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.

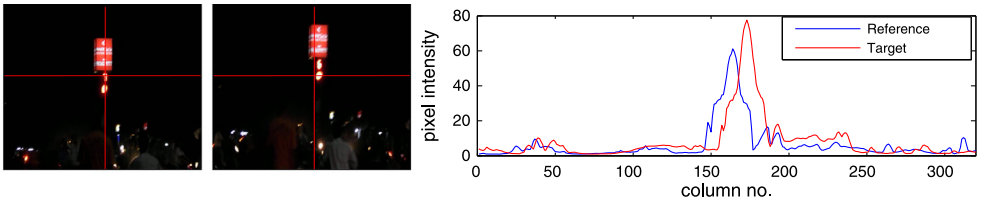
In video stabilization application, motion between the consecutive frames lies within a certain range, hence the full projection warping [21,22,24] leads to redundant processing. The constrained derivative projection warping under dedicated search space [26] eliminates the undesired matching over complete 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.

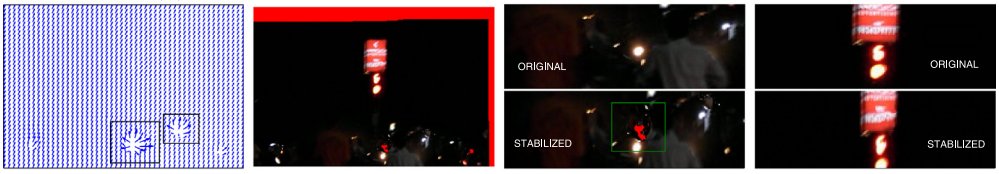
To overcome the mismatching singularity effect on black-out areas resulting to zero distance regions in distance matrix, a dual accumulation over constrained matrix has been proposed. At the first level, the zero distance regions are changed to constant value region, and the dual accumulation performed 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.



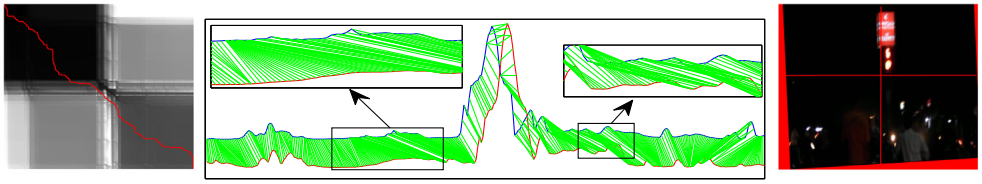

(a) Reference; Target; Corresponding horizontal projections



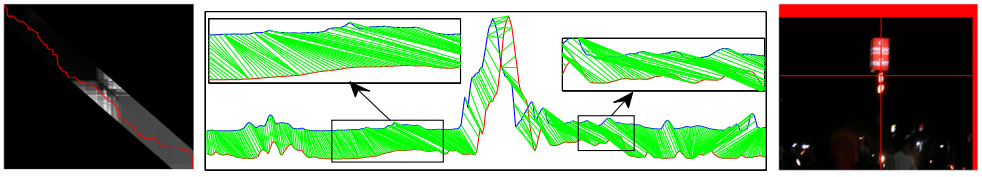

(b) Relative OF; OF stabilized target; blobbing effect; shrinking effect

(c) SURF on reference; SURF on target; Matching; SURF stabilized target

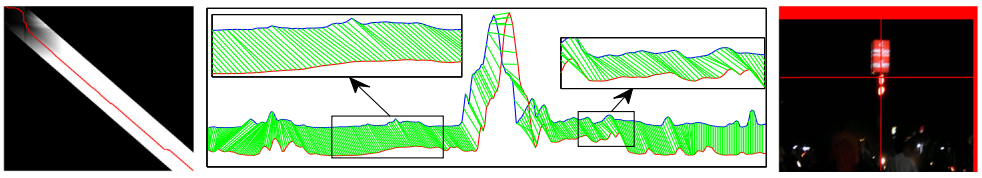
(d) MSER on reference; MSER on target; Matching; MSER stabilized target



(e) DP-DDTW matrix; projection warping with local singularity; Similarity stabilized target



(f) SA-CDDTW matrix; projection warping under constrained search space; stabilized target



(g) Proposed DA-CDDTW matrix; correctly warped projections; stabilized target

**Fig. 1.** Stabilization performance evaluation using different techniques over night-video.

In the subsequent part of the paper, Section 3 introduces the proposed Dual-accumulated constrained warping (DA-CDDTW) with a brief description of the earlier constrained projection warping technique. The proposed motion estimation and the stabilization framework are discussed in Sections 4 and 5 respectively. Section 6 presents a comparative experimental result analysis for different categories of night-shooting videos. Finally, the research outcomes of the proposed work are concluded in Section 7.

## 3. Proposed warping technique

Any translational motion between two consecutive frames can be efficiently represented in terms of the translational shift between their corresponding integral projections. The frame signatures '*S*' i.e. horizontal and vertical integral projections for a frame '*I*' of size $R \times C$ are given using (1) and (2) respectively,

$$S_H(j) = \frac{1}{R}\sum_{i=1}^{R} I(i,j); \quad \forall j \in [1:C] \tag{1}$$

$$S_V(i) = \frac{1}{C}\sum_{j=1}^{C} I(i,j); \quad \forall i \in [1:R] \tag{2}$$

The relative frame displacement in the horizontal and vertical directions can be obtained by using their corresponding signatures. In literature, the projection warping has been suggested as an efficient motion estimation alternative under poor textural conditions [17,19,22] but the technique suffers from large computational time and memory requirement. Intensity variation with in the frame results in projection mismatching and degrades motion accuracy. In the proposed method, robustness to the local intensity variation is achieved by utilizing projection-shape in terms of its derivative as the matching feature. The derivative signatures in the two directions (i.e. $DS_H$ and $DS_V$) for each frame are obtained using left point estimation given by (3).

$$\begin{aligned} DS_H(j) &= S_H(i) - S_H(i-1); \quad 1 \leqslant j \leqslant C, \quad j = i-1 \\ DS_V(j) &= S_V(i) - S_V(i-1); \quad 1 \leqslant j \leqslant R, \quad j = i-1 \end{aligned} \tag{3}$$

Relative frame motion, depending upon the applications varying from hand-recording to moving platform shooting, is observed to lie within a permissible range of 5–30 pixels. To eliminate the redundant searching over entire frame projection, Derivative Dynamic Time Warping (DDTW) [23] over constrained search space has been suggested [26]. This overcomes the extra burden of complete distance matrix estimation and gives sufficient processing time reduction. In the constrained warping approach motion accuracy degrades when the optimal warping path traces the constraint boundary, which is generally the case with most of the nighttime videos having large amount of dark regions present within frame. In this section, a brief background over existing constrained warping technique is presented for proper understanding of inherent motion estimation challenges in night videos and then the proposed modification using dual-accumulated warping over constrained region is discussed.

