NON-UNIFORM VIDEO TIME-LAPSE METHOD BASED ON MOTION SCENARIO AND STABILIZATION CONSTRAINT

Kai Guo, Nakhoon Kim, Duckchan Seo, Irina Kim, Soonkeun Chang, Dong-Ki Min and Sukhwan Lim

Corresponding author: Kai Guo, visionkai@gmail.com

Samsung electronics, South Korea

ICASSP 2020,
May 4-8, 2020

Non-uniform video time-lapse method based on motion scenario and stabilization constraint

- The necessity of collaboration between time lapse and video stabilization

- Video time-lapse becomes popular in mobile devices recently, usually time-lapse and digital video stabilization (VS) are independent sequential modules to get final output.

Video frames acquisition → Time-lapse sampling → Video stabilization → Output

The common flowchart of video time-lapse method in mobile device

- However, non-uniform sampling may produce large sampling interval and then result in larger motion, this would beyond the stabilization ability of VS and produce unpleasant output. It is necessary to collaborate time lapse and video stabilization closer so as to get better output.

Non-uniform video time-lapse method based on motion scenario and stabilization constraint

- Motivation and observations of the proposed time-lapse method

  - Use minimum frames to represent maximum contents, and achieve the best stability.

  - User-held camera is shaky in walking and running scenarios, and exhibits regular motion patterns. Sampling frames near to the camera stabilized trajectory would be helpful.

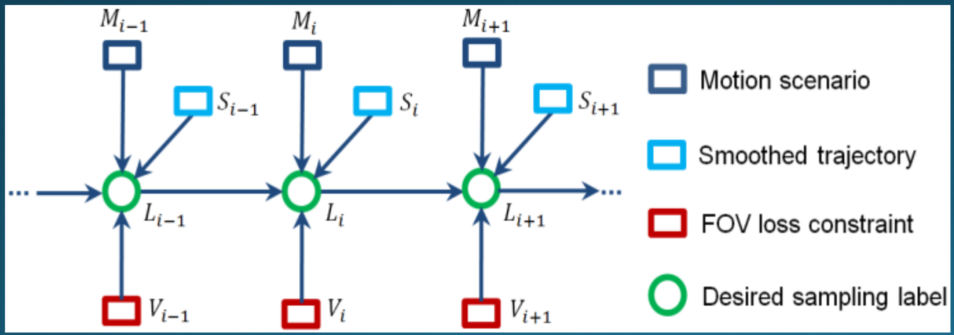
  - Sampling interval should not be too large, otherwise would produce large motion between sampled frames and be out of VS ability. In some scenarios, e.g. panning, tripod etc., there is no obvious motion pattern, we want to set a tunable constraint range for sampling interval to keep the flexibility.

  - To avoid the fluctuation of sampling, the sampling interval among consecutive frames should be smoothly changed.

Non-uniform video time-lapse method based on motion scenario and stabilization constraint

- Proposed time-lapse method based on motion scenario and stabilization constraint. The whole algorithm can work in real time during video recording on mobile device.

To fully utilize these observations and motivations, we propose a new auto time-lapse framework, which selects frames not only based on camera motion scenarios, but also refer to the smoothed camera trajectory and considering VS ability according to Field of View (FOV) loss constraint. More specific, we introduce an advanced Markov chain (MC) model, in which smoothed camera trajectory, FOV loss constraint of VS, camera motion scenario, and sampling interval similarity between consecutive frames are encoded as potential functions. Finally, dynamic programming (DP) is employed to find the best non-uniform sampling. The whole algorithm can work in real time during video recording on mobile device.



Proposed advanced Markov Chain (MC) model for auto time-lapse

Non-uniform video time-lapse method based on motion scenario and stabilization constraint

- Proposed advanced Markov Chain (MC) model

In the proposed advanced Markov chain model, with respect to *i-th* frame, the motion scenario M_i which is recognized based on camera global motion, the VS FOV loss amount V_i, and the smoothed camera trajectory S_i work as observed nodes in the proposed model. The desired sampling label L_i indicating the *i-th* frame whether is sampled is used as hidden node. The joint probability can be written as follows:

