Temporal-controlled Frame Swap for Generating High-Fidelity Stereo Driving Data for Autonomy Analysis

Yedi Luo*, Xiangyu Bai, Le Jiang, Aniket Gupta, Eric Mortin, Hanumant Singh, Sarah Ostadabbas

Electrical and Computer Engineering Department, Northeastern University

DEVCOM Analysis Center (DAC)

BMVC 2023

Motivation

Stereo vision data is critical for up-to-scale visual SLAM tasks. Synthetic data from research simulators offers a scalable solution to data scarcity, but **the realism of dedicated research simulators is not on par with state-of-the-art commercial games. Meanwhile, commerical games like GTA V lack native stereo support** and restrict multiple viewport rendering, severely limiting our access to high quality stereo data.

- **The goal is to develop a novel method to overcome the single viewport limitation and enable dynamic stereo data collection from commercial games.**

Method

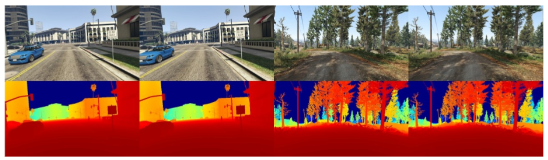
We present **Temporal-controlled Frame Swap (TeFS)**, a novel method that enables precise control over the passage of time during stereo capture process.

- Develops a customizable stereo camera setup attached to the ego vehicle.
- Leverages a single-thread architecture for precise temporal control rather than a client-server model.
- Implements a pseudo-pause function to halt the temporal state for stereo capture while keeping the rendering engine active.
- Swaps between left and right cameras at controlled moments during pseudo-pauses.
- Minimizes temporal disparity between left and right cameras from 16.7ms to 0.2-0.3ms.

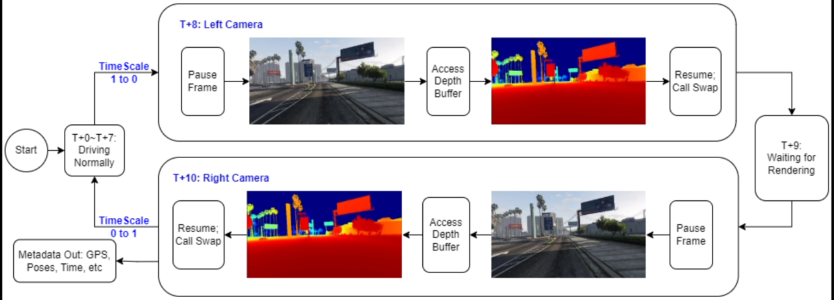
To demonstrate its generalizability, we implmented TeFS in both GTA V and CARLA. Given that CARLA is research focused and has native stereo support, it serves as a benchmark for validating TeFS Stereo.

GTAV-TeFS Dataset

We present GTAV_TeFS dataset, the first large-scale stereo driving dataset based on the high-fidelity video game, GTA V, which includes **88000 high-resolution stereo RGB image pairs, temporal information, GPS coordinates, camera poses, full-resolution dense depth maps, and challenging-case comparison groups,** facilitating up-to-scale stereo odometry tasks in a realistic commercial game environment



TeFS Stereo and Depth Capture Pipeline

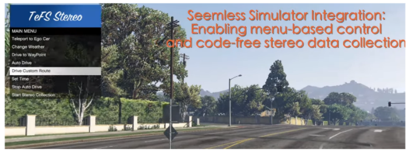


Depth and Camera fps

$$Depth = \frac{Map_{uv}}{NDC + Map_{uv} \times \frac{nc_z}{2 \times fc_z}}$$

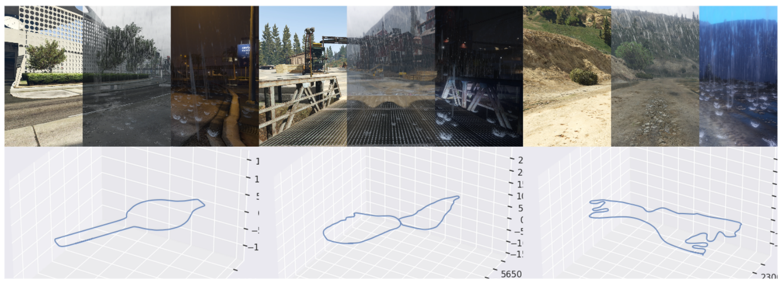
$$f_{\text{cam}} = \frac{1}{T_g \times \left(\frac{Dr}{86400}\right)}$$

Menu-based Interface



TeFS vs Native Stereo vs Simple Swap

Challenging-case Comparison Group



Experiment Results

- Validate TeFS Stereo using CARLA with ORBSLAM 3 Stereo Odometry

Map	Trajectory Distance (m)	Method	Evaluate Metrics		
			APE(m) ↓	APE(%) ↓	RPE(m) ↓
Town 03	954.41	Native	1.67±0.59	0.17%±0.06%	0.39±0.14
		TeFS	**1.82±0.65**	**0.19%±0.06%**	**0.39±0.14**
Town 06	957.97	Native	4.22±3.17	0.44%±0.33%	0.31±0.25
		TeFS	**2.30±1.67**	**0.24%±0.17%**	**0.31±0.25**
Town 10	950.05	Native	0.98±0.64	0.10%±0.07%	0.27±0.22
		TeFS	**1.01±0.63**	**0.10%±0.07%**	**0.27±0.22**

