Event cameras have low-latency, high dynamic range and no motion blur. How can we track their 6-DOF pose in natural scenes event-by-event?

Geometric Model

Given the events and scene, we can solve for the pose. The 3D scene is represented by keyframes and depth maps. Each event provides a brightness change observation $C_{th}$.

Robust Event Generation Model

Ideally, an event is generated of the brightness change is $C_{th}$. Realistically, we use a resilient mixture model:

- Good measurement: Gaussian distribution
- Noise / outliers: Uniform distribution

We model the brightness change of each event using the scene information, and we use the mixture model to provide the probability of the event being generated by the scene.

Bayesian filter:

$$q(s_k; \eta_k) \approx C \ p(o_k | s_k) q(s_k; \eta_{k-1})$$

State vector: current pose, poses and sensor parameters:

$$s = (\xi_k, \xi_{i}, \xi_{j}, C_{0i}, \pi_m, \sigma_m^2)$$

Approximate the posterior distribution in the exponential family, and minimize the Kullback-Leibler divergence to yield the filter equations:

Gain:

$$K_k = P_k J_k^T (J_k P_k J_k^T + \sigma_m^2)^{-1}$$

Weight:

$$w_k = \frac{\pi_m N(M_k, 0, \sigma_m^2)}{\pi_m N(M_k, 0, \sigma_m^2) + (1 - \pi_m) M}$$

Pose:

$$\xi_{k+1} = \xi_k + w_k K_k M_k$$

Covar:

$$P_{k+1} = (I - w_k K_k J_k) P_{k}$$

Accuracy evaluation: rocks sequences. Ground truth from mocap system.