J. Vis. Commun. Image R. 38 (2016) 217–229

Contents lists available at ScienceDirect

J. Vis. Commun. Image R.

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

Robust motion estimation for night-shooting videos using dual-accumulated constraint warping ☆

Deepika Shukla [a,*], Rajib Kumar Jha [b]

[a] PDPM Indian Institute of Information Technology, Design and Manufacturing Jabalpur, Madhya Pradesh 482005, India
[b] Indian Institute of Technology Patna, Bihar 801103, India

ARTICLE INFO

Article history:
Received 24 February 2015
Revised 3 November 2015
Accepted 1 March 2016
Available online 2 March 2016

Keywords:
Dynamic time warping
Sakoe–Chiba band
Motion estimation
Image alignment
Video stabilization
Night shooting
Low light shooting

ABSTRACT

This paper introduces a novel concept of dual-accumulated constraint projection warping, as a robust and efficient motion estimation solution for night video stabilization. Small imaging-sensors used in compact hand-held cameras become very prone to noise and blur under low illumination condition. Restricted lighting results in dark boundaries and degrades textural information of the frame. Presence of these combined textural artifacts makes night-shooting a hard problem for accurate motion estimation. At poor lighting, local intensity variations result in failure of inter-frame feature or block matching correspondence. In the proposed technique, use of projection ensures accuracy under local perturbations, noise and blur conditions, while dual-accumulation eliminates the effect of dark-regions adding robustness to night-shooting condition. Efficiency of the proposed algorithm over the existing motion estimation techniques is tested and verified over different categories of night shooting videos. In addition to night video stabilization the proposed scheme also performs well under normal illumination.

1. Introduction

Night shooting requires good photography skills for efficient scene capturing, as even small body tremors may degrade the video quality significantly. Night shooting differs from other shooting environments in terms of its limited illumination source availability and leads to poor textural detailing. Various factors affecting the motion estimation accuracy in night-video stabilization are discussed here to elaborate the hidden challenges in the application. In night shooting, most of the visibility is achieved by the neighboring lighted sources present in or around the viewing area, e.g. in on-road shooting, the lampposts, road-lights and lighted vehicles contribute as illumination source. These videos because of limited lighting suffer from large amount of dark/black-out regions and lead to scene-content deficiency within the frame. Apart from this, the videos also suffer from the local intensity variation and motion blurring due to in-scene lighted object movements (e.g. as in case of night traffic-scene moving vehicles result in local intensity variation and give rise to motion blurring effect). In view of night shooting, the videos captured using compact hand-held/mobile camera face a constraint of small imaging sensors, which under poor illumination become very prone to the noise and uniform frame blurring.

Any video stabilization system consists of mainly three parts, (1) motion estimation, (2) motion correction and (3) motion compensation. The stabilization system aims to estimate and compensate the undesired jittery motions between consecutive video frames, which are induced either due to human body tremor or unsteady platform shakiness. Out of the three, the motion estimation is considered to be the most crucial part as its accuracy highly depends on the frame content and any misleading motion parameter value will affect the system's performance adversely. The motion estimation accuracy degrades under textural artifacts like poor illumination, intensity variation, internal noise and frame blurring: the factors which are more prominent in case of low-light and night shooting, and thus limit the use of existing motion estimation techniques for night shooting application. In field of video stabilization, various optical and mechanical techniques [1–4] based on motion-sensors have been proposed. Nowadays most of the compact hand-held cameras and smartphones are being designed with a distinguished feature of low light or night shooting stabilization. This stability feature is generally achieved by incorporating the in-built motion sensors [5,6] and their driving circuitry. Sensor based techniques gain attraction for their ability to get fit in any shooting environment, but face a constraint of high cost and large space requirement. For economically efficient designs, software based solutions called *digital stabilization*

☆ This paper has been recommended for acceptance by O. Au.
* Corresponding author.
E-mail addresses: deepika.shukla0914@gmail.com (D. Shukla), jharajib@gmail.com (R.K. Jha).

http://dx.doi.org/10.1016/j.jvcir.2016.03.002
1047-3203/

techniques are suggested as smart alternative to the sensor based techniques. Digital stabilization provides a low cost solution over sensors, but are generally application specific. Current research over digital techniques focuses on achieving the better motion accuracy under the challenging textural conditions like blurring, noise, intensity variation, moving objects, moving platform and poor illumination with small processing burden for its real-time applicability.