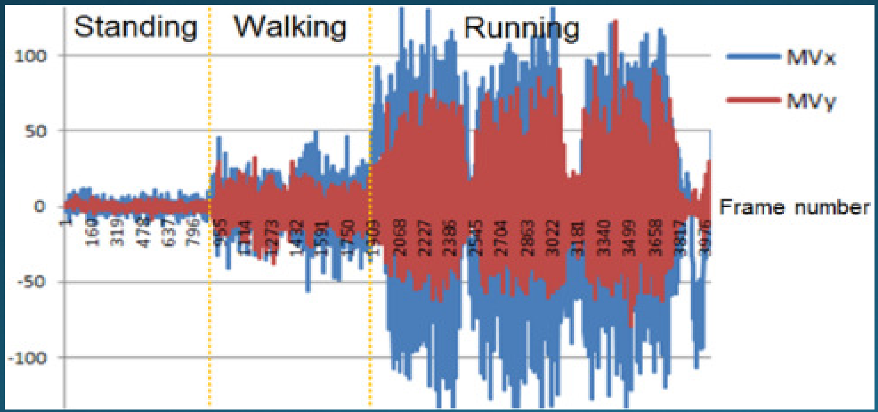
$$P(L, M, S, V) = \frac{1}{Z}\prod_{i=1}^{N} P(L_i|L_{i-1})P(L_i|V_i)P(L_i|S_i)P(L_i|M_i)$$

where Z is a normalization factor, N is the total frames number of input video. If current *i-th* frame is sampled, L_i is set to 1, otherwise, it is set to 0. We use $\mathrm{I}(L_i)$ to represent sampling interval on the basis of *i-th* frame sampling label L_i, it is the skipped frame number from last sampled frame to current frame. If the current frame is skipped, L_i becomes 0, and $\mathrm{I}(L_i)$ would be same as last sampled frame.

Non-uniform video time-lapse method based on motion scenario and stabilization constraint

- Motion scenario recognition

Camera motion magnitude can reflect different motion scenarios (as shown in Figure below), we use a linear classifier to recognize each frame motion scenario based on frame global motion magnitude, e.g. tripod, standing, walking, running. The tripod scenario is further classified into two scenarios: with local motion and without local motion, based on global motion matching residue. To detect camera panning scenario, we count the consecutive frames with the same motion direction. If the consecutive frames with the same motion direction are more than a threshold, we assume that it is a panning scenario.



Motion scenario recognition by motion magnitude

Non-uniform video time-lapse method based on motion scenario and stabilization constraint

- Motion scenario constraint

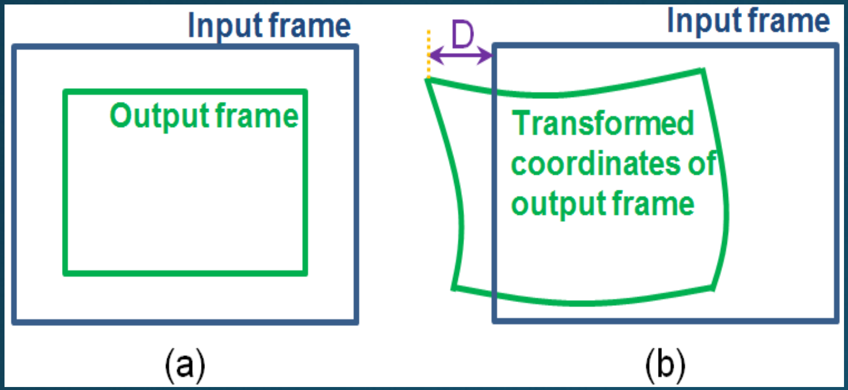
For each motion scenario M, a sampling interval range $[I_{min}(M), I_{max}(M)]$ is assigned. The sampling interval which is decided by sampling label will be uniformly distributed within the assigned interval range, so the potential function $P(L_i|M_i)$ can be written as

$$P(L_i|M_i) = \frac{1}{I_{max}(M_i) - I_{min}(M_i) + 1} \qquad L_i \in [I_{min}(M), I_{max}(M)]$$

To keep the useful information, the sampling interval in the panning scenario should be very small. Also, in order to keep the stability, the sampling interval in running scenario will be smaller compared to the standing and walking scenarios. Moreover, the static scenario without local motion (tripod mode) uses the highest sampling interval to maximize the compression ratio.

Non-uniform video time-lapse method based on motion scenario and stabilization constraint

- FoV loss constraint from video stabilization



(a) Input and output frame comparison in video stabilization. (b) Maximum distance D between transformed coordinates of output frame and input frame boundary.

