J. Vis. Commun. Image R. 44 (2017) 148–155

Contents lists available at ScienceDirect

J. Vis. Commun. Image R.

journal homepage: www.elsevier.com/locate/jvci

Novel video stabilization for real-time optical character recognition applications ☆

Yun Gu Lee

Kwangwoon University, Department of Computer Science and Engineering, Kwangwoon-ro 20, Nowon-Gu, Seoul 139-701, South Korea

ARTICLE INFO

Article history:
Received 17 May 2016
Revised 11 January 2017
Accepted 26 January 2017
Available online 1 February 2017

Keywords:
Video stabilization
Optical character recognition
Real-time

ABSTRACT

This paper presents a novel video stabilization algorithm for real-time optical character recognition (OCR) applications. The proposed method generates output frames in order to stabilize the position of a target word that will be recognized by the OCR application. Unlike in conventional algorithms, in the proposed algorithm, a causal low pass filter is not applied to the trajectory of the target word for reducing the high frequency component of camera motion. The proposed algorithm directly calculates the stable position of the word using two forces: the force used to pull the target word to the center of an output frame and a back force used to return the center of an output frame to the center of an input frame. Hence, the proposed algorithm significantly minimizes the time take to respond to sudden camera movement. Although the proposed method may not outperform state-of-the-art video stabilization in terms of video stability, the proposed technique is much more appropriate for real-time OCR applications than the conventional techniques in terms of accuracy, computational cost, processing delay, and the time taken to respond to sudden camera movement. Simulation results prove the superiority of the proposed method over conventional techniques for real-time OCR applications.

1. Introduction

The dramatic development of digital imaging technologies has increased the popularity of hand-held video cameras. Nowadays, even a high-end smartphone is equipped with a video camera capable of recording an ultra-HD video with a frame rate of 30 Hz. Besides taking pictures and videos, the smartphone camera is used for many applications, such as augmented reality, barcode reader, object recognition application (Google Goggles), and games. An interesting application is the real-time optical character recognition (OCR) in which characters (or a word) are recognized in real-time in a video captured by a smartphone camera.

In an OCR system, an image is passed through the process of binarization, line segmentation, word segmentation, character segmentation and character recognition [1]. Detection of textual regions from natural images is a challenging issue due to the variations in background, contrast, text type, font type, font size, and so on [2,3]. In the recent literature, the spatial domain characteristics of character images have been extensively studied such as perceptual quality assessment of screen content images [4] and saliency-guided quality assessment of screen content images [5]. Especially, Wang et al. provide comprehensive analyses on the unique characteristics of screen content images, including the frequency energy falloff statistics, sharpness of edges and free-energy principle based image uncertainty [6].

The real-time OCR continuously recognizes characters (or a target word) within a region of interest (ROI) in a video. A marker is usually used to point out the target word, which is displayed at the center of the preview screen. The smartphone is moved to locate the marker on the target word. However, the user's shaking hand can induce camera motion, and the video displayed on the preview screen becomes shaky. The marker will then not be able to accurately point out the target word. However, video stabilization techniques can be used to stabilize the video on the preview screen by alleviating undesirable camera motion.

Many video stabilization techniques have been developed to provide stable videos such as camera stabilization, optical image stabilization (OIS) and digital image stabilization (DIS). The camera stabilizer such as Gimbal compensates the orientation of a camera by detecting camera rotation. The OIS uses a floating lens element (or an image sensor) that is moved orthogonally to the optical axis of the lens in such a way as to compensate the camera motion. The camera stabilizer and OIS usually predict the camera motion using a gyroscope that detects the angular velocity. Hence, the camera stabilizer and OIS cannot detect pure translational camera motion.

☆ This paper has been recommended for acceptance by M.T. Sun.
E-mail address: harmony96@gmail.com

http://dx.doi.org/10.1016/j.jvcir.2017.01.027
1047-3203/

In the real-time OCR applications, a user translationally moves the smartphone while looking at the preview screen. Therefore, the capturing device for OCR applications may function better with the DIS approach to the camera stabilizer and the OIS approaches using the gyroscope.

The DIS algorithms are mainly categorized into 2D and 3D approaches. Most 2D approaches adopt 2D motion models such as the simple translational model, 2D rigid model, and 2D affine model [7–9]. As the camera motion increases, the approximation errors of the 2D motion models increase, and the methods cannot easily describe the geometric transform between two successive frames. The 3D approach essentially provides better performance than the 2D approach. One of the most significant challenges of the 3D approach is dealing with the parallax occlusion problem. The 3D approach sometimes introduces image artifacts such as a ghost effect on stabilized images. In order to reduce the image artifact, Liu introduced a content-preserving warp [10] by relaxing the constraint for a physically correct reconstruction while preserving the perceptual quality. Wang proposed a robust and efficient technique to achieve high-quality camera motion on a video where 3D reconstruction is difficult or long feature trajectories are not available [11]. Grundmann et al. proposed calibration-free video stabilization and rolling shutter compensation, based on a mixture model of homographies [12]. The above methods provide very stable videos of a quality similar to that taken with professional devices.

