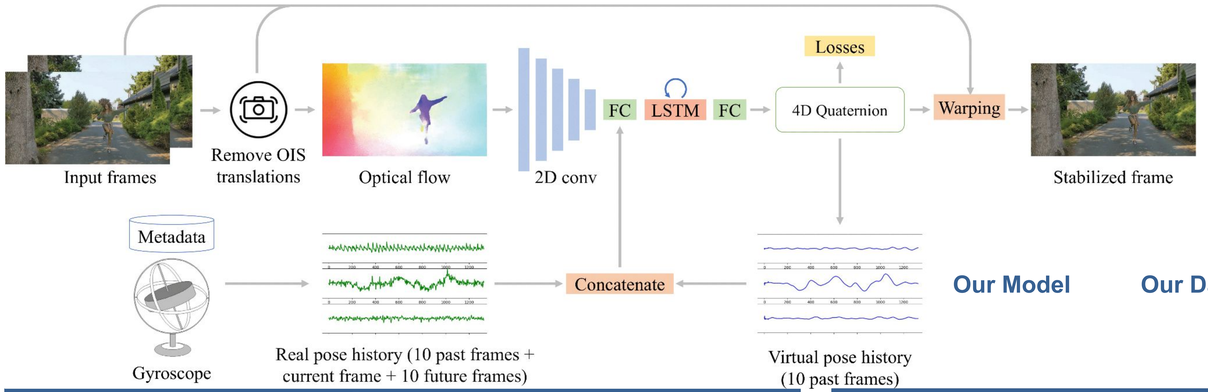
Deep Online Fused Video Stabilization

Zhenmei Shi[1]; Fuhao Shi[2]; Wei-Sheng Lai[2]; Chia-Kai Liang[2]; Yingyu Liang[1]
[1]University of Wisconsin-Madison, [2]Google

WACV VIRTUAL JANUARY 5-9



Our Model

Our Dataset

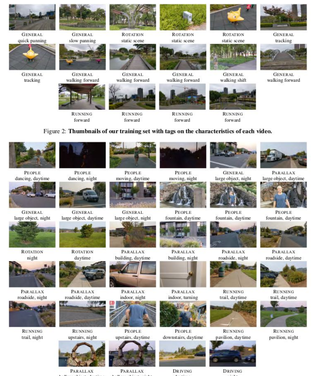
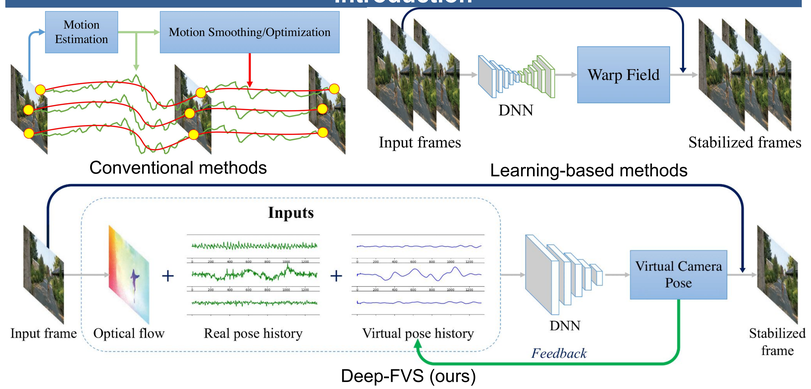


Figure 2: **Thumbnails of our training set with tags on the characteristics of each video.**

Summary

- **First DNN** framework fuses motion **sensor data** and **optical flow** for online video stabilization
- An **unsupervised** learning process with multi-stage training and relative motion representation
- A benchmark **dataset** that contains videos with gyroscope and OIS sensor data and covers various scenarios

Introduction



Method

Deep Fused Video Stabilization

- Gyroscope

$R(t_f) = \texttt{SLERP}(R(t_a), R(t_b), (t_b - t_f)/(t_b - t_a))$

- Camera pose representation

$x = K(t)R(t)X$

$x_v = K_v(t)R_v(t)R_r^{-1}(t)K_r^{-1}(t)x_r$

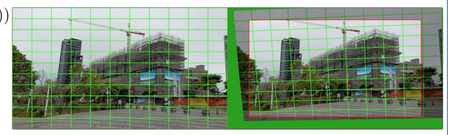
- Relative rotation based motion history

