Robust Scale Estimation in Real-Time Monocular SFM for Autonomous Driving

Shiyu Song and Manmohan Chandraker, NEC Laboratories America

NEC Laboratories America
Relentless passion for innovation

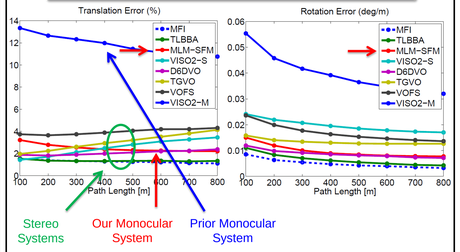
CVPR 2014

Highlights

Summary

- 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 fused estimates, which yields high accuracy and robustness.

Top Monocular Performance



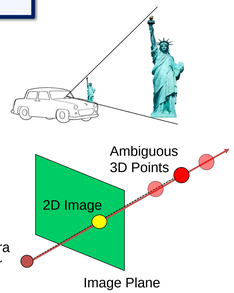
KITTI Visual Odometry Benchmark

Motivation

	LIDAR	Stereo Camera	Monocular Camera
Cost	$70,000	$1,000	$100
Maintenance	Hard	Hard	Easy

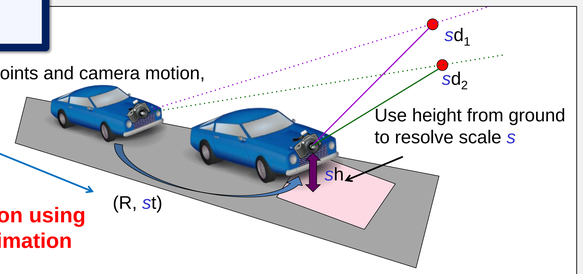
Comparisons of LIDAR, stereo and monocular System

Challenge



Method

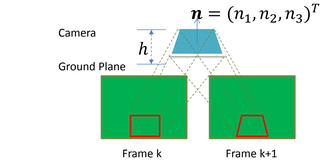
We can compute both 3D points and camera motion, *up to unknown scale factor*.



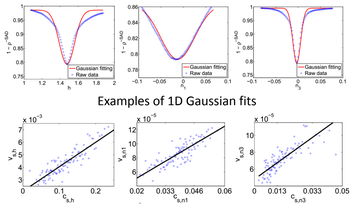
Scale Drift Correction using Ground Plane Estimation

Approach: Adaptive Cue Fusion

Cue 1: Dense Stereo

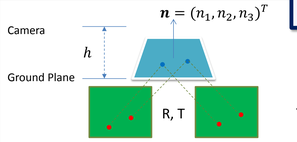


$$\min_{\boldsymbol{n},h}(1-\eta^{\mathrm{SAD}})$$

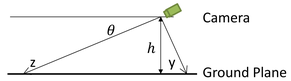


Learned linear model to relate the observation variance to the underlying variable, the Gaussian variance

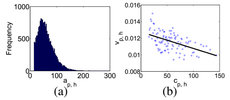
Cue 2: 3D Points



Feature matching and triangulation

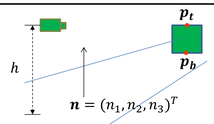


$$\mathrm{q} = \max_i\{\sum_{j\neq i}\exp(-\mu\Delta h_{ij}^2)\}$$



(a) Histogram of data. (b) Learned linear model to relate the observation variance to the underlying variable, q.

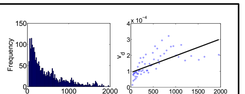
Cue 3: Object



$$\min_{n_3}(h_b-\bar{h}_b)^2$$

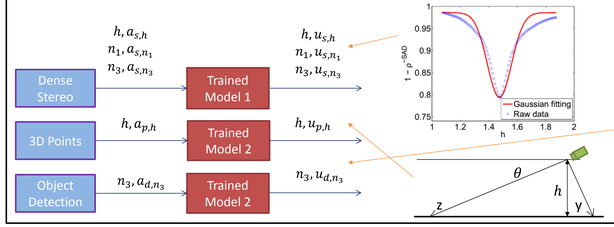
$$\min_{A_m,\boldsymbol{\mu}_m,\Sigma_m}\sum_{n=1}^{N}(\sum_{m=1}^{M}A_m e^{-\frac{1}{2}(\boldsymbol{b}_n-\boldsymbol{\mu}_m)\Sigma_m^{-1}(\boldsymbol{b}_n-\boldsymbol{\mu}_m)}-\alpha_n)^2$$

$$\Sigma_m=\begin{bmatrix}\sigma_x^2 & * & * & *\\ * & \sigma_y^2 & * & *\\ * & * & \sigma_{w_b}^2 & *\\ * & * & * & \sigma_{h_b}^2\end{bmatrix},\quad a_d^k=\frac{\sigma_y\sigma_{h_b}}{\sigma_y+\sigma_{h_b}}$$



(a) Histogram of data. (b) Learned linear model to relate the observation variance to the underlying variable, a_d^k.

Adaptive Fusion



Results

Ground Height

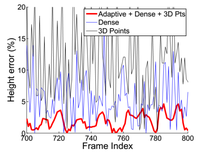
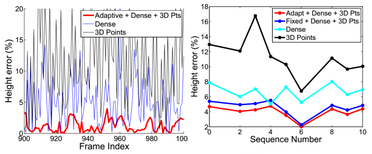


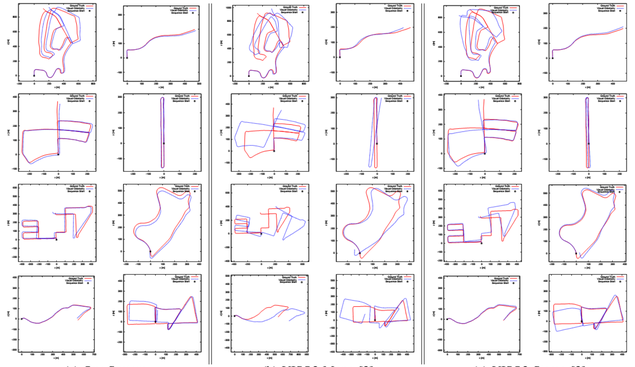
Fig. Height error relative to ground truth. The effectiveness of our data fusion is shown by less spikiness in the filter output and a far lower error.



(a) Height error (b) Success rate

Fig. (a) Average error in ground plane estimation. (b) Percent number of frames where height error is lower than 7%.

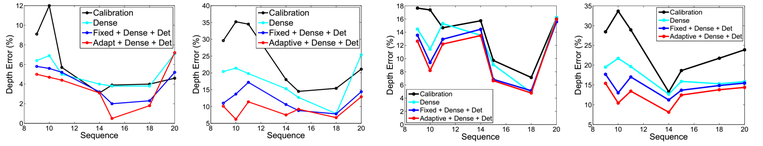
Monocular SFM



(a) Our System (b) VISO2-Mono [2] (c) VISO2-Stereo [2]

Results in the KITTI Visual Odometry Dataset

Object Localization



(a) Close objects (GT) (b) Distant objects (GT) (c) Close objects (DP) (d) Distant objects (DP)

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