Computer Vision and Image Understanding 155 (2017) 83–105

Contents lists available at ScienceDirect

Computer Vision and Image Understanding

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

A new composite multi-constrained differential-radon warping approach for digital video affine motion stabilization

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

[a] PPDM-Indian Institute of Information Technology, Design and Manufacturing Jabalpur, Madhya Pradesh, India
[b] Indian Institute of Technology Patna, Bihar, India

ARTICLE INFO

Article history:
Received 1 January 2016
Revised 21 November 2016
Accepted 30 November 2016
Available online 1 December 2016

Keywords:
Motion estimation
Video stabilization
Integral projection
Radon transform
Dynamic time warping

ABSTRACT

This paper presents a new projection based affine motion stabilizer framework for video stabilization using differential-Radon (DRadon) curve warping. Extending the translational domain of classical projection based algorithms towards affine stabilization, multiple angular curves obtained with Radon or rotated images have recently been explored for combined rotation and zoom estimation. Radon provides efficient projection extraction, but use of integral intensity under local variation degrades the desired motion accuracy. DRadon works on derivative of each angular slice to incorporate shape based matching for better projection alignment. Based on human perception of inter-frame tilt in unsteady camera recordings, the proposed angular DRadon curve estimation is confined to angular search space of [−20°, 20°] with the angular increment of 0.1°. Out of complete set of angular DRadon curves of reference frame, five key-angular slices are selected and correlated with their corresponding neighbourhood in target DRadon for inter-frame tilt estimation. Best matched key slices of reference and target DRadon are warped using a novel multi-constrained approach and the extracted warping vectors are further processed for translation and zoom estimation. A vector-slope algorithm based on relative stretching/contraction between the DRadon-projections is used for camera zoom estimation. Combining the estimated motion parameters, an affine transformation is developed for inter-frame stabilization. Comparative performance using motion accuracy and frame stability is evaluated over different categories of real-world videos.

1. Introduction

Unsteady hand-held or mounted camera recordings suffer from unwanted jerkiness due to physical or environmental vibrations. Presence of platform shakiness is very likely in surveillance, law-enforcement, and military applications. Non-professional users of smartphones and compact handycams are quite unfamiliar to the correct camera handling poses and thus induced body tremors add jittery motion effects in the captured sequence. Easiness and availability of innumerous imaging devices also led to huge amount of casually shot videos. This drastically increasing use of camera enforced imaging industry towards the cost effective high-end miniature camera design with efficient stabilization techniques. A motion-dedicated image/video stabilization system aims to estimate the undesired shakiness induced camera motions, while retaining user's intentional movements to provide a pleasant steady view of the unsteady camera recordings. The estimated jittery parts of the motions are compensated by applying reverse alignment of either imaging plane or the captured frame. Current cameras are being designed with in-built motion sensors (Karpenko et al., 2011; Jia and Evans, 2014) which provide online real-time video stabilization by removing the inherent high frequency vibrations. Sensor based stabilizations (Oshima et al., 1989; Sato et al., 1993; Zhao et al., 2009; Kinugasa et al., 1990) work on real-time processing but high user-end cost and large space requirements limit their usage in compact and low-cost stabilizer designs. Digital techniques (Uomori et al., 1990; Ratakonda, 1998; Matsushita et al., 2006; Hu et al., 2007; Pinto and Anurenjan, 2011; Okade and Biswas, 2014; Grundmann et al., 2012; Grundmann et al., 2011; Ko et al., 1998; 1999; Vella et al., 2002; Battiato et al., 2010; Brox et al., 2004; Liu et al., 2014; Liu et al., 2013; Kim and Park, 1992; Sauer and Schwartz, 1996; Piva et al., 2003; Bosco et al., 2008; Shukla and Jha, 2015; Shukla and Jha, 2013a, b; Veldandi et al., 2013; Mohamadabadi et al., 2012a, b; Puglisi and Battiato, 2011) (http://iplab.dmi.unict.it/download/Video/TCSVT2011) in form of their program-scripts provide economically efficient solutions for video stabilization, but these techniques mostly work offline to process the video after it has been recorded. Frame buffer for required processing over stored frames, is adapted for real-time stabilization.

* Corresponding author.
E-mail address: deepika.shukla@iiitdmj.ac.in (D. Shukla).

http://dx.doi.org/10.1016/j.cviu.2016.11.009
1077-3142/

Existing literature on digital stabilization can be broadly classified as *feature* based and *image* based techniques. Feature based stabilization techniques (Uomori et al., 1990; Ratakonda, 1998; Matsushita et al., 2006; Hu et al., 2007; Pinto and Anurenjan, 2011; Okade and Biswas, 2014) work on distinct image-feature extraction, which are matched to find correct transformation between the reference and target images. These techniques provide affine or spatial motion stabilization and but their reliability is highly dependent on the quality of image. Poor textural conditions like noise, blur, low intensity, and presence of homogeneous regions may lead to motion failure as a consequence of limited number of distinct or reliable features (Ratakonda, 1998). Adding robustness to degraded textural conditions some modified feature matching based solutions are reported. Instead of working over feature trajectories, Grundmann et al. (2012) presented mesh based analysis, where feature estimation and outlier rejection are performed over small grids and the estimated adaptive feature points provide improved results for texture less regions. The scheme is further modified using L1 optimization over estimated feature path (Grundmann et al., 2011) but these solutions may also fail under significant visual degradation caused by blur and missing features.

