Event-based Visual Inertial Odometry

Alex Zhu, Nikolay Atanasov, Kostas Daniilidis
University of Pennsylvania

Penn Engineering | GRASP Laboratory
General Robotics, Automation, Sensing & Perception Lab

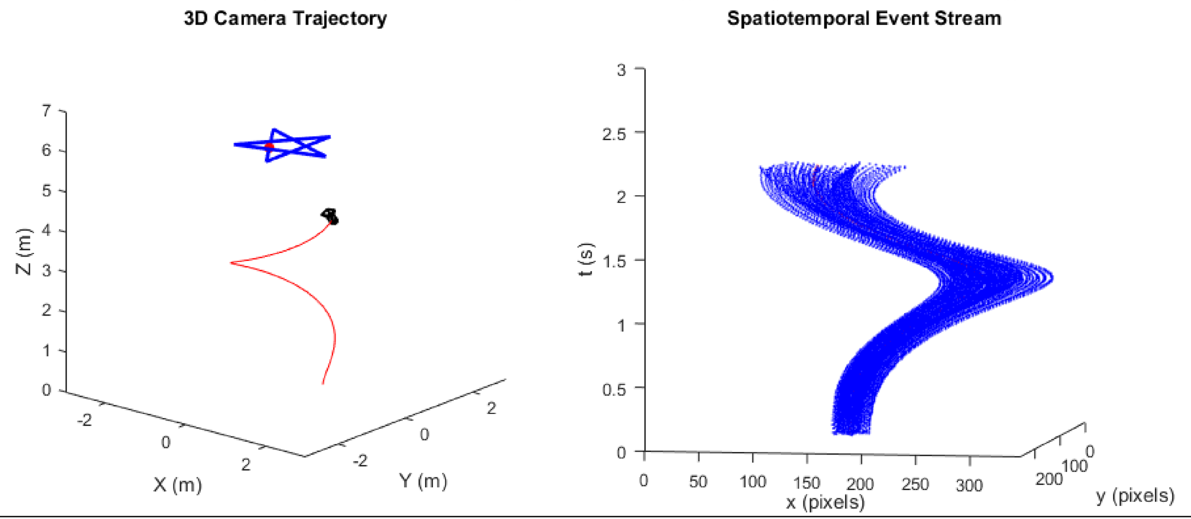
IEEE 2017 Conference on Computer Vision and Pattern Recognition
CVPR July 21-26 HONOLULU 2017

Event-based Cameras

Event-based cameras **asynchronously** capture **changes in log light intensity**. Whenever the log light intensity over any pixel changes over a set threshold, the camera **immediately** returns the pixel location of the change, a timestamp with **microsecond** accuracy, and the direction of the change (+ or -).

$$\{(x_i, t_i, p_i): |\log(I(x_i, t_i)) - \log(I(x_{i-1}, t_{i-1}))| \geq \theta\}$$

The camera exhibits extremely **low latency** and **high dynamic range**.



Motivation

Objective
Given a set of event and inertial measurements, estimate the sensor state $s(t) := [\bar{q}\ b_g\ v\ b_a\ p]$ over time.

Challenges
Unknown **data association** between events over time, no intensity information.

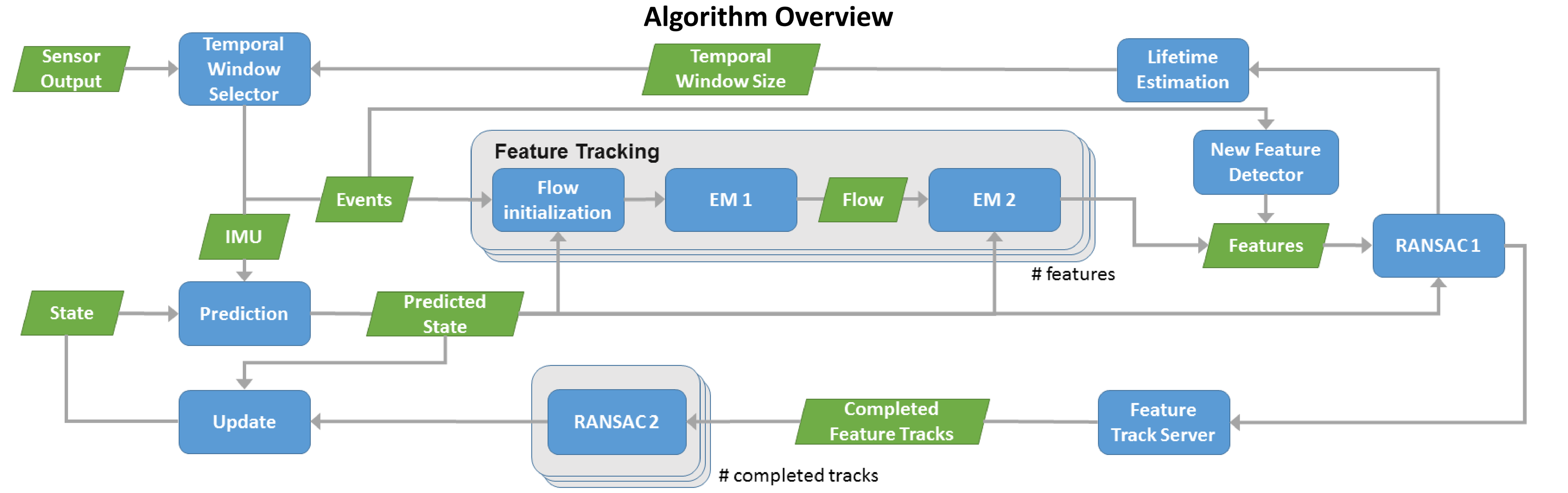
Prior work
Event-based visual odometry given a prior map [1], combined with grayscale images [2], event-based SLAM methods [3], [4].