Generally, digital VS achieves stability at the expense of FOV loss, the input frame resolution of VS is larger than the stabilized output (as shown in Figure above). VS FOV loss amount V_i can be used as a constraint, so as to not only avoid the motion between sampled frames go out of VS compensation ability, but also let the subsequent VS fully utilize the FOV loss amount.

Non-uniform video time-lapse method based on motion scenario and stabilization constraint

- FoV loss constraint from video stabilization

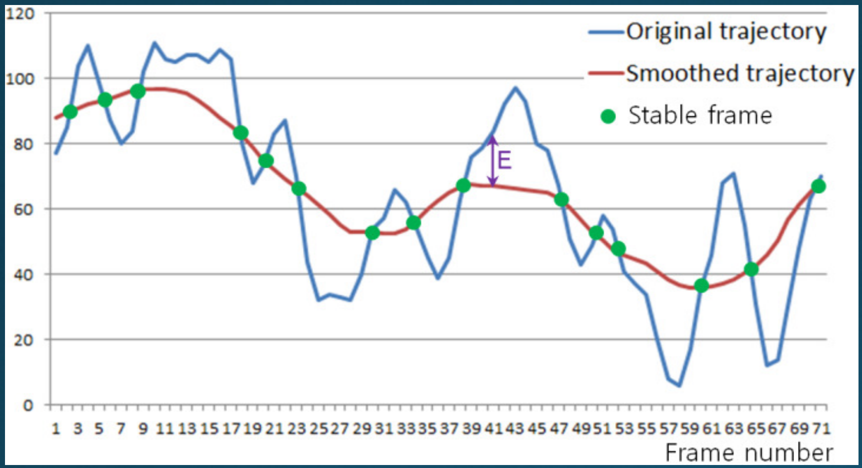
Then the potential function $P(L_i|V_i)$ can be written as

$$P(L_i|V_i) = exp\left(-\frac{\left\|D\left(W_{L_i}(x,y)\right)\right\|^2}{2\sigma_\phi^2}\right)$$

where W is the transformation between *i-th* frame and the last sampled frame. Because rolling shutter distortion is taken into account, W is the combination of several warping matrices in a scan-line by scan-line manner [6]. The interval between *i-th* frame and last sampled frame is determined by sampling label L_i. D is the function which measures the distance between the warped coordinates by W and the input frame boundary of VS (as shown in Figure in last page). Given pixel coordinates (u,v) of the output frame and the transformation matrix W, if all transformed coordinates are inside the input frame boundary, D is set to 0; otherwise, D is calculated as a maximum distance between $W(u,v)$ and input frame boundary. As D is smaller, the factor $P(L_i|V_i)$ gives higher probability. In other words, this function encourages the sampling process use largest sampling interval to achieve the highest stability. σ_ϕ is the standard variance to tune the strength of D.

Non-uniform video time-lapse method based on motion scenario and stabilization constraint

- Smoothed trajectory constraint



Original and smoothed trajectory. Stable frames near with smoothed trajectory are encouraged to be selected. Euclidean distance E is measured between two trajectories.

As aforementioned, we want to sample stable frames which are near with the smoothed trajectory. Also for commercialization, our auto time-lapse algorithm should work in real time. To these ends, we first accumulate camera trajectory T from frames global motion vectors. Then the 1st order IIR (Infinite Impulse Response) filter [7] is employed to get smoothed trajectory S.

Non-uniform video time-lapse method based on motion scenario and stabilization constraint

- Smoothed trajectory constraint

After that the Euclidean distance E_i (as shown in Figure of last page) is used to measure how far the current i-th frame is away from the smoothed trajectory:

$$E_i = \sqrt[2]{\left(T_i(x) - S_i(x)\right)^2 + \left(T_i(y) - S_i(y)\right)^2 + \left(T_i(z) - S_i(z)\right)^2}$$

where x, y and z represents trajectory along horizontal translation, vertical translation and z rotation, respectively. The potential function $P(L_i|S_i)$ can be written as

$$P(L_i|S_i) = exp\left(-\frac{(E_i)^2}{2\sigma_{S_i}^2}\right)$$

where σ_s is the standard variance to control the efficiency of E.
This potential function value would be larger as E_i becomes smaller. It means that this function prefers to select frames which are closer with the smoothed trajectory.