Image based techniques (Ko et al., 1998; 1999; Vella et al., 2002; Battiato et al., 2010; Brox et al., 2004; Liu et al., 2014; Liu et al., 2013; Kim and Park, 1992; Sauer and Schwartz, 1996; Piva et al., 2003; Bosco et al., 2008; Shukla and Jha, 2015; Shukla and Jha, 2013a, b; Veldandi et al., 2013; Mohamadabadi et al., 2012a, b; Puglisi and Battiato, 2011) (http://iplab.dmi.unict.it/download/Video/TCSVT2011 bib35) do not rely on salient regions/points, rather consider whole image either in blocks or in full-area for relative motion estimation. Spatial processing in the image techniques proves to be better alternative under challenging scene capture conditions, as it is less probable to motion failure. Various block based and projection based techniques of this category have their application domain confined to translation motion only. Block matching techniques (Ko et al., 1998; 1999; Vella et al., 2002) use sliding window concept for finding the best match between test block of reference frame and corresponding target block under dedicated search space. Block shift required in finding the best match is considered as local motion of the respective test block. Accuracy of block matching techniques degrades with local intensity variations and homogenous regions. Modifying this, Battiato et al. (2010) have addressed the performance of block matching using pre-filtering criterion to eliminate erroneous local block motion vectors for challenging scene capture conditions like illumination change, blur and moving objects, but the approach has limitation over periodic patterns and homogeneous region. Instead of block correlation, optical flow technique (Brox et al., 2004) using relative pixel velocity is explored for relative image affine motion estimation. The technique provide good stabilization by neighbour-pixel warping but sharp local intensity variations in the captured scene sometimes lead to inaccurate flow estimation and results distorted view of the warped target frame. Overcoming the hidden limitations of optical flow, Liu et al. (2014) have suggested the mesh based warping using smoothed pixel profiles of estimated optical flow vectors, which provides robustness over textureless regions and moving object boundaries. Use of multiple camera paths with constrained pixels profile of optical flow, provides added robustness to occlusion, blur and rapid camera motion (Liu et al., 2013).

Integral frame projections obtained by averaged cumulative pixel intensities along a fixed angular direction provide a simple and efficient alternative to computationally poor and local mismatch prone block-matching algorithms. Projections based techniques (Kim and Park, 1992; Sauer and Schwartz, 1996; Piva et al., 2003; Bosco et al., 2008; Shukla and Jha, 2015; Shukla and Jha, 2013a, b; Veldandi et al., 2013; Mohamadabadi et al., 2012a, b; Puglisi and Battiato, 2011) (http://iplab.dmi.unict.it/download/Video/TCSVT2011 bib35, Shukla et al., 2016) instead of working on local areas, use a global distribution of frame content. These techniques are thus less affected by local scene variations and provide better performance under above mentioned challenging motion estimation cases. Instead of working on entire frame as whole, block-projections have also been proposed as a time efficient alternative to block-matching techniques (Kim and Park, 1992; Sauer and Schwartz, 1996). In literature, motion estimation based on projection are explored with correlation (Piva et al., 2003) and warping (Bosco et al., 2008; Shukla and Jha, 2015; Shukla and Jha, 2013a, b) techniques. Extension of translational domain of projection techniques towards rotation estimation has been reported recently using multiple angular projections obtained with rotated images (Veldandi et al., 2013) or Radon based angular slices (Mohamadabadi et al., 2012a, b). Multiple angular curves are matched with a fixed reference curve and the angular position corresponding to the best matched projection is termed as the inherent rotational tilt. Rotating an image at different angles for possible projection extraction overburdens the processing time, while Radon transform based on Fourier implementation provides fast processing and thus chosen as better option for angular curve extraction. Extending the projection based scheme towards affine motion stabilization, Puglisi and Battiato (2011) combined the block based approach with derivative block projection error and least square estimation for affine transformation. Different categories of challenging scene capture conditions have been addressed and stabilization results of the approach with state-of-the-art are publicly available (http://iplab.dmi.unict.it/download/Video/TCSVT2011 bib35).

In this paper, a differential-Radon (DRadon) made up of the derivative curves for each angular slice has been proposed as efficient alternative to classical intensity integral based Radon projections (Mohamadabadi et al., 2012a). The derivative in form of shape feature in the test and target DRadon provides improved correlation values and ensure better rotational accuracy. DRadon gives dominating correlation peak at the desired image-tilt angle, but use of single correlation curve may mislead the true angular motion due to peak allocation error (Shukla et al., 2016). Overcoming this, multiple key-slices are chosen instead of using single correlation curve to provide improved motion angular motion accuracy. In unsteady platform recordings, the rotation induced due to inherent shakiness is observed within a small range, hence instead of estimating angular curves over half-circle region as in (Veldandi et al., 2013; Mohamadabadi et al., 2012a, b), the curve extraction in proposed DRadon is confined under the dedicated search range of $[-20^{o}, 20^{o}]$. Intermediate resolution between angular DRadon slices is set to 0.1^{o} for improved rotational accuracy. Relative frame translation is reflected as the corresponding projection shift, while the camera zooming can be efficiently observed as the relative stretching/compression of the projection curves. Translational shift and stretching/contraction between best matched horizontal and vertical DRadon slices of the reference and target frames are used to find relative translation and zoom parameters respectively. Veldandi et al. (2013) have suggested the least square solution of estimated projection warping vectors for combined translational and zoom motion estimation, but the presence of miswarped singularity vectors as one-to-many match between curve elements affects motion accuracy severely and limits the intended range of both the translation and zoom motion extraction.

In the proposed affine video stabilization technique, best matched angular slices of the test and the target DRadon are warped using novel multi-constrained dynamic time warping (MC-DTW) approach. A combination of Sakoe-Chiba search band (Shukla and Jha, 2013b, Keogh and Pazzani, 2001) and the new boundary constraint in proposed multi-constrained region is used to suppress the misaligned boundary vectors corresponding to