PROBE: Predictive Robust Estimation for Visual-Inertial Navigation

Valentin Peretroukhin, Lee Clement, Matthew Giamou, and Jonathan Kelly

IROS 2015, Hamburg, Germany

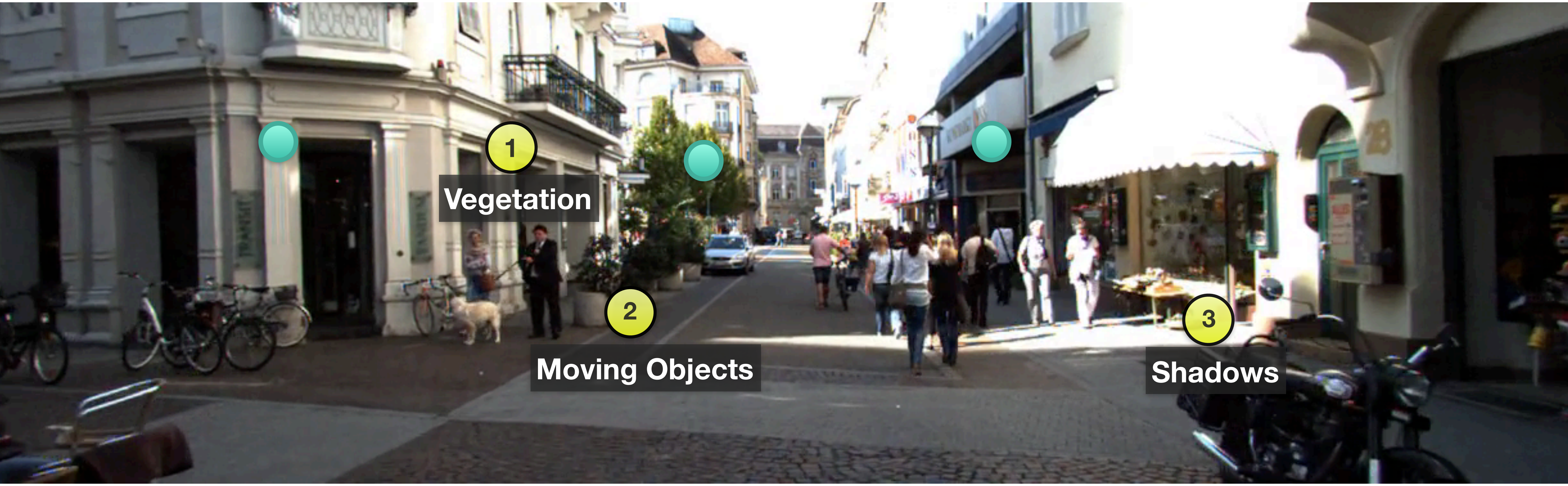
Visual Navigation II

Institute for Aerospace Studies
UNIVERSITY OF TORONTO

STARS
LABORATORY

Visual Inertial Navigation

Question: Are all visual features created equal?