Non-uniform video time-lapse method based on motion scenario and stabilization constraint

- Smoothing sampling constraint

As mentioned, the sampling interval among consecutive frames should be smoothly changed. We define the potential function $P(L_i|L_{i-1})$ which constrains the sampling interval change between successive frames, and it is written as:

$$P(L_i|L_{i-1}) = exp\left(-\frac{\|I(L_i) - I(L_{i-1})\|^2}{2\sigma_\varphi^2}\right)$$

where σ_φ is the standard variance. This potential function will return the larger value as sampling interval similarity between successive frames becomes higher.

Non-uniform video time-lapse method based on motion scenario and stabilization constraint

- Dynamic programming

Once all potential functions are determined, the objective is to maximize $P(L,M,S,V)$ so as to infer the best sampling interval for each frame. Because the potential functions$P(L_i|V_i)$, $P(L_i|S_i)$ and $P(L_i|L_{i-1})$ have exponential forms, the objective can be transformed to minimize the following cost function:

$$C(L,M,S,V) = \sum_{i=1}^{N}\left(a \times D\left(W_{L_i}(x,y)\right)^2 + b \times (E_i)^2 + c \times \left(I(L_i) - I(L_{i-1})\right)^2\right)$$

where $L_i \in [I_{min}(M),\ I_{max}(M)]$. The weighting coefficients a, b and c determine the trade-off among FOV loss constraint, stable frames selection and sampling interval continuity, and they are controlled by the standard variance $\sigma_{\emptyset}, \sigma_S, \sigma_{\varphi}$. In the standing scenarios, because the value of cost W and E is smaller compared with walking, running and panning scenarios, the coefficients a and b can be set to larger values.

Non-uniform video time-lapse method based on motion scenario and stabilization constraint

- Dynamic programming

The Dynamic Programming (DP) method [8] is employed to get best sampling interval (as shown in Figure 5). Given scenario recognition results M, camera original trajectory T, smoothed trajectory S, assigned sampling interval range [$I_{min}(M)$, $I_{max}(M)$] and total frames number N, DP can be performed in a piece-by-piece manner, where each piece size is $I_{max}(M)$, and there are two stages in each piece. In stage 1, all the frames cost values $C_i(L, M, S, V)$ are calculated. In stage 2, a frame with minimum cost value is selected as a best sample. Because all of the M, T, S, $I_{min}(M)$, $I_{max}(M)$ can be estimated with frame-by-frame manner, sampling the best frames by dynamic programming can be performed in real time.

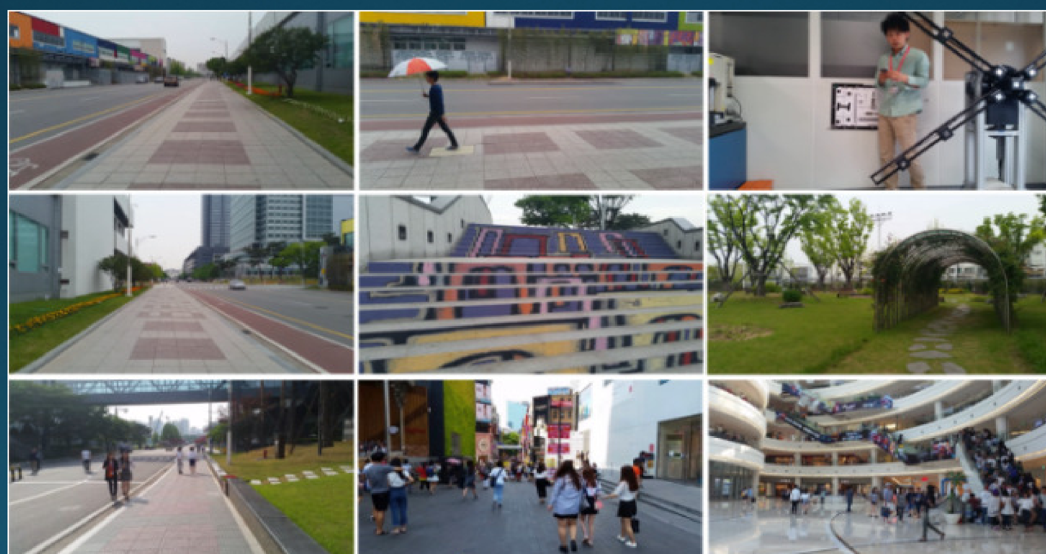
Input: M, T, S, $I_{min}(M)$, $I_{max}(M)$, N
Initialization:
  Set the 1st frame label as 1
  p = 1
while $\left(p + I_{max}(M)\right) < N$ **do**
  First stage: calculate cost function
    for j = 1 to $I_{max}(M)$ **do**
      i = p + j
      calculate $D\left(W_{L_i}(x, y)\right)^2, (E_i)^2, \left(I(L_i) - I(L_{i-1})\right)^2$
      calculate $C_i(L, M, S, V)$
    end for
  Second stage: select sample with minimum cost
    $v = arg \min_{j=I_{min}(M) \text{ to } I_{max}(M)} C_{j+p}(L, M, S, V)$
    p = p+v
end while