### 3.1. Background: constrained DDTW

The jitter induced frame displacement generally lies within a certain range depending upon the application, thus instead of finding match over entire projection length, the search can be confined within a specified range. To provide equal search space in both the directions, Sakoe–Chiba window (SC-window) [23] of size '*w*' has been chosen as a constrained band in both the left and right sides of the matrix diagonal. This actually corresponds to a total search window of size $(2w+1)$ imposed around each projection element, where the windowing parameter $w$ is chosen to be 35. Within the SC-window the cost matrix $D^f_{f-1}$ between the derivative horizontal signatures $DS_H^{f-1}$ and $DS_H^f$ of any two consecutive frames is given by (4).

$$D^f_{f-1}(i,j) = \left| DS_H^{f-1}(i) - DS_H^f(j) \right| \tag{4}$$

where $\forall i \in [1:C-1]$ and $i - w \leqslant j \leqslant i + w$. For the sake of convenience, we shall be using $D(i,j)$ in place of $D^f_{f-1}(i,j)$, whenever there is no ambiguity in representation. Using the dynamic programming (DP) implementation [22], the elements of single accumulation distance matrix (*SAccD*) inside the constrained window are obtained as (5). At the constraint boundary, the lower and upper bounds are incorporated by choosing the minimum of two neighboring elements lying within the confined region.

$$\begin{aligned} SAccD(i,1) &= \sum_{k=1}^{i} D(k,1); \quad \forall i \in [1,w] \\ SAccD(1,j) &= \sum_{k=1}^{j} D(1,k); \quad \forall j \in [1,w] \\ SAccD(i,j) &= min[SAccD(i-1,j-1), SAccD(i-1,j), SAccD(i,j-1)] \\ &\quad + D(i,j); \quad \forall i \in [2:C-1] \text{ and } i - w \leqslant j \leqslant i + w. \end{aligned} \tag{5}$$

In the night shooting videos, dark regions present along the boundary and upper part of the frames, lead to zero distance portions in the cost matrix '*D*'. Single accumulated-constrained DDTW (SA-CDDTW) [26] over constrained matrix '*D*' converts the zero distance region into a constant distance region, but the homogeneous effect in the matrix remains unchanged. When such regions fall on the constrained boundary, the optimal path tracks the boundary which further leads to wrong motion estimation. To overcome this problem an efficient solution of applying repeated accumulation over matrix '*SAccD*' has been proposed.

### 3.2. Proposed algorithm: dual accumulated CDDTW

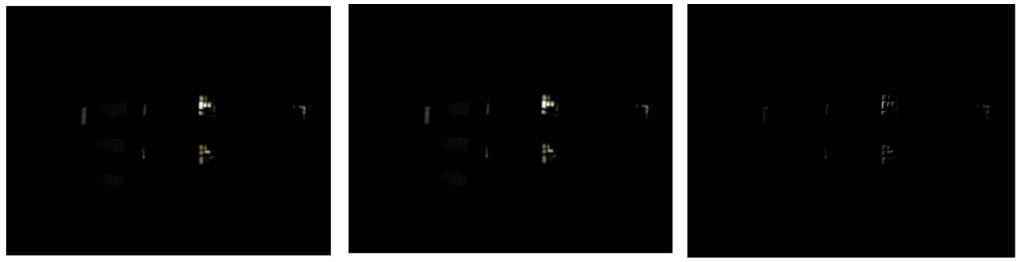
The zero valued regions in distance matrix '*D*' corresponding to black-out frame portions, are converted to constant value region in *SAccD*. These constant valued regions in *SAccD* cause one-way tracking of the optimal path in search of minimum cost elements. By applying the recursive DP algorithm (i.e. dual accumulation), at each move in raster-scan, distance value of matrix elements falling within the constant value region of *SAccD* matrix is increased by the amount of the minimum distance value of its three previous neighbors. This converts the constant distance regions of *SAccD* into a varying distance regions, while preserving the inherent distribution within the distance matrix *D*. The matrix *DAccD* obtained using the dual accumulation CDDTW (DA-CDDTW) is given as (6).

$$\begin{aligned} DAccD(i,1) &= \sum_{k=1}^{i} SAccD(k,1); \quad \forall i \in [1,w] \\ DAccD(1,j) &= \sum_{k=1}^{j} SAccD(1,k); \quad \forall j \in [1,w] \\ DAccD(i,j) &= min[DAccD(i-1,j-1), DAccD(i-1,j), DAccD(i,j-1)] \\ &\quad + SAccD(i,j); \quad \forall i \in [2:C-1] \text{ and } i - w \leqslant j \leqslant i + w. \end{aligned} \tag{6}$$

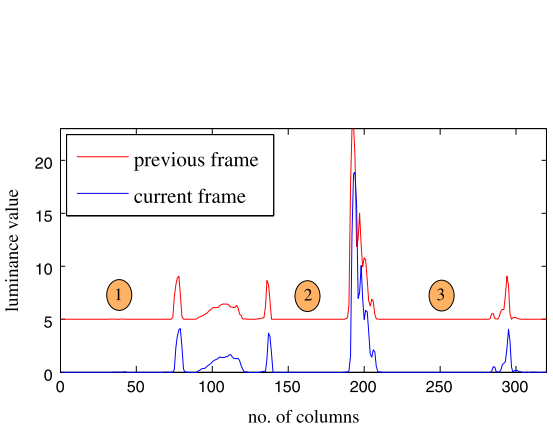
The optimal warping path $P$ between the two projections extracted using backward tracking of minimum valued neighbors within the dual-accumulated SC-window is given by (7).

$$P := \begin{cases} (1,j-1); & \text{if } i = 1, \\ (i-1,1); & \text{if } j = 1, \\ argmin\{DAccD(i-1,j-1), DAccD(i-1,j), \\ DAccD(i,j-1)\}; & \text{otherwise}. \end{cases} \tag{7}$$

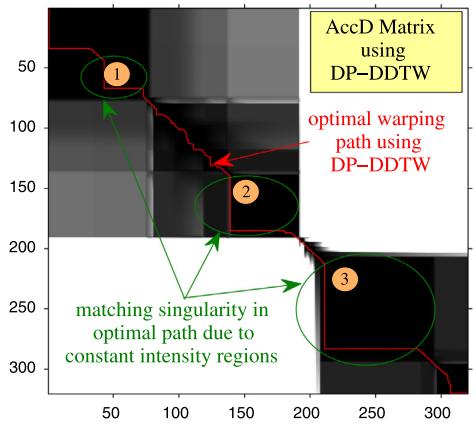
where $P \in \mathbb{Z}^{L\times 2}$. Optimal path $P$ contains the index-pair list of the selected minimum valued elements obtained in (6) and can be represented as a $L \times 2$ matrix, where $L$ is the length of extracted warping path and the two columns contains the row and column indices of the selected matrix elements.



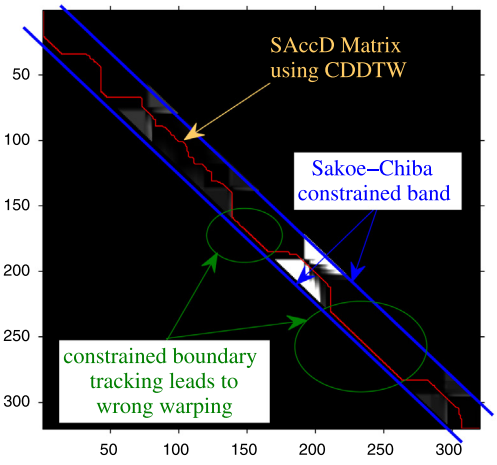
(a) frame 9, frame 10 and the difference frame



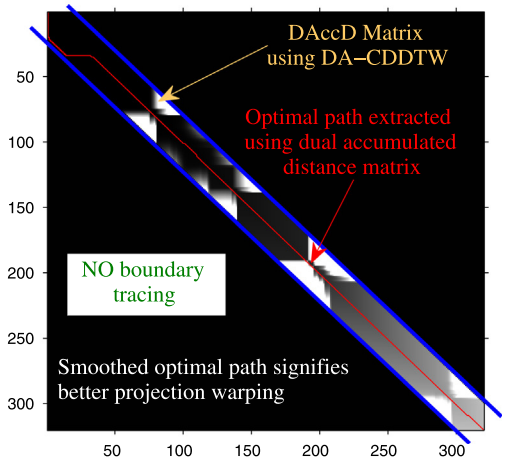

(b) Horizontal projections




(c) DP-DDTW




(d) SA-CDDTW




(e) DA-CDDTW

**Fig. 2.** Motion estimation analysis under dark in-frame regions: (a) frame 9; frame 10; difference frame of a test night-video, (b) corresponding horizontal projections; Optimal warping path estimation using (c) DP-DDTW (d) SA-CDDTW (e) Proposed DA-CDDTW technique.