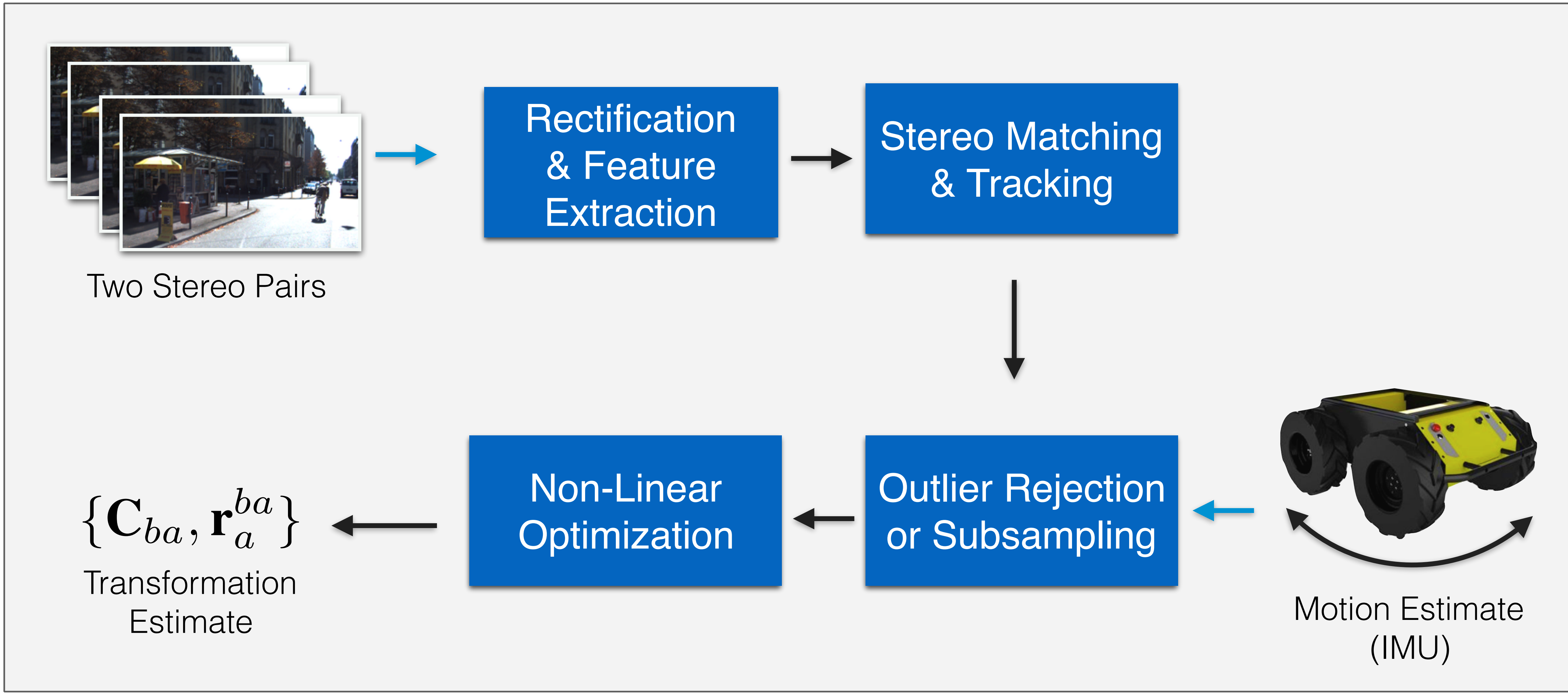


KITTI Dataset. Sequence 2011_09_29_drive_0071.

Hypothesis: Inlier point features are not all equally informative.

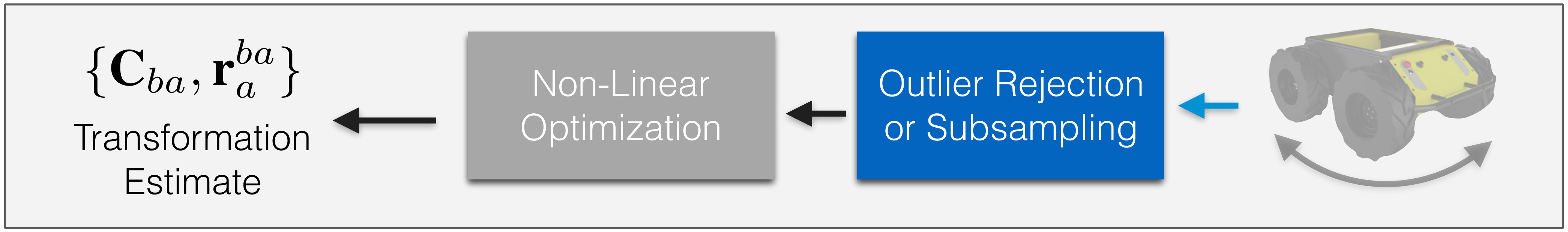
Visual Inertial Navigation

Frame-to-frame visual-inertial navigation with sparse visual features.



Feature Selection

How do we deal with less informative features?



RANSAC (and variants)

- Front-end, **binary** technique
- Binary outlier rejection based on Random Sample Consensus
- Fischler (1981)

M-Estimation

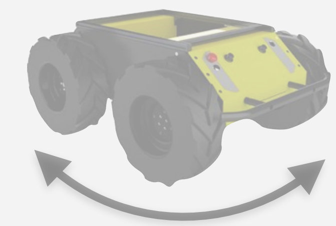
- Back-end, **reactive** technique
- Robust cost functions reduce influence of outliers
- See Latif et al. (2013)

Feature Selection

How do we deal with less informative features?