Real-time Dynamic programming for our time-lapse

Non-uniform video time-lapse method based on motion scenario and stabilization constraint

- Experimental results

We first justify the effectiveness of adopting motion-scenario-determined sampling interval and the effectiveness of constraint of sampling interval similarity. Then we verify the effectiveness of FOV loss constraint. At last, we compare the proposed method with the regular time-lapse method to validate the superiority of our proposed algorithm.

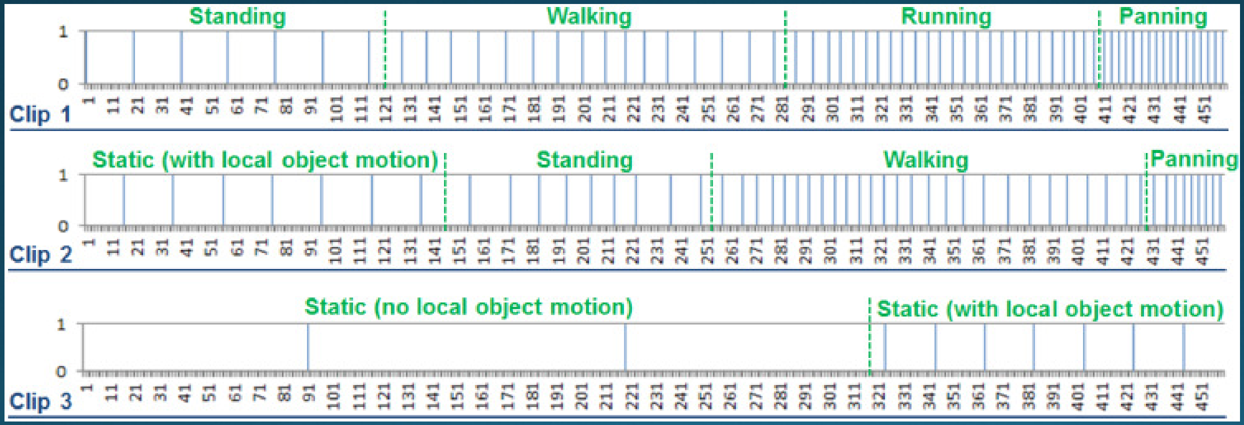


Test videos for performance evaluation.

Non-uniform video time-lapse method based on motion scenario and stabilization constraint

- Experimental results

In the first experiment, we apply our algorithm on three user-captured short video clips, including static, standing, walking, running, panning scenarios etc. (as shown in top row of Figure of last page). The visualizations of sampling interval results are shown in Figure below. We can see that our algorithm not only adaptively estimates sampling interval for different motion scenarios, but also generates smooth sampling interval change, especially during scenario transition.



Sampling interval. Each vertical and horizontal line represents a sampled frame and frame number, respectively.

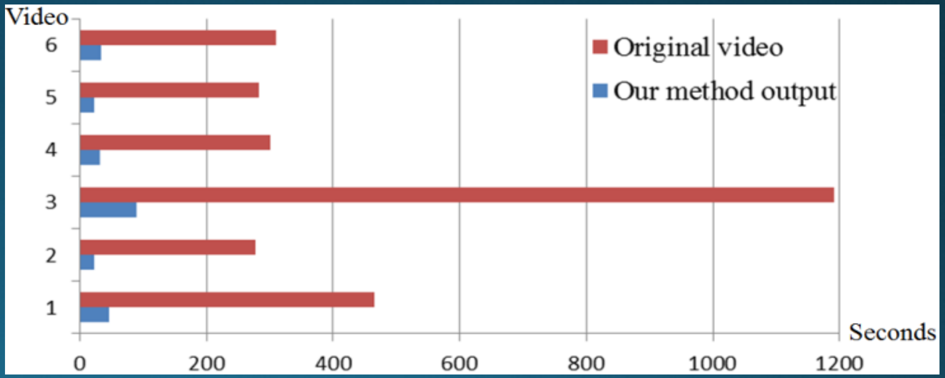
Non-uniform video time-lapse method based on motion scenario and stabilization constraint

