Computer Vision and Image Understanding 160 (2017) 45–56

Contents lists available at ScienceDirect

Computer Vision and Image Understanding

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

Model-based motion blur estimation for the improvement of motion tracking

Clemens Seibold[a,*], Anna Hilsmann[a], Peter Eisert[a,b]

[a] Fraunhofer HHI, Einsteinufer 37, 10587 Berlin, Germany
[b] Humboldt University Berlin, Unter den Linden 6, 10099 Berlin, Germany

ARTICLE INFO

Article history:
Received 29 July 2016
Revised 23 December 2016
Accepted 29 March 2017
Available online 31 March 2017

Keywords:
Motion blur
Multi-frame deconvolution
Motion tracking

ABSTRACT

Video tracking is an important task in many automated or semi-automated applications, like cinematic post production, surveillance or traffic monitoring. Most established video tracking methods fail or lead to an inaccurate estimate when motion blur occurs in the video, as they assume, that the object appears constantly sharp in the video. In this paper, we present a novel motion tracking method with explicit modeling of motion blur, estimating the continuous motion of a rigid 3-D object with known geometry in a monocular video as well as the sharp object texture. Instead of treating motion blur as a potential source of errors, we take advantage of it and consider motion blur as an additional information source, providing information about the motion of the tracked object during the exposure. In an analysis-by-synthesis approach we explicitly model the effects of motion blur reconstructing the captured frames, in order to accomplish a more accurate estimation. We design our algorithm to be capable to run in parallel on the GPU using the common rendering pipeline and considering each frame individually to handle also long videos. We tested our approach on both synthetic and real videos. In both cases, we achieve significant improvements of accuracy and reductions of frame reconstruction error compared to the estimated motion of a rigid body tracker, without motion blur handling.

1. Introduction

Motion blur is a frequently occurring effect in video recording, arising whenever an object is moving relative to the camera during the exposure time. Its strength depends on the duration of the exposure as well as on the speed of the object relative to the camera. Thus, it often arises significantly in cinematic productions, where long exposure times are common, and in amateur videos, which are mainly captured with shaking hand-held cameras. Today's high video resolution further increases the effect of motion blur, making it more important to take this effect into account when performing typical computer vision tasks like motion tracking or pattern recognition. The conventionally employed methods for these tasks assume constant image characteristics for object points in each frame and thus fail if these assumptions are violated, for example, when striking edges perpendicular to the motion direction are low-pass filtered due to motion blur.

In this paper, we introduce a new model-based motion estimation approach for monocular videos with explicit modeling of motion blur. On the one hand, explicit modeling of motion blur makes the motion estimation more robust and on the other hand, it allows us to exploit the information that motion blur carries about motion during the exposures in order to achieve more accurate results. Finally, it allows us to reconstruct a sharp and improved texture of the object. In order to describe the object's motion, we employ a continuous function to describe the object's pose during the whole video. This function separately describes the trajectory of the object's center using a cSpline and the orientation of the object using spherical linear interpolation. The object's texture is described by a 2-D texture map and a mapping from the 3-D coordinates of the object to this 2-D texture map. We assume that the object's geometry is known and use a triangle mesh for its representation.

This paper is organized as follows. Section 2 gives an overview of recent work on motion blur estimation and modeling. Section 3 outlines our motion blur and texture estimation framework and formalizes the motion model. The estimation of the sharp texture is described in Section 4. In Section 5, the estimation

* Corresponding author.
E-mail address: clemens.seibold@hhi.fraunhofer.de (C. Seibold).

http://dx.doi.org/10.1016/j.cviu.2017.03.005

of the motion model parameters and thus the motion and motion blur estimation is presented. In Section 6, the introduced approach is tested on synthetic and real videos. Finally, results, advantages, and limitations of the introduced approach are discussed.

2. Related work

Motion blur estimation, utilization and compensation have been examined in numerous publications with various requirements and aims. In the following, we give a short overview of different methods estimating or utilizing motion blur in monocular videos or images as well as their applications.