$H_v = H_{v,\text{absolute}} * R_r^{-1}(t)$

- Multi-stage training

  Sensor ⇨ Sensor + Optical Flow

- Grid-based frame warping



Loss

- Smoothness losses

$L_{C^0} = ||R_v(t) - R_v(t-\Delta t)||^2$

$L_{C^1} = ||R_v(t)R_v^{-1}(t-\Delta t) - R_v(t-\Delta t)R_v^{-1}(t-2\Delta t)||^2$

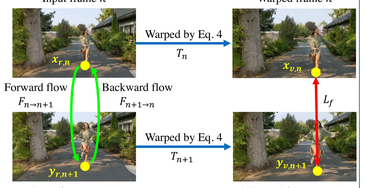
- Distortion loss

$L_d = \Omega(R_v, R_r)/(1 + e^{(-\beta_1(\Omega(R_v,R_r)-\beta_0)})$

- Protrusion loss

$L_p = \sum_{i=0}^{N} w_{p,i}||\texttt{Min}(\texttt{protrude}(P_v(t), P_r(t+i\Delta t))/\alpha, 1)||^2$

- Optical flow loss



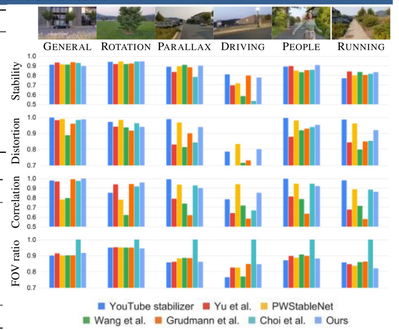
Quantitative Results

Method	Stability	Distortion	Correlation	FOV
YouTube stabilizer	0.834	**0.978**	0.969	0.977
Grundmann et al. [7]	0.818	0.896	0.948	0.635
Wang et al. [29]	0.836	0.850	0.877	0.753
PWStableNet [34]	0.830	0.965	0.973	0.934
Yu et al. [32]	0.842	0.854	0.941	0.793
Choi et al. [4]	0.781	0.875	0.916	**1.0***
Ours (sensor only)	0.846	0.888	**0.976**	0.827
Ours (sensor + flow)	**0.853**	0.937	**0.976**	0.906

Compare with SOTA

	More stable	Less distortion	Larger FOV
vs. Grundmann et al. [7]	98.4±2.3%	95.9±3.5%	82.1±6.9%
vs. Wang et al. [29]	98.4±2.3%	95.1±3.9%	77.2±7.5%
vs. PWStableNet [34]	91.1±5.1%	89.4±5.5%	48.0±9.0%
vs. Yu et al. [32]	92.7±4.7%	93.5±4.4%	38.2±8.7%
vs. Choi et al. [4]	96.7±3.2%	97.6±2.8%	27.6±8.0%
vs. YouTube stabilizer	96.7±3.2%	88.6±5.7%	23.6±7.6%
Average	95.7±1.5%	93.4±1.8%	49.5±3.6%

Results of user study



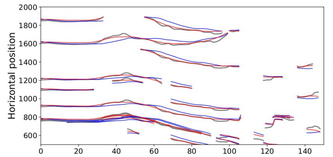
Per-category quantitative evaluation

Qualitative Results

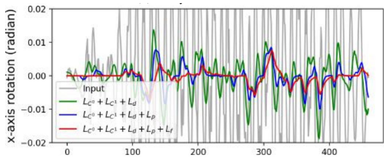
Visual comparisons

Stability comparisons

Feature trajectories before & after stabilization, Grey: input, Blue: stage 2, Red: stage 3.



Analysis loss function



References

[4] Jinsoo Choi and In So Kweon. Deep iterative frame interpolation for full-frame video stabilization. ACM TOG, 39(1):1–9, 2020.
[7] Matthias Grundmann, Vivek Kwatra, and Irfan Essa. Auto-directed video stabilization with robust l1 optimal camera paths. In CVPR, 2011.
[29] Miao Wang, Guo-Ye Yang, Jin-Kun Lin, Song-Hai Zhang, Ariel Shamir, Shao-Ping Lu, and Shi-Min Hu. Deep online video stabilization with multi-grid warping transformation learning. IEEE TIP, 2018.
[32] Jiyang Yu and Ravi Ramamoorthi. Learning video stabilization using optical flow. In CVPR, 2020.
[34] Minda Zhao and Qiang Ling. Pwstablenet: Learning pixel-wise warping maps for video stabilization. IEEE TIP, 2020.

Project Page

https://zhmeishi.github.io/dvs/

Code & Dataset & More Result

© POSTER TEMPLATE BY GENGRAPHICS: 1.800.790.4001 WWW.GENGRAPHICS.COM