- ## Experimental results

Then we validate the stabilization ability of the proposed method by comparing to disabling FOV loss constraint. A VS method [9] is used to stabilize the output. Then inter-frame transformation fidelity (ITF) is calculated on the output videos, which is the average value of PSNR between successive frames. The ITF comparison results on remaining 6 user captured videos clips (including static, standing, walking, running, panning scenarios etc) are shown in Table below. It can be seen that enabling FOV loss constraint achieves higher ITF value and thus produces more stabilized video. The video lengths of our method and these 6 original videos are shown in Figure below.

	Without FOV	With FOV	Improvement
Video 1	19.7536	19.9024	0.1488
Video 2	13.4753	13.5006	0.0253
Video 3	17.7563	17.9417	0.1854
Video 4	22.2465	22.2679	0.0214
Video 5	18.8641	18.8853	0.0212
Video 6	18.9179	19.0454	0.1275
Average	18.5023	18.5906	0.0883

ITF comparison

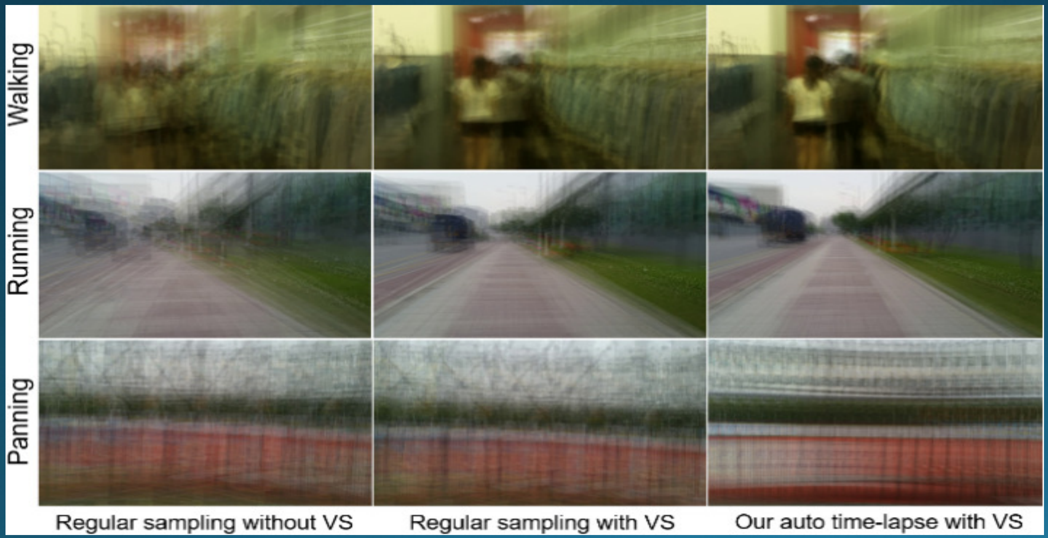


Video length comparison.

Non-uniform video time-lapse method based on motion scenario and stabilization constraint

- Experimental results

We select three videos which are corresponding to walking, running and panning scenarios, respectively. Then 10~15 consecutive frames in each video output are integrated into one image. The integrated images comparisons are shown in Figure below. Compared with regular sampling, our auto time-lapse in walking and running scenarios can generate sharper image and less ghost artifacts in the integrated image. In panning scenario, our proposed method can keep more meaningful frames, therefore the integrated image of our method is more elegant and reflects more useful structure information. It means that our method can select more stable frames and thus generate more stable and comfortable output.



Integrated frames comparison.

Non-uniform video time-lapse method based on motion scenario and stabilization constraint

- Conclusion

In this paper, we proposed a real-time auto time-lapse method, that achieves the meaningful non-uniform sampling based on motion scenario, and provides satisfactory stabilization performance by stable frames selection. Moreover, the proposed method can work in real time during video recording on mobile device, thus the extra memory space for original video is saved. This is very useful when user want to capture long videos. In the future, we will develop event and saliency detection based auto time-lapse, to extract more meaningful stabilized frames.

Thanks