Fig. 2 shows the optimal path extracted using DP-DDTW [19,24], SA-CDDTW [26] and the proposed DA-CDDTW techniques. The repeated accumulation does not affect the local distribution of the distance matrix (The distribution of the gray shades remains at the same place but with increased intensity). Two consecutive frames and the respective difference frame for a static night test video are shown in Fig. 2(a). Fig. 2(b) shows the corresponding horizontal projections (one curve is intensity lifted for better visualization), which contains large zero intensity portions and a relative shift of 1 pixel. The DDTW extracted optimal path (red) corresponding to these zero intensity regions, as shown in Fig. 2(c), results to the encircled horizontal and vertical segments within path. These encircled segments of the path correspond to the projection mismatching singularities as one-to-many match within the projection's constant intensity sections. Fig. 2(d) shows that these portions cause boundary tracking (encircled) for the SA-CDDTW [26]. Optimal warping path using the proposed DA-CDDTW technique is shown in Fig. 2(e), which is free from the boundary tracking and results in better curve warping even in poor textural information.

## 4. Proposed motion estimation approach

The optimal warping path $P$ obtained by the derivative signature warping is used to get the relative frame displacement. $P$ consists of a sequence of different matrix elements having their

corresponding row and column values as its index pairs. In each element of path $P$ represented as $p_{i,j}$, the indices '$i$' and '$j$' refer to the individual warping vector drawn between the corresponding reference projection element of frame $f-1$ and its matched point in the target projection of frame $f$ respectively. Local motion vectors (LMVs) between the two frames are obtained by finding the difference of two index values of each element of path $P$. Out of these *LMV*s the most frequent value is considered as the global motion vector (GMV) between the two frames. A method for finding *GMV*s in the two directions $dr \in$ [H,V] i.e horizontal and vertical, for different consecutive pairs of frames is given by (8).

$$\begin{aligned} P^{dr} &= DA\text{-}CDDTW\left(S_{dr}^{f-1}, S_{dr}^{f}\right) \\ LMV_{dr}(i) &= p_{i,1}^{dr} - p_{i,2}^{dr}; \quad \forall i \in [1:L] \\ GMV_{dr} &= mode(LMV_{dr}(g+1:L-g)) \end{aligned} \tag{8}$$

where $L$ is the length of optimal warping path. In unsteady videos the frame boundary variations resulting from the newly added content, make end parts of the frame projections to be the most mismatching prone areas. To avoid this, *LMV*s for a length of '$g$' at both the ends of the matched projections are discarded and the most frequent motion vector within the constraint range given by $LMV(g+1:L-g)$ is chosen as the resultant *GMV*. The parameter '$g$' is termed as projection-bounding constraint and can be chosen independently to the motion search-constraint parameter $w$. In this paper, both the projection-bounding and search constrained window are chosen to be 35, but this can be varied depending upon the applications like hand-held shooting or moving platform shooting. For hand-held recordings because of confined range of jittery motions small search-window is preferred.

Fig. 3 shows the resulting local motion vectors (green[1] lines) of the two horizontal projections (Fig. 2(b)) using DP-DDTW [24], SA-CDDTW [26] and proposed DA-CDDTW techniques. At the constant intensity regions the DP-DDTW and SA-CDDTW contain large mismatching singularities. The DP-DDTW gives accurate motion estimation, but warping over complete signature adds unnecessary processing burden. The SA-CDDTW gives a time efficient solution but the presence of boundary-traced section in warping path results in similar *LMV* values and choosing the most frequent *LMV* of this leads to wrong motion estimation. The proposed method inherits the constrained motion search space for reduced processing time and dual accumulation eliminates the boundary tracing problem, which results in accurate projection warping and gives the correct intended motion estimation (i.e. the imposed displacement of 1 pixel between test frames).

## 5. Proposed stabilization system framework

A framework of the proposed translational motion stabilization system dedicated to low quality night videos is shown in Fig. 4. At the initial stage, the derivative frame projections of the first frame are extracted using (1) and (2) and are stored in their corresponding buffers. The next frame is then processed and its projections are warped with the stored ones using the proposed DA-CDDTW. At the motion estimation step, the *LMV*s and the *GMV*s are estimated in the two directions using (8). The projection warping results in a set of local motion vector *LMV*s out of which the most frequent vector is extracted as Global motion vector and gives a representative motion between the two frames. To get the motion with respect to a reference frame, an accumulated *GMV* (AGMV) in the two directions $dr \in$ H,V i.e. horizontal and vertical, is calculated using (9).

$$AGMV_{dr}^{f} = AGMV_{dr}^{f-1} + GMV_{dr}^{f} \tag{9}$$

*AGMV* is a $1 \times F$ sized matrix, where $F$ is the total number of frames. The $AGMV(1)$ is always set to zero representing no relative motion of the first frame with respect to itself. The estimated motions contain some intentional movements along with the unwanted jitter. Jitter being a high frequency oscillatory motion causes the *AGMV* to vary around a certain value, while the intentional motion causes additive increase in a single direction.

A threshold over the accumulated *GMV* is used to separate out the panning motion from the undesired jitter [11,22]. Panning makes *AGMV* to increase in a single direction exceeding the range of jitter induced displacement and results a large change in the corresponding accumulated motion. Here *AGMV* threshold given by (10), is used for motion correction.

$$AGMV_{dr}^{f} \begin{cases} \geqslant \text{threshold} \Rightarrow \text{Panning}, \\ < \text{threshold} \Rightarrow \text{Jitter}, \end{cases} \tag{10}$$

When the *AGMV* crosses the threshold in either horizontal or vertical direction (here the threshold is taken as one-fourth of the corresponding frame dimensions), the intentional panning is reported and the *AGMV* of the current frame is fixed to its previous value, while the reference frame is updated from the first frame to the current frame and this reference-frame-updation continues until the *AGMV* again goes below the threshold level. For any inherent jitter combined with the panning motion, the shakiness effect can be efficiently corrected because of its oscillatory deviation from the previous *AGMV* level which was set earlier for the panning motion. The motion compensation block utilizes the negative of these corrected *AGMV*s for repositioning the corresponding frames at their correct locations. The resultant boundary gaps due to motion compensation can be filled either using previous stabilized frame as the background or by existing frame-inpainting technique [27].

## 6. Experimental results and discussion

As stated in the Section 2, the night videos captured using compact cameras suffer from various constraints like poor lighting conditions, noise, blur, local intensity variation and put a limit over estimated motion accuracy. Being a critical case, the field has not yet been specifically analyzed for efficient motion estimation solution using digital algorithms. The existing block-matching [8–11] and feature based methods [13–15] fail to stabilize the night videos because of the dark or low intensity regions (homogeneous effect) and the blurring effects respectively. In case of poor quality night videos, the feature or interest point based techniques also fail due to lack of reliable feature extraction. The integral projection [17] technique gives good results for static-scene condition and gets affected under intensity variation and presence of moving objects. Motion estimation accuracy of Radon Transform based technique [19] degrades under large amount of combined translational and rotational motions. The DP-DTW [22] and DP-DDTW [24] produce better stabilization under these cases but require large processing time. A similarity stabilization [25] for low light videos has also been proposed using DDTW. The technique performs well in case of low light condition, but the motion estimation for night videos is affected badly due to large number of mismatched warping vectors resulting from dark in-frame regions (see Fig. 3). The SA-CDDTW [26] approach reduces the warping time but at the cost of degraded motion accuracy. The proposed dual-accumulated derivative warping under constrained search space provides accurate and time efficient motion estimation solution and in comparison to existing techniques gives better night video stabilization. In this section, considering the failure of existing feature and pixel based approaches, a comparative stabilization

[1] For interpretation of color in Figs. 3, 6–9, and 11, the reader is referred to the web version of this article.