$\{\mathbf{C}_{ba}, \mathbf{r}_a^{ba}\}$

Transformation Estimate ← Non-Linear Optimization ← Outlier Rejection or Subsampling ←



C (and variants)

, **binary** technique

tlier rejection based

m Sample

us

981)

PROBE

- Hybrid, **predictive** technique
- Inflate image covariance based on a learned model from visual and inertial data
- Vega-Brown et al. (2013), Peretroukhin et al. (2015)

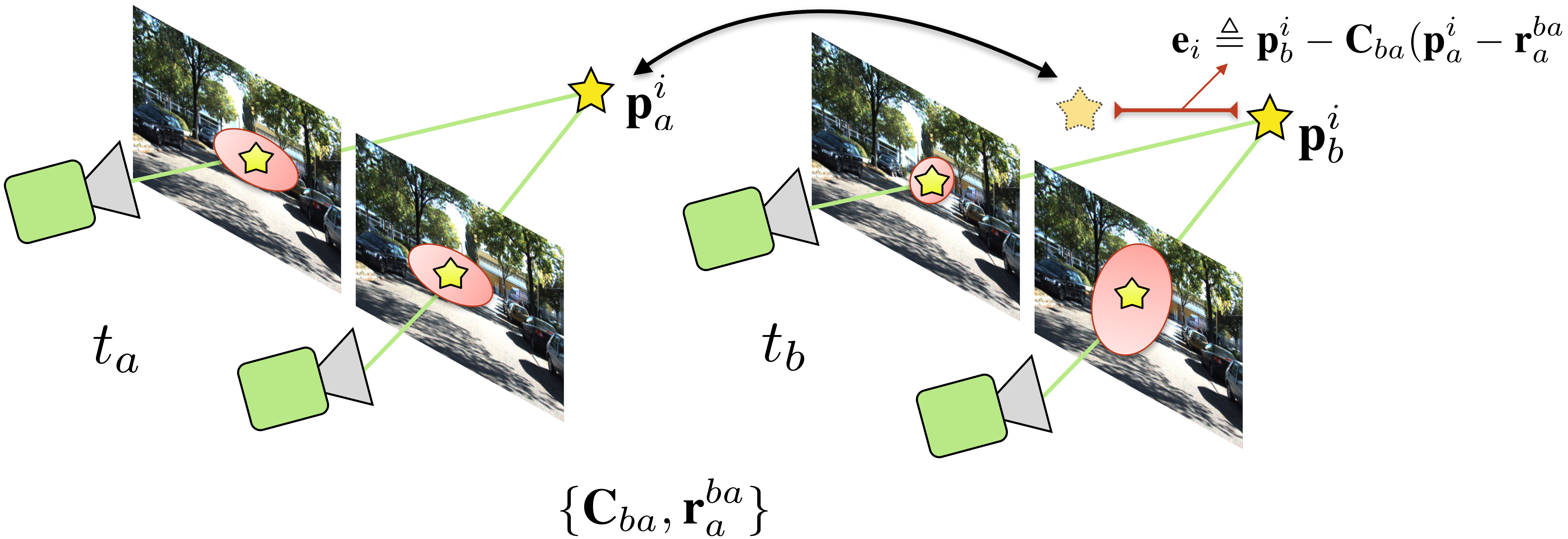
M-Estimat

- Back-end, **reactiv**
- Robust cost functi influence of outliers
- See Latif et al. (201

PROBE

Varying image covariances.

Key Idea: $\mathbf{R} = \mathbf{R}(\phi)$



- We perform non-linear optimization on the weighted sum of 3D errors:

$$\mathcal{L} = \frac{1}{2}\sum_{i=1}^{N} \mathbf{e}_i^T \mathbf{\Gamma}^i \mathbf{e}_i$$

- Each weight is found by propagating image space **covariances** through a stereo camera model:

$$\mathbf{\Gamma}^i = f(\mathbf{R}) = \left(\mathbf{G}_b^i \mathbf{R}_b^i \mathbf{G}_b^{i^T} + \mathbf{C}_{ba} \mathbf{G}_a^i \mathbf{R}_a^{i^T} \mathbf{G}_a^{i^T} \mathbf{C}_{ba}^T\right)^{-1}$$

PROBE: Predicting Feature Quality

Idea: Scale image covariance as a function of a prediction space.

