RGB-D-E: Event Camera Calibration for Fast 6-DOF Object Tracking

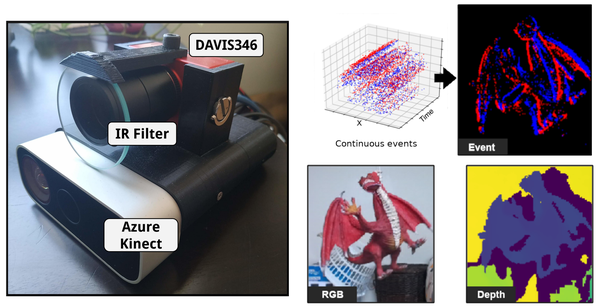
Etienne Dubeau*, Mathieu Garon*, Benoit Debaque†, Raoul de Charette¶, J.-F. Lalonde*
*Université Laval, †Thales Digital Solutions, ¶Inria

Context

Current 6-DOF trackers use mostly frame-based camera which imposes a **high data bandwidth and power usage**. Also, those methods are strongly limited by the **framerate** and **motion blur** for high-speed tracking.
We propose to use an event-based camera to increase the speed of 3D object tracking in 6 degrees of freedom.
We make the following contributions:

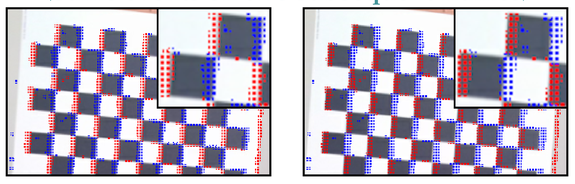
- Show how to **calibrate a RGB-D-E setup**
- Provide a new **RGB-D-E 6-DOF evaluation dataset**
- **Increase object tracking performance** on high-speed scenarios

Capture Setup



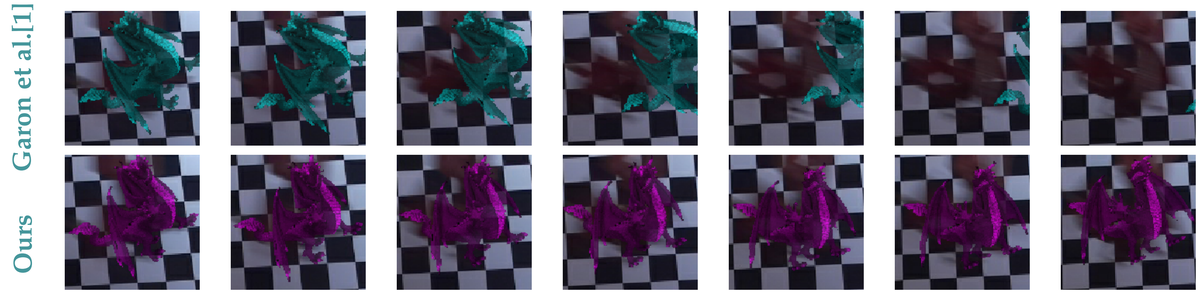
- Combine Kinect Azure (**RGB-D**) with a DAVIS346 (**event-based**)
- Spacially calibrated and temporally synchronized
- Apply linear depth correction

Demonstrated temporal and spatial calibration (Same events on two subsequent frames)



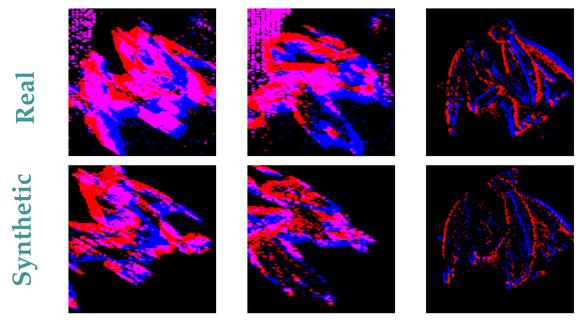
Frame t_0 Event $[t_0, t_0 + \Delta t]$ | Frame $t_0 + \Delta t$ Event $[t_0, t_0 + \Delta t]$

Result on Sequence



Training Dataset

- Train on synthetic data only
- Events are generated using **ESIM** [2] simulator