- GTAV-TeFS Stereo Odometry Eval on Feature-based vSLAM Models

Scene	Length(m)	Weather	Model	Metrics			
				APE(m)↓	APE(%)↓	RPE(m)↓	Loop Detected
City 04	760.0	Extra Sunny	ORBSLAM3	3.20±0.85	0.42±0.11	2.78±0.16	✓
			OV²SLAM	**2.53±1.21**	**0.33±0.16**	**0.06±0.09**	✓
			VINS Fusion	2.95±1.36	0.39±0.18	0.09±0.26	✓
		Cloudy with Rain	ORBSLAM3	**2.13±1.19**	**0.28±0.16**	**0.03±0.04**	✓
			OV²SLAM	21.01±9.33	2.77±1.23	0.29±2.28	✓
			VINS Fusion	3.63±2.00	0.48±0.26	0.21±0.48	✓
		Night Thunderstorm	ORBSLAM3	Lost track of the map			
			OV²SLAM	**54.19±48.40**	**7.11±6.35**	**0.59±4.15**	
			VINS Fusion	Map range unreachable			
Offroad 01	1206.0	Extra Sunny	ORBSLAM3	**3.16±1.65**	**0.26±0.14**	**0.11±0.12**	✓
			OV²SLAM	4.81±2.21	0.39±0.18	0.21±0.19	
			VINS Fusion	4.27±1.73	0.35±0.14	0.16±0.25	✓
		Cloudy with Rain	ORBSLAM3	**4.19±2.39**	**0.35±0.20**	**0.07±0.40**	✓
			OV²SLAM	6.60±4.19	0.54±0.34	0.20±0.25	
			VINS Fusion	49.75±32.98	4.12±2.73	0.38±4.73	✓
		Night Thunderstorm	ORBSLAM3	Lost track of the map			
			OV²SLAM	**91.90±45.50**	**7.57±3.79**	**0.29±2.31**	
			VINS Fusion	Map range unreachable			
Offroad 02	3230.0	Extra Sunny	ORBSLAM3	**8.19±3.95**	**0.25±0.12**	**0.39±0.26**	✓
			OV²SLAM	15.47±12.69	0.48±0.39	0.36±2.07	✓
			VINS Fusion	15.44±8.99	0.48±0.39	0.16±0.21	✓
		Cloudy with Rain	ORBSLAM3	**9.45±2.77**	**0.29±0.09**	**0.16±0.14**	
			OV²SLAM	85.62±65.35	2.65±2.03	0.14±0.21	
			VINS Fusion	72.47±61.96	2.26±1.93	0.22±0.70	
		Night Thunderstorm	ORBSLAM3	Lost track of the map			
			OV²SLAM	**160.33±82.39**	**4.97±2.49**	**0.34±3.54**	
			VINS Fusion	Map range unreachable			

- GTAV-TeFS Stereo and RGB-D Eval on Learning-basd DROIDSLAM

Scene	Weather	Input Mode	Metrics(Corrected)			Metrics(Raw)
			APE(m)↓	APE(%)↓	RPE(m)↓	APE(m)↓
City 04	Extra Sunny	Stereo	1.27±0.58	0.23±0.12	0.63±0.14	73.69±33.71
		RGBD	**0.86±0.45**	**0.11±0.06**	**0.63±0.10**	**15.77±6.79**
	Cloudy with Rain	Stereo	**1.73±0.95**	**0.23±0.12**	0.63±0.14	73.75±33.65
		RGBD	2.99±1.19	0.39±0.16	**0.62±0.10**	**13.16±5.48**
	Night Thunderstorm	Stereo	6.42±2.09	0.84±0.27	0.63±0.20	74.61±33.79
		RGBD	**4.91±2.39**	**0.64±0.31**	**0.62±0.15**	**10.29±4.77**
Offroad 01	Extra Sunny	Stereo	3.61±1.15	0.30±0.10	0.58±0.19	98.68±49.18
		RGBD	**1.35±0.54**	**0.11±0.04**	**0.58±0.18**	**19.72±9.85**
	Cloudy with Rain	Stereo	6.23±3.66	0.51±0.30	0.59±0.18	99.44±49.29
		RGBD	**2.64±1.13**	**0.22±0.09**	**0.58±0.18**	**18.59±9.23**
	Night Thunderstorm	Stereo	18.24±6.11	1.50±0.50	0.62±0.31	100.32±50.09
		RGBD	**6.68±4.42**	**0.55±0.35**	**0.62±0.19**	**21.48±12.89**

Acknowledgement: Research was sponsored by the DEVCOM Analysis Center and was accomplished under Cooperative Agreement Number W911NF-22-2-0001. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Army Research Office or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation herein.