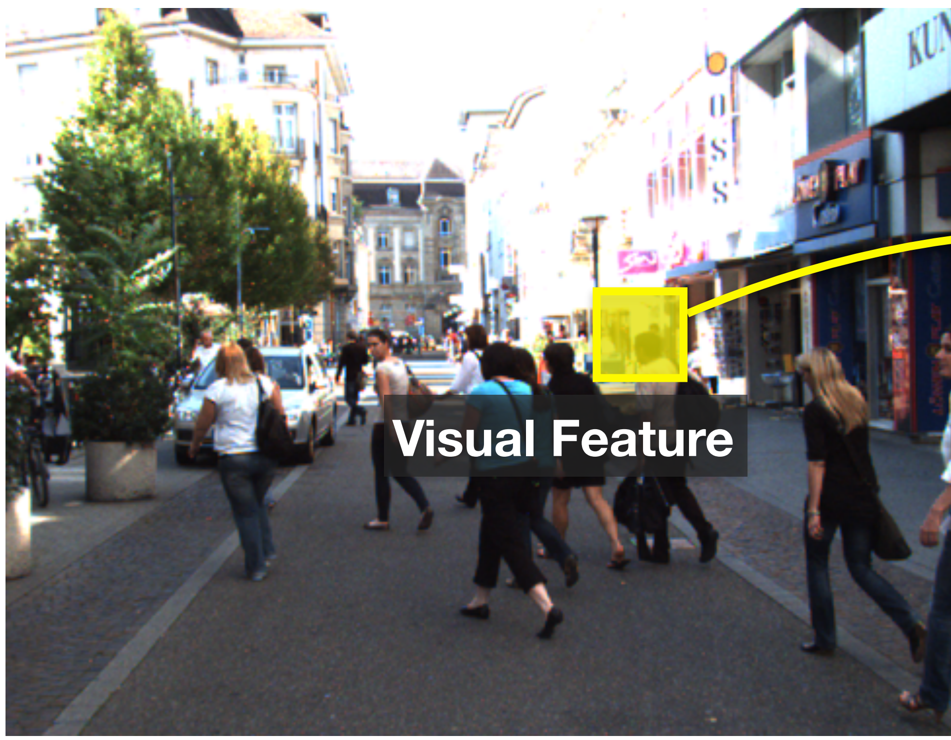
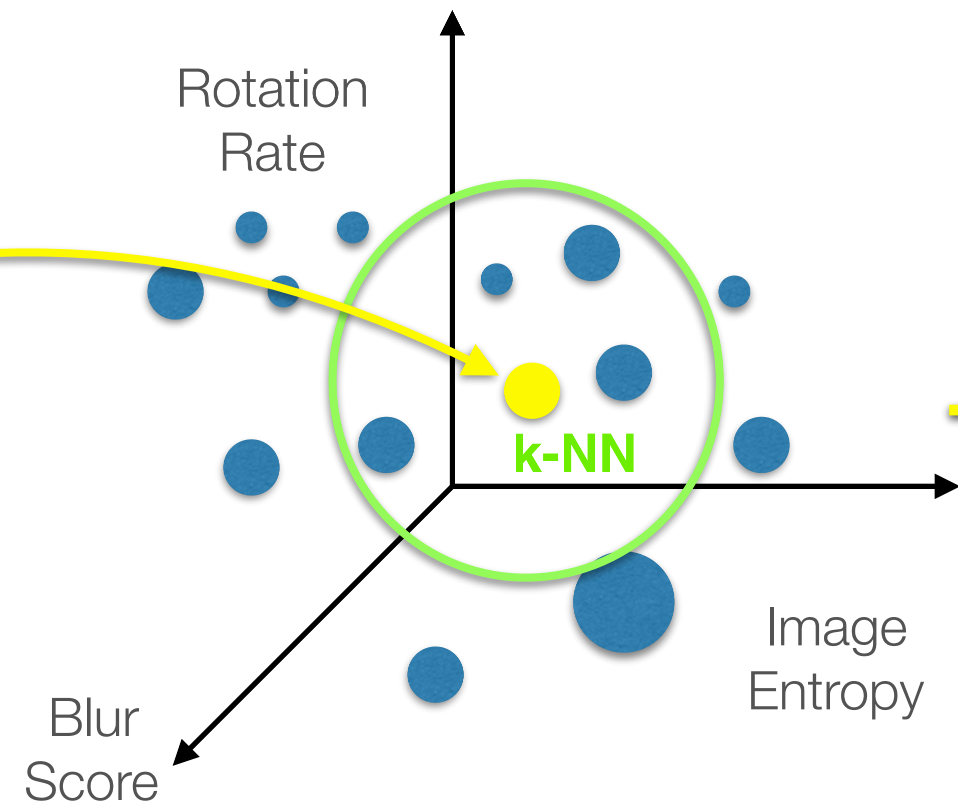