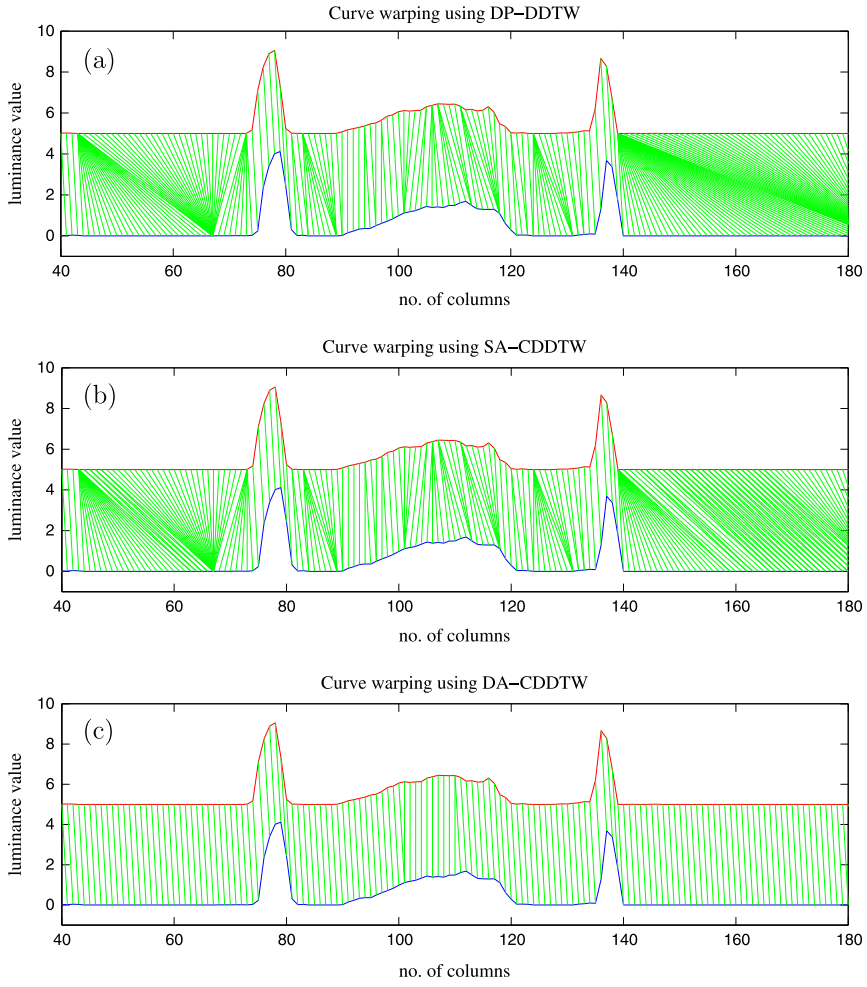



**Fig. 3.** Comparative projection warping analysis using (a) DP-DDTW, (b) SA-CDDTW and (c) Proposed DA-CDDTW technique for night shooting video.

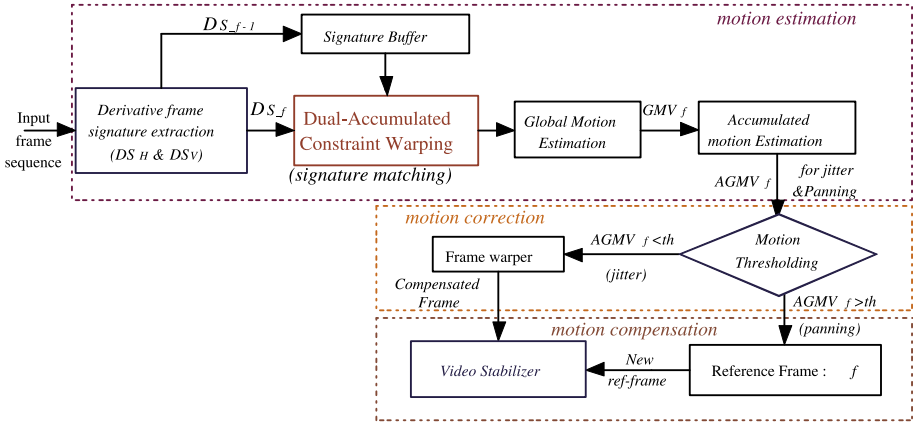



**Fig. 4.** Proposed video stabilization system framework.

result analysis with different projection based stabilization techniques is presented for various categories of night videos.

### 6.1. Motion estimation error analysis

The motion accuracy is evaluated using accumulated global motion (*AGMV*) error analysis. A night video captured from steady platform is artificially destabilized with known motion vectors. Motions extracted using different projection based stabilization algorithms are compared with the imposed ones. Fig. 5 shows a comparative plot between the imposed and extracted accumulated motions. Integral projection correlation [17] and projection warping [22] based techniques give satisfactory results, but due to error accumulation effect motion accuracy degrades at the larger number of frames and the corresponding *AGMV* curves deviate from the imposed motion curve. Motion accuracy using the least square

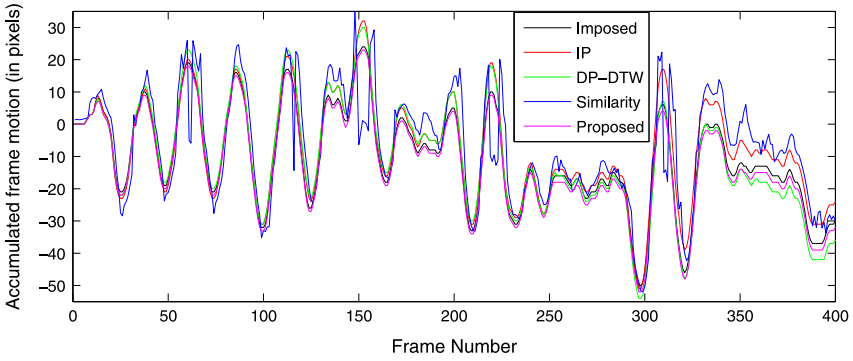



**Fig. 5.** Accumulated-motion error analysis over artificially destabilized video.

estimation [24] over warping vectors suffer from large motion errors due to existing mismatching singularities. The proposed DA-CDDTW warping technique in comparison to other methods gives better accuracy and the corresponding extracted accumulated motion curve closely follows the imposed motion curve over complete sequence of frames.

### 6.2. Subjective performance analysis

Subjective performance in terms of visual-content stabilization is evaluated over different categories of night shooting videos. Each considered category has its own sets of challenges and limitations to accurate motion estimation. The first four cases of stabilization consider the videos *DarkStatic*, *SmallMovObject*, *LargeMovObj*, and *LocalIntVar* captured using hand-held camera at steady platform, while the videos *OnRoadMov* and *GlobalIntvar* are captured from an on-road forward moving bus.