2. State-of-the-art and problem formulation

Night-video stabilization due to its inherent textural deficiency turns out to be a challenging motion estimation case in comparison to the low light or day-time indoor/outdoor shooting. Dark homogeneous regions, local intensity variations, noisy and blurry conditions make night shooting a hard problem in the field of digital stabilization. Some solutions for low-light/night-shooting have been suggested at the sensor-end [7] but the field of night video stabilization has not been specifically explored for digital techniques. The homogeneity effect of dark boundary regions in the night shooting videos results in a serious cause of motion error for most of the block matching techniques [8–11], while the optical flow [12] and feature based methods [13–15] fail due to the presence of noise, local variation and blurring effect. Under poor texture condition projection based methods like integral [16,17] and threshold-projection [18] work good but the motion accuracy degrades in presence of local intensity variations and in-scene moving objects. In literature, Radon Transform based on projection correlation [19] has also been suggested for combined translational and angular motion estimation, but the technique gives limited angular resolution and its motion accuracy degrades for large translational shift. Efficient projection matching under small perturbation is ensured using classical projection warping [20,21] and dynamic programming based dynamic time warping (DP-DTW) [22]. The CW [21] has a drawback of large processing time and the optimal path ambiguity for the case where multiple path having similar accumulated cost exist, and this condition is very likely in case of night videos due to the zero valued regions present in distance matrix. The dynamic programming based DTW [22] overcomes the path ambiguity issue and gives efficient motion estimation under degraded texture, but small variations in the projection-shape resulting from in-scene moving objects and intensity variation affect its warping accuracy. Use of derivative instead of intensity values for warping [23] by incorporating projection-shape as the matching feature, overcomes the effect of local variations within the projection [24]. Recently, a combination of angular projection and derivative warping vectors providing similarity stabilization [25] is reported for low light shooting environment. Night shooting differs from the low-light shooting in view of its limited textural information. The similarity technique [25] performs well in case of low light environment, but fails under night shooting due to large amount of mismatched warping vectors.

In video stabilization application, motion between the consecutive frames lies within a certain range, hence the full projection warping [21,22,24] leads to redundant processing. The constrained derivative projection warping under dedicated search space [26] eliminates the undesired matching over complete projection length and gives improved motion accuracy with significant processing time reduction. The constrained warping method estimates the relative frame motion efficiently for most of the day or low light cases, but the presence of dark frame-boundaries in night video produces zero-distance regions around the constraint-motion boundary and results in wrong motion due to optimal path tracing along lower or upper constraint-bound.

A failure case analysis of various motion estimation techniques is illustrated in Fig. 1 for a hand-recorded night traffic footage having small moving objects with local intensity variations Fig. 1(a) shows the reference, target frame and their corresponding horizontal projections sequentially, where the horizontal frame shift is clearly visible in their projections. Fig. 1(b) shows the optical flow (OF) [12] between the two frames, OF stabilized target frame, and comparative zoomed-frame areas highlighting the OF induced blobbing effect near moving lighted vehicles (marked as big black box in OF diagram and the green block in stabilized part of frame) and the shrinking effect on the lighted traffic-signal pole. Fig. 1(c) and (d) shows the Speeded-Up Robust Features (SURF) [14] and Maximally Stable Extremal Regions (MSER) [15] for the two frames. Both the feature based techniques suffer from feature mismatching due to in-scene lighted moving vehicles. Fig. 1(e) shows the DP based derivative dynamic time warping (DP-DDTW) matrix, the corresponding warped projections and the stabilized frame using similarity transformation over warping vectors [25]. Matching singularities present in the warped projections (highlighted by zooming over small regions shown under boxes) result in wrong transformation estimation. Frame stabilization using single-accumulated constrained DDTW (SA-CDDTW) [26] is shown in Fig. 1(f), where miswarped vectors lead to wrong motion estimation and result in poor stabilization.

To overcome the mismatching singularity effect on black-out areas resulting to zero distance regions in distance matrix, a dual accumulation over constrained matrix has been proposed. At the first level, the zero distance regions are changed to constant value region, and the dual accumulation performed at second stage produces a variance over the specified regions. Fig. 1(g) shows the proposed dual-accumulation distance matrix, the corresponding warped projections with reduced mismatching vectors and the stabilized target frame.

2.1. Key contribution

The paper highlights the need for improved motion estimation techniques under challenging conditions like low lighting, frame blurring and environmental and/or imaging sensor noise, since most of the real-world footages are affected by these conditions. Nighttime shooting because of the limited illumination sources becomes very susceptible to all the mentioned challenging conditions and hence considered as a hard problem for accurate motion estimation. This paper presents a novel concept of dual accumulated constrained projection warping as a robust solution to night video stabilization. Dark regions having very less or no textural information lead to constant intensity level in the frame projections and result in projection mismatching. Curve warping is chosen to provide motion accuracy under local perturbations, and robustness to mismatched projection vectors under constant intensity effect is achieved by applying dual-accumulation over distance matrix. Processing efficiency is ensured by incorporating a confined search region over projection matching. Smoothed derivative of the intensity-projection is used to overcome the intensity variation effects. The optimal warping path of the proposed dual-accumulated constrained derivative dynamic time warping (DA-CDDTW) matrix is utilized for the relative frame motion estimation. Optimal warping path estimated at this step is free from mismatching and boundary tracing problems, and gives accurate motion estimation between the low illumination frames. It is worthwhile to mention here that the proposed scheme efficiently covers both the night-shooting and normal illumination videos, unlike the existing techniques, which handle videos of normal illumination only.