The video stabilization method is highly relevant to the viewing conditions. For the purpose of video recording, once a video is captured, it can be stabilized and re-encoded. In that viewing condition, computational burden is not a critical issue, and most conventional video stabilization with complex motion models can provide outstanding performance in terms of display performance. However, in OCR applications, captured frames should be displayed on the preview screen immediately. Hence, the video stabilization should consider accuracy, complexity, processing delay, and the response time to the sudden camera movement, which will be discussed in detail in Section 2. Therefore, this paper proposes a new video stabilization algorithm for the purpose of real-time OCR applications. Although the proposed algorithm may not outperform state-of-the-art video stabilization in terms of video stability, it is more suitable for real-time OCR applications. Another application of the proposed algorithm is a magnifier to enlarge and enhance text on the phone screen for making people easier to see, as shown in Fig. 1. Since the preview screen is magnified, even small amount of the user's shaking hand makes the preview screen very shaky. The proposed method can be used to stabilize the video on the preview screen.

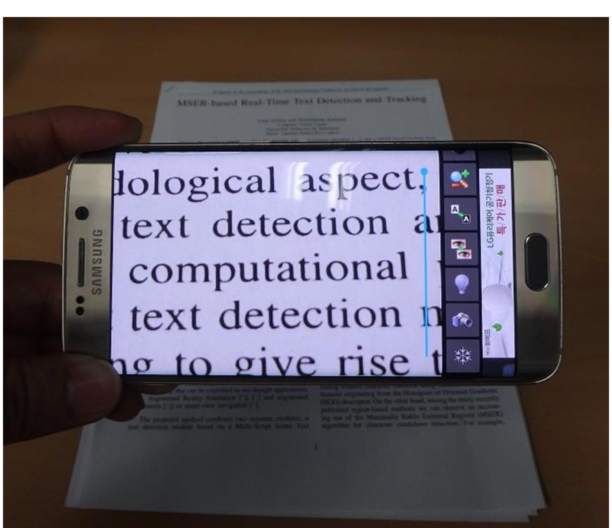

Fig. 1. Magnifier application for Google Android.

This paper is organized as follows. Section 2 will discusses the limitation of conventional video stabilization and the requirements of video stabilization for real-time OCR. Section 3 will introduces the proposed algorithm. Simulation results are will be given in Section 4. and Finally, Section 5 will concludes the paper.

2. Problem definition

2.1. Limitation of existing video stabilization methods for the purpose of optical character recognition

Conventional video stabilization methods may not be suitable for real-time OCR applications for mobile devices. The first issue is the processing delay. Conventional methods usually employ feature extraction and tracking techniques. Since objects appear and disappear over time, the feature points do not typically last a long time along the frames. For example, Liu uses only the feature points that last for at least 20 frames to solve the problem [10]. Therefore, the conventional methods should buffer a sufficient number of input frames to check whether or not the feature points last, which induces a processing delay. Then, the current frame cannot be displayed on the preview screen at the right time. For example, while smartphones such as Apple iPhone and Samsung Galaxy support video stabilization, they do not display the stabilized video on the preview screen. The video stabilization is applied on the recorded video only. Video stabilization for OCR applications should be capable of displaying the stabilized video on the preview screen without delay.

The second issue is the time taken to response to sudden camera movement. Conventional video stabilization methods usually use a low-pass filter to remove the high-frequency jitter of camera motion. For example, one method can calculate a new camera trajectory by applying a low-pass filter to the original camera trajectory, where the camera trajectory represents a position of a camera with time. Many approaches can be used to compute the trajectory, such as dense trajectory [13] and scale invariant feature transform (SIFT) flow trajectory [14]. The camera trajectory can be directly predicted from the structure-from-motion [10], or simply calculated by accumulating the predicted camera motion between consecutive frames [15]. Since an ideal low pass filter having infinite delay cannot be used in a real-time system, outputs of the low-pass filter should depend on past and current inputs but not on future inputs. This causal low-pass filter induces a slow response to sudden camera movement as shown in Fig. 2. In the figure, the x and y axes represent a frame number and the position of a camera, respectively Although the stabilized trajectory #2 significantly removes high frequency components of the original trajectory, it cannot immediately follow the original camera trajectory. It will then appear as though while the preview screen is very stable, it does not update the views corresponding to camera movement. A weak low pass filter may shorten the response time, as the stabilized trajectory #1 in the figure. However, it will not effectively remove high frequency components of the original trajectory. Also, there is still a considerable time gap between the original and stabilized trajectories around $\boldsymbol{R}$ in the figure. Video stabilization for OCR applications is required to provide fast response to sudden camera movement.

The third issue is their computational burden. Conventional methods adopt computationally expensive tools such as feature extraction and tracking, structure-from-motion, image warping, and view interpolation, which may not be appropriate for mobile