Contributions
A novel event association scheme resulting in robust feature tracks by employing two EM-steps and variable temporal frames depending on flow and rotation estimates obtained from the odometry filter.
The first visual odometry system for event-based cameras that makes use of inertial information.

References

[1] A. Censi et al. Low-latency event-based visual odometry. *ICRA 2014.*
[2] B. Kueng, et al. Low-latency visual odometry using event-based feature tracks. *IROS 2016.*
[3] H. Kim et al. Real-time 3D reconstruction and 6-dof tracking with an event camera. *ECCV 2016.*
[4] H. Rebecq et al. EVO: A geometric approach to even-based 6-dof parallel tracking and mapping in real time. *RAL 2017.*
[5] A. Mourikis et al. A multi-state constraint Kalman filter for vision-aided inertial navigation. *ICRA 2007.*
[6] C. Troiani et al. 2-point-based outlier rejection for camera-imu systems with applications to micro aerial vehicles. *ICRA 2014.*

Algorithm Overview



Feature Tracking

EM1: Flow Estimation

For each spatiotemporal window, propagate the events to a common time using the current estimate of the flow, and then jointly estimate the **soft** data association, r_{kj}, between the propagated events from the current and last windows, and the optical flow, v. The optical flow is used to **estimate the future position** of the feature.

$$\min_{r,v} \sum_{k=1}^{n}\sum_{j=1}^{n} r_{kj}||(x_k - \bar{t}_k v) - l_j||^2$$

EM2: Template Alignment

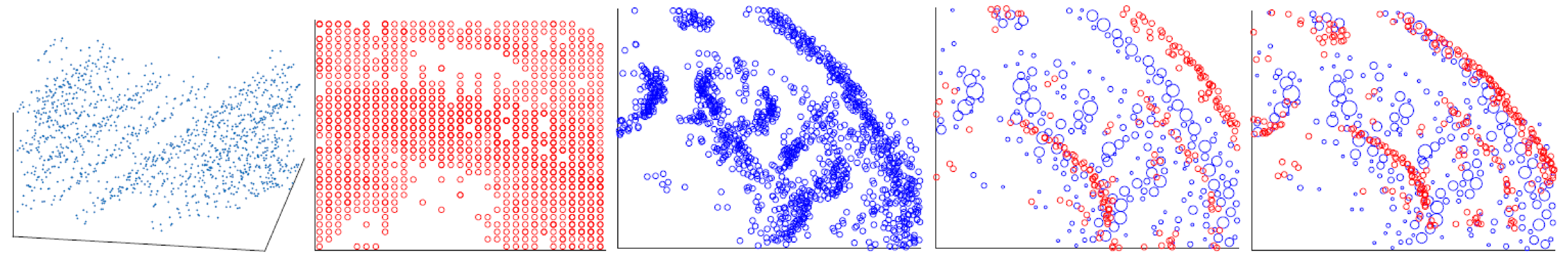
Given a set of propagated events, warp them according to the rotation between the current state rotation and the rotation of the first window, $\bar{l}_j^{i*}$, and then jointly estimate the **soft** data association, r_{kj}, scaling, σ, and translation, b, between the current events and those from the first window, y_k^i. This **corrects for drift** and **detects failed tracks**.

$$\min_{r,\sigma,b} \sum_{k=1}^{n}\sum_{j=1}^{m} r_{kj}||\sigma y_k^i - b - \bar{l}_j^{i*}||^2$$

Temporal Window Size: Lifetime Estimation

The size of the next temporal window, τ, is set to the time taken for a point to move 3 pixels. This constrains the amount of motion in the window to be small, to satisfy the **constant optical flow assumption** within each window.

$$\tau = \frac{3}{\|v\|_2}$$



Spatiotemporal window. | Propagated events, zero flow. | Propagated events, correct flow. | Before template alignment. | After template alignment.