Single-image methods. In image processing, blind deconvolution is the estimation of a sharp image and a blur kernel given a blurry input image. The blur kernel describes the blur caused by the capturing process, without explicit knowledge about this process. One of the first blind deconvolution methods addressed problems in seismology (Wiggins, 1978) and astronomy (Bates, 1982) and took advantage of the special properties of these images, like minimal entropy in the underlying sharp image or a uniform background. Other early methods rely on the absence of noise, assume that the kernel can be described by a parametric model or have hard assumptions on the image content, e.g. blur invariant edges. For an extensive survey on these methods see Kundur and Hatzinakos (1996). Recent methods focus on the deconvolution of still images, taken with hand-held devices, with completely different image characteristics and require other deconvolution methods. These methods are often based on image priors derived from natural image statistics. Exploiting these priors, Fergus et al. (2006) developed a blind deconvolution approach based on a Variational Bayesian method. Shan et al. (2008) proposed a method based on a maximum a posteriori probability (MAP) estimator. While the Variational Bayesian method estimates the blur kernel regardless of an estimate for the sharp image, the MAP estimator combines the blur kernel and sharp image estimation. The data-adaptive kernel regression technique of Takeda et al. (2008) is capable of simultaneously denoising and deblurring images. In contrast to Fergus et al. (2006) and Shan et al. (2008) they focused on blur kernels, which can be described by a Gaussian or moving average filter. A model-free deblurring method was developed by Kim and Lee (2014). They estimate a 2-D motion field describing motion blur and underlying sharp image based on a TV-L1 model. Gast et al. (2016) combined foreground segmentation and estimation of non-uniform motion blur resulting in robust foreground segmentation also in the presence motion blur. Other recent deconvolution methods take advantage of step edges in the underlying sharp scene (Cho et al., 2011) or estimate the motion blur kernels for moving objects in front of a static scene using a non-binary alpha mask (Dai and Wu, 2008). Also, Convolutional Neural Networks were successfully used to estimate patch wise motion blur kernels (Sun et al., 2015) and for the deconvolution of images (Xu et al., 2014).

Template-matching methods. In template matching-based tracking, a warping function transforms a template to fit it into an image with a minimal difference between the transformed template and the image. Common template-matching algorithms like Benhimane and Malis (2004); Lucas and Kanade (1981) can fail if the template and the image are blurred in different ways. Jin et al. (2005) study this problem in order to achieve a robust template matching, in which both, the template and the target image, can be blurred by motion. In addition to the affine warping, defining the position of the template in the image, they estimate a Gaussian motion blur kernel for the template as well as the target image. Park et al. (2009; 2012) modify the ESM algorithm to consider motion blur and be able to perform motion blur-aware template matching in real-time and use it for augmented reality tasks. Their main contributions are a more flexible blur kernel, which can also model rotational blur, and an efficient calculation of the motion during the exposure.

Multi-image methods. If more than one image is used to estimate the motion or underlying sharp scene, on the one hand more information is provided, but on the other hand fusion of these information must be considered. Bascle et al. (1996) fuse the information of several frames in order to deblur videos and even get a high resolution image. They describe the motion between frames by homographies and use the same homographies to describe the motion during the frame. Thus, they require the relative exposure to calculate the pixel motion during the exposure, which is needed for the deblurring process. Li et al. (2010) expand this approach by the estimation of the relative exposure time and adapt it for the generation of sharp panoramas. Cai et al. (2009) estimate a static sharp scene from multiple blurred frames by using one spatially invariant kernel per frame to describe the motion blur and perform the deblurring in framelet domain. Other methods, Lee et al. (1997), Onogi and Saito (2005) and Cho et al. (2012), divide the frames into several pieces and describe the motion individually for each piece. While Lee et al. (1997) and Onogi and Saito (2005) deblur the pieces to estimate sharp frames, Cho et al. (2012) use Lucky Imaging (Fried, 1978): They assume that for each patch in a frame, there is at least one patch in the video containing the same region but without motion blur. Kim and Lee (2015) move away from a parametric model to describe the motion for a whole image or at least connected components in an image and estimate a pixel-wise motion and thus motion blur kernel. Sorel and Flusser (2008) and Lee and Lee (2013) study the problem of motion blur in depth estimation of static scenes. Sorel and Flusser (2008) estimate in a first step the motion, describing it by a non-parametric kernel, which is estimated by a blind-deconvolution approach applied on a user selected region with negligible depth variation. Using this kernel, depth and underlying sharp scene is estimated in an iterative fashion. The method of Lee and Lee (2013) gets along without estimation of the underlying sharp scene and estimates only its depth. The camera motion is estimated in advance neglecting motion blur. Wulff and Black (2014) proposed a technique for combined segmentation and motion blur estimation in dynamic scenes. They segment the scene in a foreground and background and estimate an affine motion for both parts and the sharp underlying scene.

Methods with specialized capture setups. In case of a specialized capture process, the relation between the captured data can be defined and thus exploited without an additional estimation of the camera positions or temporal aspects. In most cases the capture process is designed, such that a short exposed sharp frame is captured during or immediately before or after a long exposed frame and the motion during the long exposed frame is estimated. Favaro and Soatto (2004) use two frames, with one being captured at a subinterval of the other, to estimate the motion, the underlying sharp scene and borders of several objects moving parallel to the image plane. Their experiments show that this setup can be simulated by capturing and averaging three consecutive frames. Sellent et al. (2009a; 2009b) proposed non-parametric approaches to estimate the motion during a long exposed image, utilizing two short exposed images, one captured immediately before and one immediately after the long exposed image. They describe the motion pixel-wise by a motion field, relative to the short exposed frames. The approach of Yuan et al. (2007) uses two images, one sharp but containing strong image noise and the other blurred