Robust Scale Estimation in Real-Time Monocular SFM for Autonomous Driving
Shiyu Song and Manmohan Chandraker, NEC Laboratories America

Highlights

- Highly accurate monocular SFM with comparable accuracy to stereo.
- Ground plane estimation through different cues, such as dense stereo, 3D points and object detection.
- A novel data-driven framework that adaptively combines multiple cues based on per-frame observation covariances estimated by rigorously trained models.
- Scale drift is corrected by the optimal estimated by rigorously trained models.

Approach: Adaptive Cue Fusion

1. **Cue 1: Dense Stereo**
   - Use height from ground to resolve scale $s$

2. **Cue 2: 3D Points**
   - Object height estimation $h_{p,k} = h_{p,k-1} + \delta h_{p,k}$

3. **Cue 3: Object**
   - Histgram of data.
   - Learned linear model to relate the observation variance to the underlying variable, the Gaussian variance $\Sigma_{x_{p,k}}$.

Method

- We can compute both 3D points and camera motion, up to unknown scale factor.

Object Localization

- Comparison of 3D object localization errors for calibrated ground, stereo cue only, fixed covariance fusion and adaptive covariance fusion of stereo and detection cues.

Results

- Ground Height
  - Fig. Height error relative to ground truth. The effectiveness of our data fusion is shown by less spikiness in the filter output and a fair lower error.

- Monocular SFM
  - (a) Height error
  - (b) Success rate

- Object Localization
  - (a) Close objects (GT)
  - (b) Distant objects (GT)
  - (c) Close objects (DP)
  - (d) Distant objects (DP)