Image Space
(with Inertial Cues)



Prediction Space (ϕ)

W. Vega-Brown et al. “CELLO: A fast algorithm for Covariance Estimation,” ICRA 2013.

$$\beta$$

$$\left(\frac{1}{\overline{\alpha} K} \sum_{k=1}^{K} \alpha_k \right)^{\gamma}$$

Uncertainty
Factor

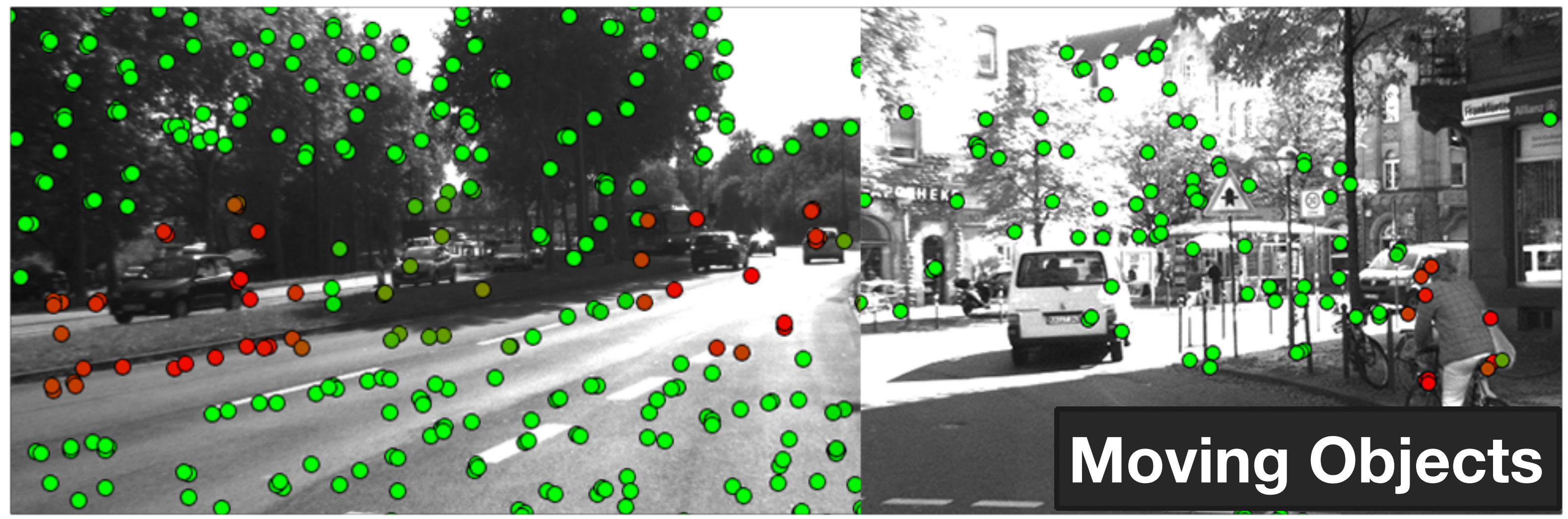
Scale Feature Covariance $\mathbf{R}(\phi) = \beta(\phi)\mathbf{R}_{\text{fixed}}$