#### 6.2.1. Static scene

This category of night video can be considered as the most simple case for motion estimation. Static videos do not contain any sharp intensity variation as most of the lighted objects are retained at their positions. Jerky motions may induce a blurring effect due to small camera sensor limitations. Fig. 6(a) shows four different frames of hand-recorded video *DarkStatic*, having motion induced frame blurring. Low illumination causes large dark portions at the frame boundary. Fig. 6(b) shows the similarity stabilization results for each respective frame, where the least square estimation over miswarping vectors leads to failure of similarity motion estimation [25]. The corresponding stabilized frames using proposed method are shown in Fig. 6(c), where the lighted tree (highlighted in blue box) is positioned correctly irrespective of the blurring and dark frame boundary.

#### 6.2.2. Small moving objects

Presence of small moving objects results local shape variations in the frame projections. Fig. 7(a) shows the different frames of night-traffic video *SmallMovObject*, captured using hand-held camera. Lower part of the frame contains moving objects (marked within green box) and the upper portion contains zero-intensity regions. The corresponding similarity stabilized results are shown in Fig. 7(b), which suffer from false zooming due to miswarping singularity caused by illumination variation at lower part of frame. Fig. 7(c) shows stabilized frames using proposed DA-CDDTW method, in which the blue box covering the lighted signal-pole are aligned w.r.t. red lines.

#### 6.2.3. Large moving objects

Another night-traffic video *LargeMovObj* containing a moving car with jerky hand movements has been chosen for this category. The moving car (within green box) covers a large portion of the lighted section and the corresponding area changes as the car goes away. Some blurring is also present due to hand motion. Fig. 8(a) shows four different frames of *LargeMovObj* video, in which the background is quite visible and remains unchanged in comparison to other night videos. The moving car and motion blur in the video lead to local variations and produces error in case of similarity stabilization [25]. Fig. 8(b) shows the similarity stabilized frames, which are affected by false zooming due to miswarped vectors. Stabilization results of proposed DA-CDDTW method, are shown in Fig. 8(c), where the lighted background marked with blue line and rectangle, are aligned efficiently irrespective of the moving car.

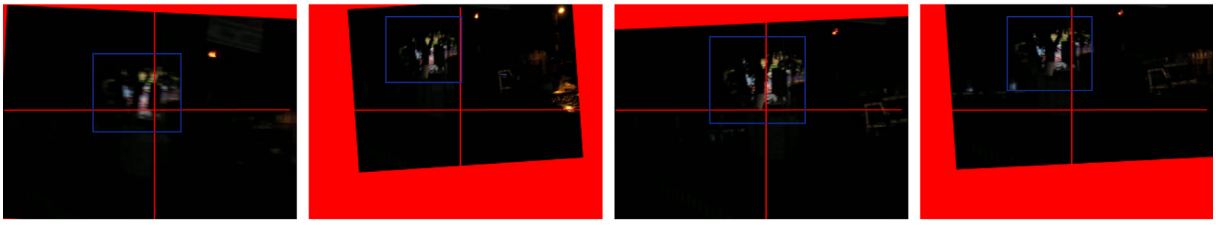
#### 6.2.4. In-scene local intensity variation

Local intensity variation is illustrated with a scene captured at a 4-way traffic signal. In this case all the moving vehicles (marked with green boxes) contribute to the intensity variation at their local positions, which affect both the shape and illumination level of the projection posing a challenge over correct motion estimation. Fig. 9(a) shows four frames of the traffic video *LocalIntVar*, captured using hand-held camera. Lower and upper part of the frame contains moving objects and zero intensity regions respectively with small frame blurring due to hand movements. Fig. 9 (b) shows the corresponding stabilized frames using Similarity method and suffers from wrong motion estimation. The proposed method is able to stabilize the sequence efficiently under mentioned constraints. Fig. 9(c) shows the stabilized frames, in which the signal-pole (marked with blue box), is aligned correctly w.r.t. red lines.

#### 6.2.5. On-road moving platform

The on-road videos captured at nighttime suffer from boundary variation along with the intensity variation resulting from in-scene road-lights and moving vehicles. Fig. 10(a) shows four different frames of *OnRoadMov* video, captured from a moving bus. Rapid platform jerks cause frame blurring and lighted poles at the road-divider (highlighted with slant line and centered blue box) produce intensity variation with small zooming effect. Left half of the frame suffers from low illumination. Fig. 10(b) shows the corresponding stabilized frames using similarity method. The method gives wrong rotation due to in-scene variation caused by forward platform motion. The proposed stabilization results are shown in Fig. 10(c), where the centered box over far-away lights and the slant lines drawn on right sided lighting billboards are positioned correctly in each corresponding frame.

(a) Original frames



(b) Similarity stabilized frames

(c) Proposed stabilized frames

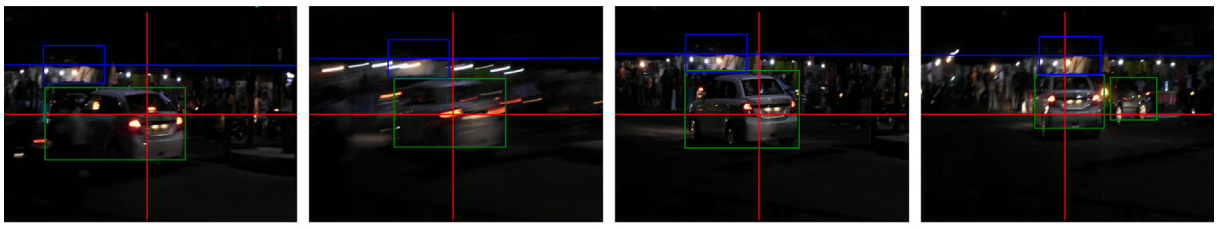
**Fig. 6.** Original and stabilized frames for video *DarkStatic*.