Visual Inertial Odometry

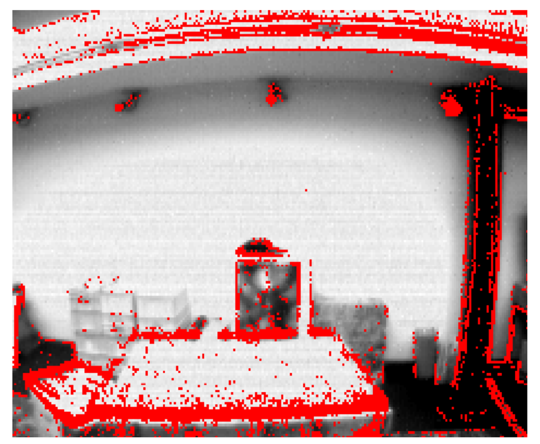
Filter
The feature positions and IMU measurements are fused using the MSCKF [5] framework, which optimizes the sensor **without optimizing over the 3D feature positions**. This method provides a fast state update over a window of feature observations.

Outlier Rejection
After each feature tracking update, two-point RANSAC [6] is used to remove outliers and failed tracks. Given rotation from the IMU, the largest set of point pairs that correspond to the same translation between the current and last frame is kept as inliers.

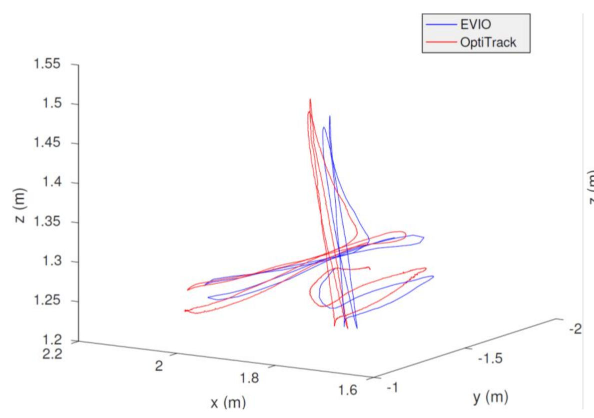
In addition, before each feature track is marginalized, the largest set of observations that project to the same 3D point is kept as inliers.

Results

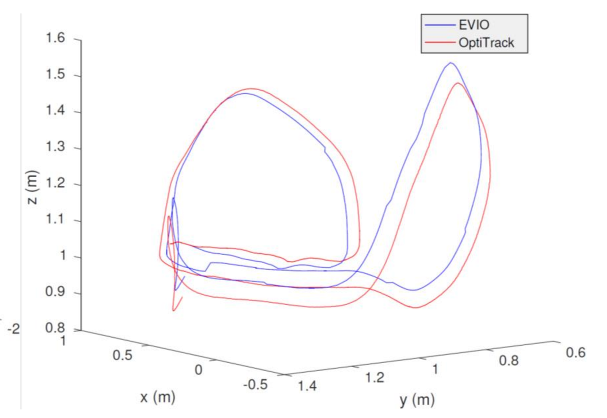
Challenging low texture scene.