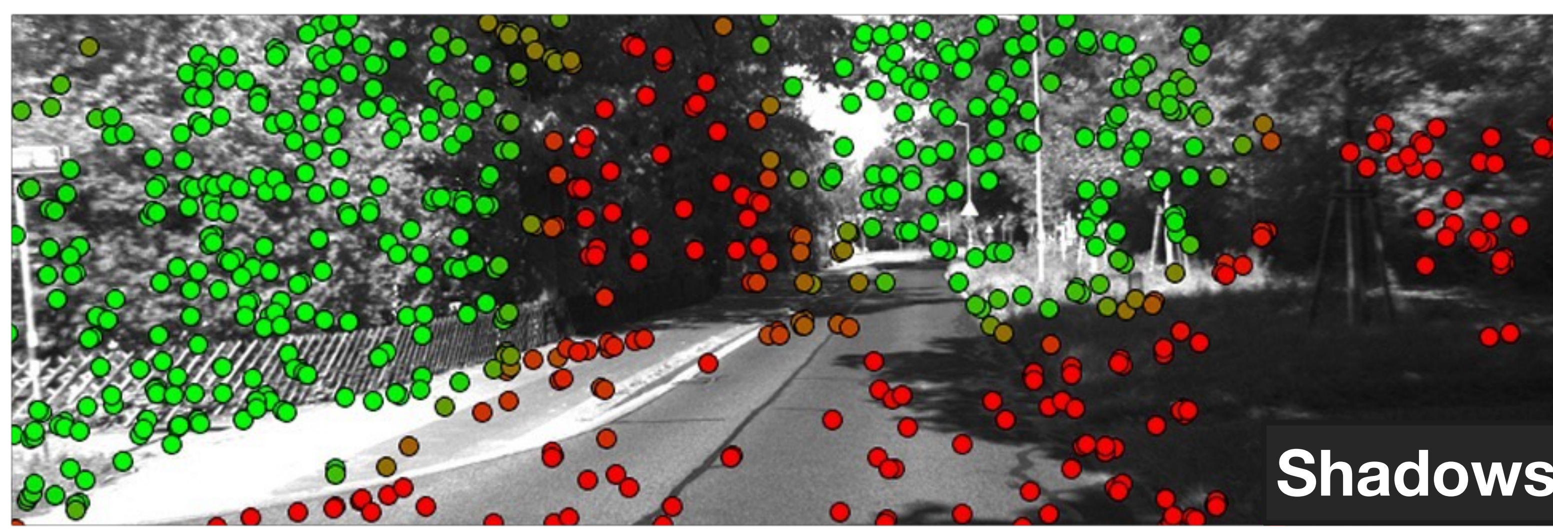
PROBE: Predictor Selection

Goal: Prediction space should identify moving objects, shadows, motion blur.

We use a combination of visual and inertial predictors.

1. Optical Flow Variance
2. Inertial Magnitudes
3. High and Low Frequency Content
4. Motion Blur Score [1]
5. Image Entropy