(a) Original frames

(b) Similarity stabilized frames

(c) Proposed stabilized frames

**Fig. 7.** Original and stabilized frame for video *SmallMovObj*.



(a) Original frames

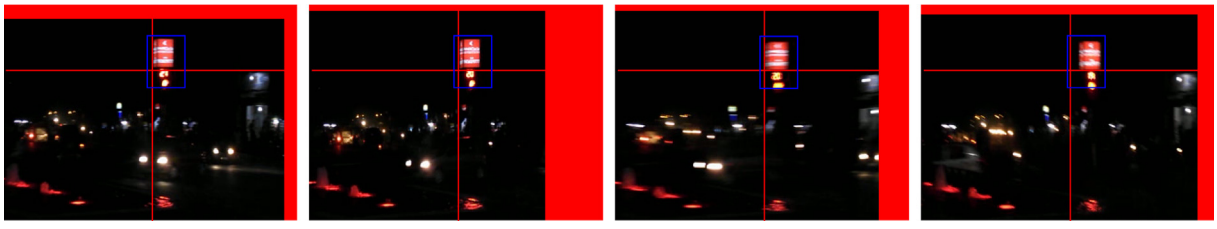
(b) Similarity stabilized frames

(c) Proposed stabilized frames

**Fig. 8.** Original and stabilized frame for video *LargeMovObj*.

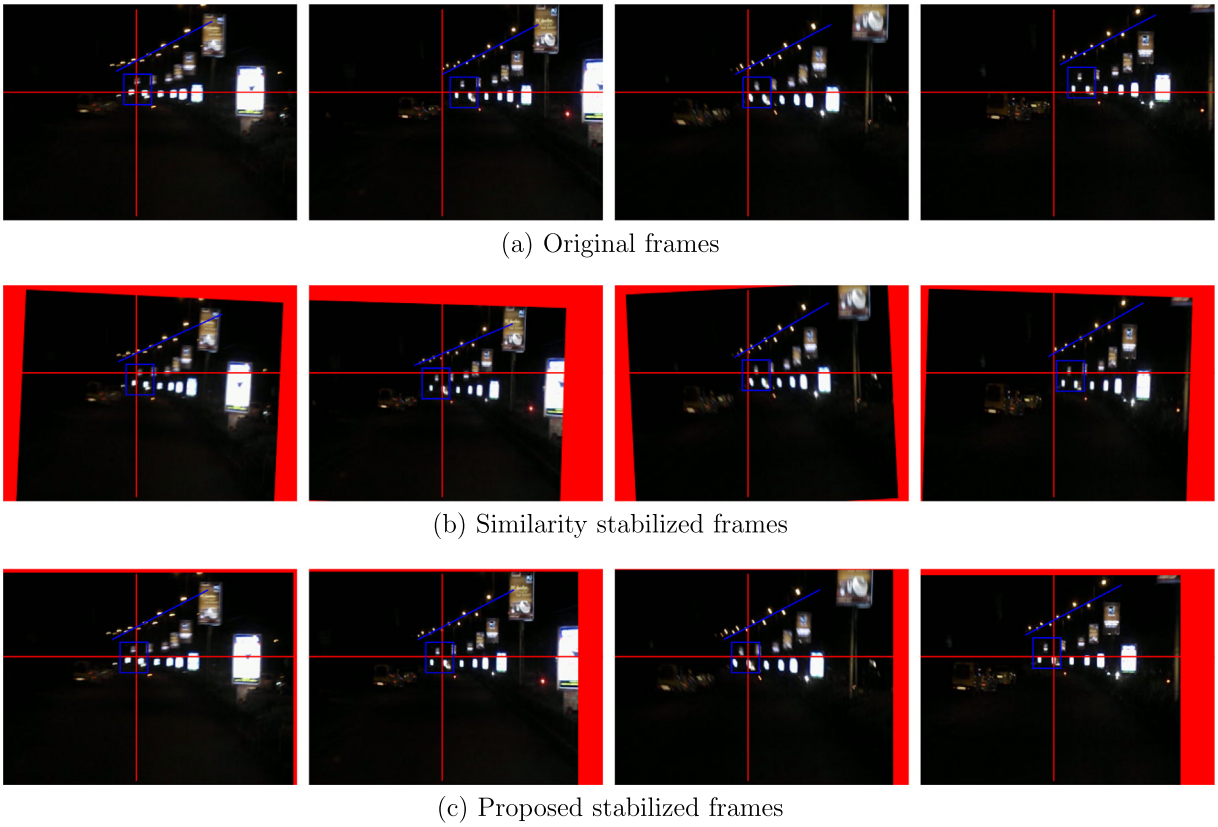
(a) Original frames

(b) Similarity stabilized frames



(c) Proposed stabilized frames

**Fig. 9.** Original and stabilized frame for video *LocalIntVar*.



(a) Original frames

(b) Similarity stabilized frames

(c) Proposed stabilized frames

**Fig. 10.** Original and stabilized frame for video *OnRoadMov*.

*6.2.6. In-scene global intensity variation*

The on-road night videos suffer from limited illumination sources and incur sudden intensity lift in case some bright light come on the way (e.g. road-lamps located at sufficient distance marked with yellow boxes and front-lighted moving vehicles marked in green box contribute to intensity variation). The particular case has been illustrated using the *GlobalIntVar* video having intensity variation due to incoming road-lights and the platform jerk induced frame burring. Fig. 11(a) shows four different frames of the video. The corresponding similarity stabilized results are shown in Fig. 11(b). These results suffer from wrong motion estimation due to limited content detailing available within the scene. In Fig. 11(c) the corresponding proposed stabilized frames with proper street-pole alignment (marked with blue box) are shown to demonstrate the stabilization performance.

*6.3. Quantitative performance analysis*

A quantitative performance evaluation of different stabilization algorithms is performed using the peak signal-to-noise ratio (PSNR) and the structural similarity (SSIM) index [28] chosen as the interframe alignment error measures. *PSNR* between the frames $I_f$ and $I_{f-1}$ is given as (11).

$$PSNR(I_f, I_{f-1}) = 10log_{10}\frac{I_{max}^2}{MSE(I_f, I_{f-1})} \tag{11}$$

where $MSE(I_f, I_{f-1})$, i.e. the mean square error between the frames is obtained using (12), and $I_{max}$ is the maximum intensity value of frame of size $R \times C$.

$$MSE(I_f, I_{f-1}) = \frac{1}{RC}\sum_{r=1}^{R}\sum_{c=1}^{C}\left[I_f(r,c) - I_{f-1}(r,c)\right]^2 \tag{12}$$

The *PSNR* and the *MSE*, being based on direct intensity differences, lack the human visual consideration for better perceived visual quality. The *SSIM* considers perceptual degradation as structural distortion. *SSIM* attempts to measure the perceptual changes by estimating the deviation in luminance, contrast, and structure within the image and is given as a product of the three distortions, where the luminance is modeled as average pixel intensity, contrast by the variance between the reference and target image, and structure by the cross-correlation between the two images. For two local patches $x$ and $y$ of common size of $8 \times 8$, the local *SSIM* index is defined as (13).

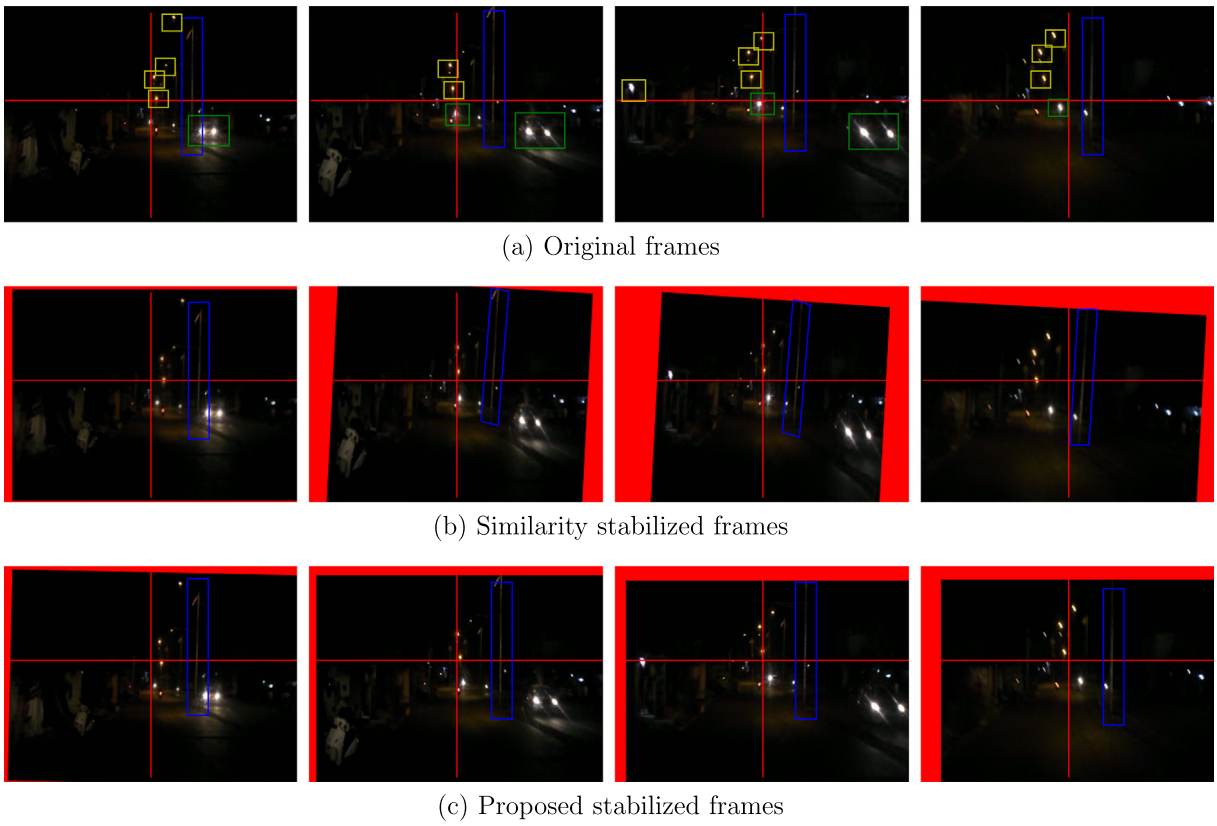
$$SSIM(x,y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{\left(\mu_x^2 + \mu_y^2 + C_1\right)\left(\sigma_x^2 + \sigma_y^2 + C_2\right)} \tag{13}$$

