System Overview

Kinematic and Measurement Models for Monocular Visual-Inertial Fusion

- Measurements of linear acceleration and rotational velocity from an inertial measurement unit (IMU) are used as inputs to a kinematic-statistical model:

\[
\begin{align*}
\dot{T} &= v \\
\dot{R} &= R(\omega_{\text{imu}} - \omega_b) + n_R \\
\dot{\omega}_b &= \omega_b \\
\dot{\alpha}_b &= \alpha_b.
\end{align*}
\]

- Acceleration and rotational velocity are known, but biased, inputs with gravity as a known parameter. Model driven by Gaussian process noise.

- Feature tracks from a monocular camera rigidly attached to the IMU provide measurements of relative pose and world structure:

\[
X_t^i = X_t^i(g_t, y_t, Z_t) = g_t, y_t, Z_t
\]

\[
\begin{align*}
\dot{g}_i^b &= 0, & i = 1, \ldots, N(j) \\
\dot{Z}_i^b &= 0, & j = 1, \ldots, K(t).
\end{align*}
\]

- Additional complexities:
  1) camera-IMU alignment: \( \dot{g}_i^b = 0 \)
  2) temporal misalignment: \( \dot{\omega}_b = 0 \)

- World scale recovered by fusion with accelerometers.

- Fully realtime implementation using an Extended Kalman Filter.

- Further details to appear in [1] and [2].

Key Novelty in Measurement Processing: Outlier Rejection

- Flow-based tracking provides long tracks for accurate depth estimation, but are prone to drift
- Track lengths limited only by visibility
- Unpredictable drift can cause a feature to become an outlier at any time, e.g. occlusion boundaries, specular highlights, non-rigid scene

Goal:
- Ensure features are not outliers at all times before use
- Perform updates at the current time using only features represented in the state

Key Modeling Assumption:
- Residual process should be temporally uncorrelated, not just instantaneously small
- Typical robust filtering strategies rely on instantaneous tests of the residual
- The residual process being temporally white is a stronger condition
- Perform outlier testing on the residual process rather than on individual samples instantaneously

Map Building from Filter Output

- Estimated 3D coordinates of features are stored in their initial coordinate frames and saved in a relative pose & co-visibility graph.
- Lines denote co-visibility edges and store relative transformations, and each 3D point in the map is associated with a node of the graph.
- Detected loop closures add an edge to the graph, enabling efficient local map correction.

Results

Crowded poster hall (many moving people)

- Instantaneous Testing - drift: 4.33m (~2.72%), wd: 6.76
- Process Testing - drift: 0.18m (~0.11%)
- Origin of ~160m path

Indoor building environment (many specular highlights)

- Instantaneous Testing - drift: 2.65m (~0.96%), wd: 6.02
- Process Testing - drift: 0.47m (~0.37%)
- Origin of ~265m path

Outdoor natural environment (many occluding boundaries due to foliage)

- Instantaneous Testing - drift: 3.44m (~1.91%), wd: 12.24
- Process Testing - drift: 0.67m (~0.37%)
- Origin of ~190m path