Architecture

Input: $\mathbf{e}_{[t-1,t]}$

kernel-5
conv-3-64
fire-32-64
fire-64-128
fire-128-256
fire-128-512
FC-500
FC-6

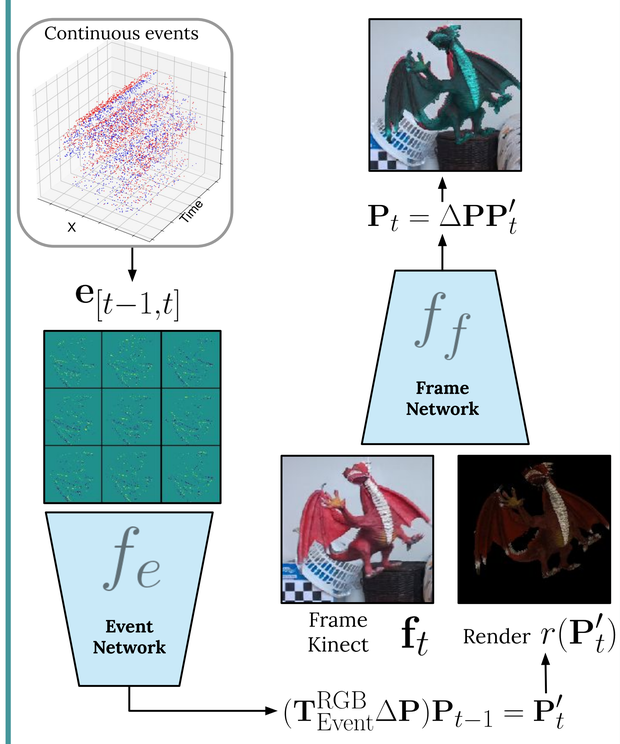
Output: ΔP

- EventSpikeTensor as input representation (**discretize in 9 temporal bins**)
- **Runtime of 3.96 ms** for the event network on Nvidia GTX 1060 GPU
- L2 loss on ΔP

[1] Garon *et al.*, A framework for evaluating 6-DOF object trackers, ECCV, 2018.
[2] Rebecq, H. *et al.*, ESIM: an Open Event Camera Simulator, CoRL, 2018.

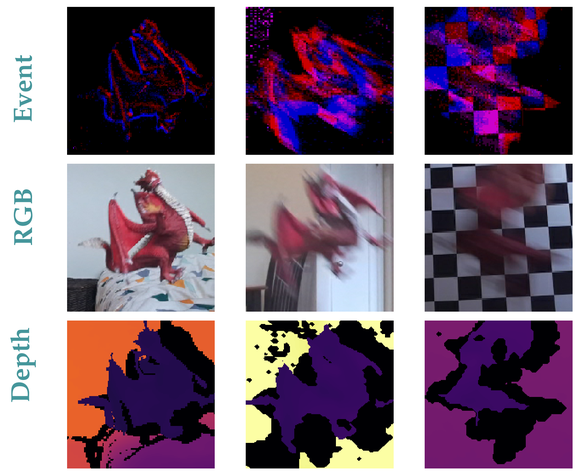
Method

- We combine two networks in cascade to compute the final position
- The **Event network** predicts the coarse position from event information
- A state of the art **RGB-D network** refines the final pose

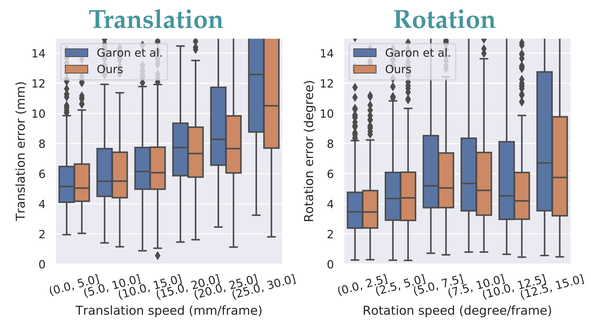


Evaluation Dataset

- We capture an RGB-D-E evaluation dataset
- **10 different sequences** with more than 2400 frames
- **Ground truth** annotated with ICP



Qualitative Result



Tracking Failures

- Tracking failures are **reduced by more than 60%**

Tracking Failures	30fps	15fps	10fps
Garon et al. [1]	83	130	166
Ours	28	48	64

This work was supported by FRQNT, FRQ-NT Samuel de Champlain grants and Thales. We thank nVidia for the donation of GPUs. Project page: `https://github.com/lvsn/rgbde_tracking`