where $\mu_x, \sigma_x$ and $\sigma_{xy}$ are the mean, standard deviation and cross correlation between two patches, and constants $C_1 = (K_1L)^2$ and $C_2 = (K_2L)^2$ having $L = 255$, $K_1 = 0.01$ and $K_2 = 0.003$ are used to avoid the null denominator. *SSIM* over full frame measured as the mean of local *SSIM* indices (MSSIM), obtained with a sliding image window, is given by (14).

$$MSSIM(I_f, I_{f-1}) = \frac{1}{B}\sum_{j=1}^{B} SSIM\left(I_f^j, I_{f-1}^j\right) \tag{14}$$

where $I_f$ and $I_{f-1}$ are the present and the previous frames, respectively; $I_f^j$ and $I_{f-1}^j$ are the frame-contents at the $j$th local window, and $B$ is the number of local windows in the frame [28].

Interframe Transformation Fidelity (ITF) in terms of averaged *PSNR* over entire frame sequence has been widely used for video stability evaluation, but due to it's limited perceptual quality assessment, a new *MSSIM* based video-stability measure *Interframe*



(a) Original frames

(b) Similarity stabilized frames

(c) Proposed stabilized frames

**Fig. 11.** Original and stabilized frame for video *GlobalIntVar*.

*Stability Factor* (ISF) is hereby defined as the average *MSSIM* taken over all the frames. *ITF* and *ISF* are given by (15) and (16) respectively.

$$ITF = \frac{1}{F-1}\sum_{f=2}^{F} PSNR(I_f, I_{f-1}) \tag{15}$$

$$ISF = \frac{1}{F-1}\sum_{f=2}^{F} MSSIM(I_f, I_{f-1}) \tag{16}$$

where $F$ is the total number of frames in the video sequence. *ITF* and *ISF* objectively assess the stabilization brought by the video stabilization algorithm and produces higher values for stabilized video than the original video.

A comparative result analysis for the *ITF* (in *decibel*) and *ISF* values of different real-world night videos, computed using various projection based stabilization algorithms is presented in Tables 1 and 2 respectively. The warping based methods give improved *ITF* and *ISF* in comparison to integral projection correlation (IP) [17] method. IP provides comparable stabilization to classical warping [21] for static-scene and small moving objects but the results degrade under intensity variation and large moving objects. The CW [21] has a drawback of large processing time and the optimal path ambiguity, which is very dominant in with night videos due to the zero distance regions. The dynamic programming based DTW [22] overcomes the path ambiguity issue and gives efficient motion estimation under poor texture conditions. Differential projections using the derivative as its shape feature provide better warping under unsteady moving platform [25], and give higher *ITF* and *ISF* values in comparison to the intensity projections [17,21,22]. The least square solution over the derivative warping vectors has been utilized for similarity motion stabilization [24], but miswarped vectors lead to wrong motion estimation and degrades the video quality adversely. Smoothed derivative warped under constrained-search regions are suggested for better projection alignment with sufficient processing time reduction [26]. Overcoming the boundary tracing problem of SA-CDDTW (see Section 3.1), the proposed method dual accumulation method results better under poor textures like noise and blur and gives added robustness to dark regions. The boundary vector elimination in the proposed technique discards most of the boundary mismatched vectors leading to better accuracy and gives higher *ITF* and *ISF* of all.

A comparative processing time analysis for the different methods is given in Table 3. All the experiments are done using MATLAB on a 2.40 GHz processor. The Bosco *et al.* technique [21] takes large processing time due to classical warping approach. Other full-frame warping algorithms take more time in comparison to the constrained ones. The IP [17] takes smallest processing time but at the cost of limited motion accuracy. The proposed DA-CDDTW method requires more processing time in comparison to SA-CDDTW [26] due to repeated matrix accumulation, but provides better stability under the challenging cases.

## 7. Conclusion

A robust digital motion estimation technique dedicated to night shooting videos has been presented. This paper investigates that the presence of frame blur, noise and large dark areas pose challenges for accurate image alignment in the night videos. A new projection warping algorithm based of dual-accumulation over constrained warping matrix has been presented to overcome these challenges and provides accurate motion estimation in case of night recordings. Use of projections handles blur and noise (imaging and/or sensor) in the frames while the dual accumulated warp-

**Table 1**
Interframe Transformation Fidelity Factor (ITF) values of different night videos.

| Test videos | Interframe transformation fidelity (decible) | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | Orig. ITF | Stabilized video ITF | | | | | | |
| | | [17] | [21] | [22] | [24] | [25] | [26] | Proposed |
| *DarkStatic* | 28.063 | 30.590 | 30.628 | 30.895 | 25.160 | 31.082 | 29.935 | 31.694 |
| *SmallMovObj* | 23.409 | 25.836 | 25.859 | 25.981 | 24.290 | 26.228 | 26.373 | 26.763 |
| *LargeMovObj* | 22.552 | 26.473 | 26.953 | 26.988 | 20.013 | 27.118 | 26.219 | 27.526 |
| *LocalIntVar* | 22.529 | 25.951 | 26.186 | 26.325 | 22.991 | 26.868 | 26.280 | 26.907 |
| *OnRoadMov* | 20.302 | 25.133 | 25.668 | 25.906 | 18.639 | 25.902 | 25.935 | 25.986 |
| *GlobalIntVar* | 28.645 | 32.306 | 32.914 | 32.945 | 26.828 | 33.157 | 33.281 | 33.579 |

**Table 2**
Interframe Stability Factor (ISF) values of different night videos.

| Test videos | Interframe stability factor | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | Orig. ISF | Stabilized video ISF | | | | | | |
| | | [17] | [21] | [22] | [24] | [25] | [26] | Proposed |
| *DarkStatic* | 0.9137 | 0.9378 | 0.9385 | 0.9391 | 0.7704 | 0.9403 | 0.9359 | 0.9427 |
| *SmallMovObj* | 0.8961 | 0.9005 | 0.9052 | 0.9068 | 0.8981 | 0.9073 | 0.9061 | 0.9099 |
| *LargeMovObj* | 0.7802 | 0.8646 | 0.8683 | 0.8695 | 0.6435 | 0.8730 | 0.8621 | 0.8748 |
| *LocalIntVar* | 0.8543 | 0.9043 | 0.9082 | 0.9088 | 0.8690 | 0.9091 | 0.9064 | 0.9096 |
| *OnRoadMov* | 0.8455 | 0.8954 | 0.8979 | 0.8983 | 0.7211 | 0.8928 | 0.8991 | 0.8999 |
| *GlobalIntVar* | 0.8913 | 0.9148 | 0.9179 | 0.9185 | 0.7338 | 0.9189 | 0.9194 | 0.9213 |