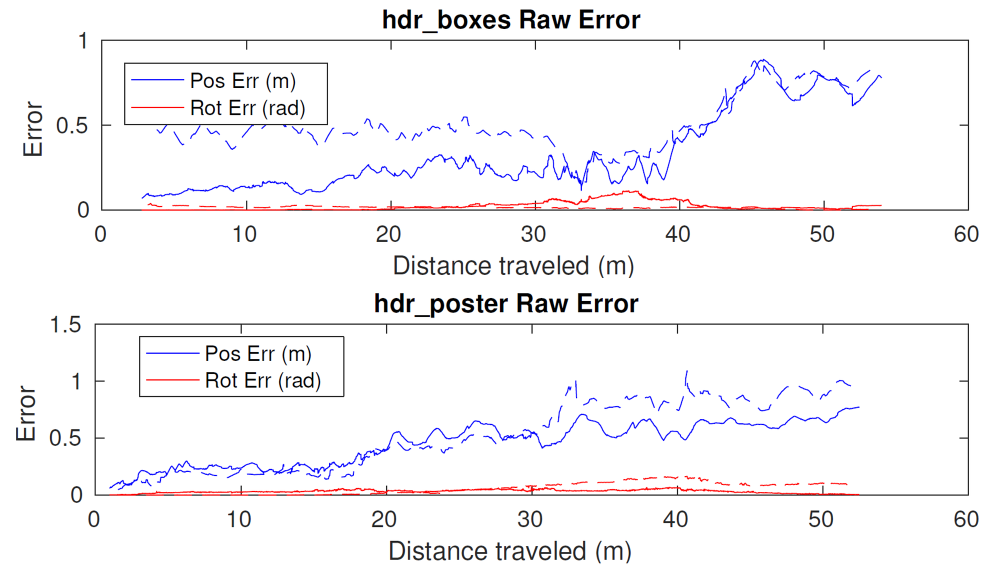


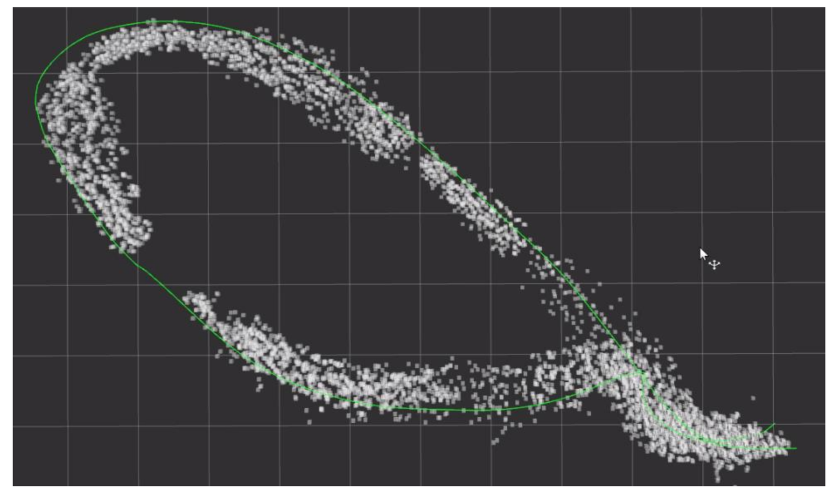
shapes_translation



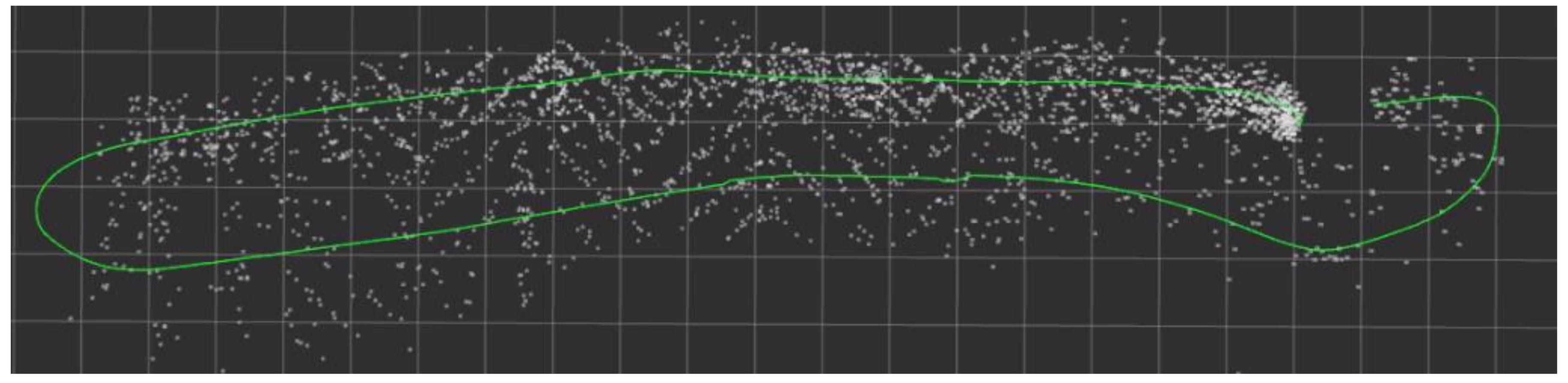
hdr_boxes







DAVIS-346B mounted downwards on a quadrotor flying over grass.



DAVIS-346B mounted downwards on a quadrotor flying over tiled surface.