblurring framework[7].
- Space representation was declared due to sparse properties of the natural scenes. It can be applied in many computer vision applications, such as denoising, inpainting, superresolution, and deblurring[8].
- Homography-based modelling was generally proposed to simulate the blur effect induced by the camera shake as a spatially variant deblurring using a set of the multiple kernels or homographies[9]. Usually this approach is amplified by the “temporal” homographies, if it is possible.
- Region-based methods create the blur model based on the local consistent kernels in each region, even at each point. The efficient filter flow model was proposed by Hirsch et al.[10]. In the case of the object motion blur, the