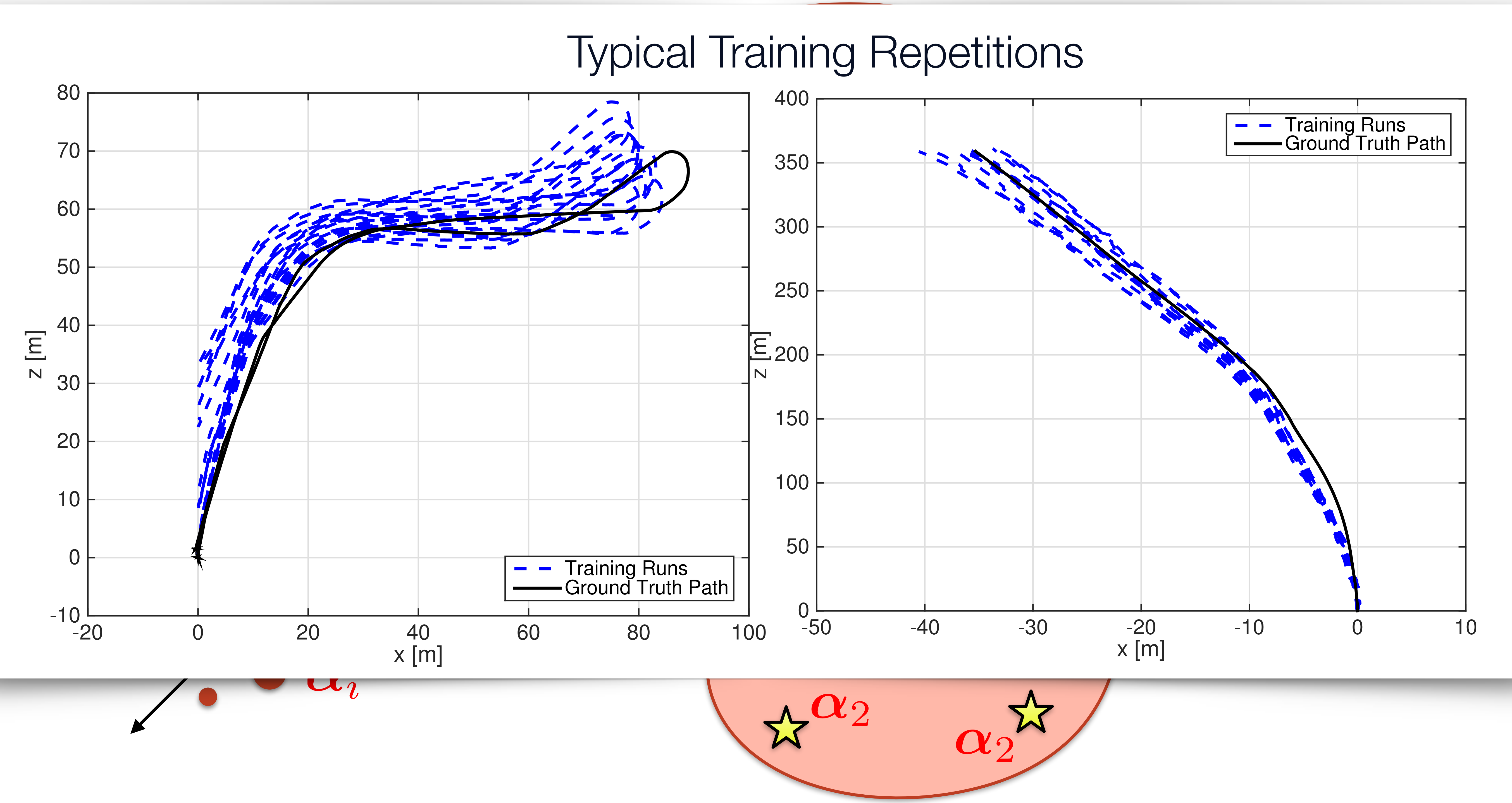




[1] F. Crete et al, "The blur effect: perception and estimation with a new no-reference perceptual blur metric," Electronic Imaging 2007, vol. 6492,, Feb. 2007.