**Table 3**
Comparative results of processing speed.

| Test videos $240 \times 320$ | Processing time (seconds/frame) | | | | | | |
|---|---|---|---|---|---|---|---|
| | [17] | [21] | [22] | [24] | [25] | [26] | Proposed |
| *DarkStatic* | 0.064 | 36.02 | 0.194 | 0.461 | 0.196 | 0.071 | 0.086 |
| *SmallMovObj* | 0.066 | 37.91 | 0.193 | 0.490 | 0.203 | 0.076 | 0.093 |
| *LargeMovObj* | 0.073 | 39.01 | 0.198 | 0.503 | 0.213 | 0.081 | 0.096 |
| *LocalIntVar* | 0.069 | 38.23 | 0.190 | 0.483 | 0.199 | 0.074 | 0.091 |
| *OnRoadMov* | 0.068 | 37.59 | 0.195 | 0.459 | 0.198 | 0.073 | 0.088 |
| *GlobalIntVar* | 0.071 | 38.36 | 0.196 | 0.488 | 0.201 | 0.074 | 0.091 |

ing handles the effect of dark regions. The stabilization scheme can be used for normal illumination videos as well without degrading its motion accuracy. Efficiency and robustness of the proposed techniques are validated using experimentation over various categories of real-world night-shooting videos.

## References


[1] M. Oshima, T. Hayashi, S. Fujioka, T. Inaji, H. Mitani, J. Kajino, K. Ikeda, K. Komoda, VHS camcorder with electronic image stabilizer, IEEE Trans. Consumer Electron. 35 (4) (1989) 749–758.

[2] K. Sato, S. Ishizuka, A. Nikami, M. Sato, Control techniques for optical image stabilizing system, IEEE Trans. Consum. Electron. 39 (3) (1993) 461–466.

[3] Y.-X. Zhang, W.-G. Zhang, X.-X. Zhao, H.-M. Yuan, Study on electronic image stabilization system based on MEMS gyro, in: Proc. Int'l Conf. on Electronic Computer Technology (ICECT), 2009, pp. 641–643.

[4] T. Kinugasa, N. Yamamoto, H. Komatsu, S. Takase, T. Imaide, Electronic image stabilizer for video camera use, IEEE Trans. Consum. Electron. 36 (3) (1990) 520–525.

[5] A. Karpenko, D. Jacobs, J. Baek, M. Levoy, Digital video stabilization and rolling shutter correction using Gyroscopes, Stanford CS Tech Report, 2011.

[6] C. Jia, B.L. Evans, Online camera-gyroscope auto-calibration for cell phones, IEEE Trans. Image Process. 23 (12) (2014) 5070–5081.

[7] G. Hua, J. Jose, Breakthroughs in low-light performance illuminate IP video camera applications, Texas Instruments White Paper, 2012.

[8] K. Uomori, A. Morimura, H. Ishii, T. Sakaguchi, Y. Kitamura, Automatic image stabilizing system by full-digital signal processing, IEEE Trans. Consum. Electron. 36 (3) (1990) 510–519.

[9] S.-J. Ko, S.-H. Lee, K.-H. Lee, Digital image stabilizing algorithms based on bit-plane matching, IEEE Trans. Consum. Electron. 44 (3) (1998) 617–622.

[10] S.-J. Ko, S.-H. Lee, S.-W. Jeon, E.-S. Kang, Fast digital image stabilizer based on Gray-coded bit-plane matching, IEEE Trans. Consum. Electron. 45 (3) (1999) 598–603.

[11] F. Vella, A. Castorina, M. Mancuso, G. Messina, Digital image stabilization by adaptive block motion vectors filtering, IEEE Trans. Consum. Electron. 48 (3) (2002) 796–801.

[12] J.Y. Chang, W.F. Hu, M.H. Cheng, B.S. Chang, Digital image translational and rotational motion stabilization using optical flow technique, IEEE Trans. Consum. Electron. 48 (1) (2002) 108–115.

[13] R. Hu, R. Shi, I.-F. Shen, W. Chen, Video stabilization using scale-invariant features, in: Proc. 11th Int'l Conf. on Information Visualization, 2007, pp. 871–877.

[14] B. Pinto, P.R. Anurenjan, Video stabilization using speeded up robust features, in: Proc. Int'l Conf. on Communications and Signal Processing, 2011, pp. 527–531.

[15] M. Okade, P.K. Biswas, Video stabilization using maximally stable extremal region features, Mult. Tools Appl. 68 (3) (2014) 947–968.

[16] K. Ratakonda, Real-time digital video stabilization for multi-media applications, in: Proc. IEEE International Symposium on Circuits and Systems (ISCAS'98), 1998, pp. 69–72.

[17] S. Piva, M. Zara, G. Gera, C.S. Regazzoni, Color-based video stabilization for real-time on-board object detection on high-speed trains, in: Proc. IEEE Conf. on Advanced Video and Signal Based Surveillance (AVSS), 2003, pp. 299–304.

[18] F. Albu, C. Florea, A. Zamfir, A. Drimbarean, Low complexity global motion estimation techniques for image stabilization, in: Proc. Int'l Conf. on Consumer Electronics (ICCE'08), 2008, pp. 1–2.

[19] B.B.-Mohamadabadi, A.B.-Khaligh, Digital video stabilization using radon transform, in: Proc. International Conference on Digital Image Computing Techniques and Applications (DICTA'12), 2012, pp. 1–8.

[20] M. Muller, Information Retrieval for Music and Motion, Springer-Verlag New York, Inc., Secaucus, NJ, 2007.

[21] A. Bosco, A. Bruna, S. Battiato, G. Bella, G. Puglisi, Digital video stabilization through curve warping techniques, IEEE Trans. Consum. Electron. 54 (2) (2008) 220–224.

[22] D. Shukla, R.K. Jha, A robust video stabilization technique using integral frame-projection warping, Signal Image Video Proces. 9 (6) (2015) 1287–1297.

[23] E.J. Keogh, M.J. Pazzani, Derivative dynamic time warping, in: Proc. The First SIAM Int'l Conf. on Data Mining (SDM'2001), 2001, pp. 1–11.

[24] M. Veldandi, S. Ukil, K.G. Rao, Video stabilization by estimation of similarity transform from integral projection, in: Proc. IEEE Int'l Conf. on Image Processing (ICIP'13), 2013, pp. 785–789.

[25] D. Shukla, R.K. Jha, A robust on-road moving platform video stabilization using derivative curve warping, in: Proc. 5th Int'l Conf. on Pattern Recognition and Machine Intelligence (PReMI'13), LNCS, vol. 8251, Springer-Verlag, Berlin, Heidelberg, 2013, pp. 343–348.

[26] D. Shukla, R.K. Jha, An optimized derivative projection warping approach for moving platform video stabilization, in: Proc. Fourth National Conference on Computer Vision, Pattern Recognition, Image Processing and Graphics (NCVPRIPG'13), 2013, pp. 1–4.

[27] Y. Matsushita, E. Ofek, W. Ge, X. Tang, H.-Y. Shum, Full-frame video stabilization with motion inpainting, IEEE Trans. Pattern Anal. Mach. Intell. 28 (7) (2006) 1150–1163.

[28] Z. Wang, A.C. Bovik, H.R. Sheikh, E.P. Simoncelli, Image quality assessment: from error visibility to structural similarity, IEEE Trans. Image Process. 13 (4) (2004) 600–612.