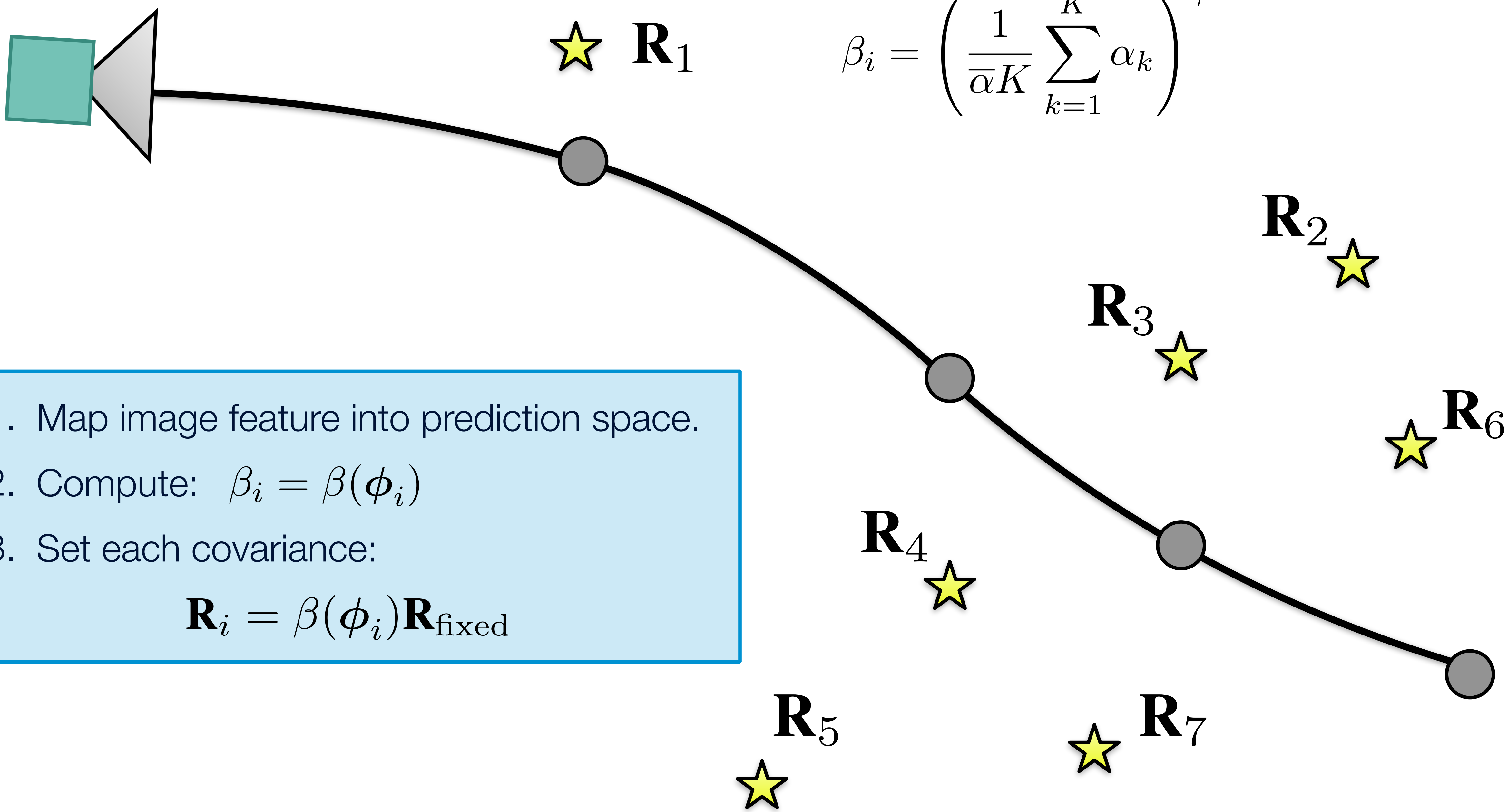
PROBE: Predictors

PROBE: Training Procedure



PROBE: Implementation

$$\beta_i = \left(\frac{1}{\overline{\alpha} K} \sum_{k=1}^{K} \alpha_k \right)^{\gamma}$$



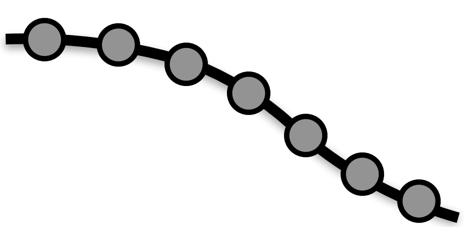
1. Map image feature into prediction space.
2. Compute: $\beta_i = \beta(\boldsymbol{\phi}_i)$
3. Set each covariance:

$$\mathbf{R}_i = \beta(\boldsymbol{\phi}_i)\mathbf{R}_{\text{fixed}}$$

PROBE: Datasets

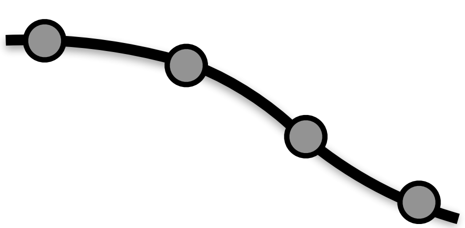
KITTI Dataset
> 4 km in 3 environments

<u>Frequent Ground Truth RMSE</u>



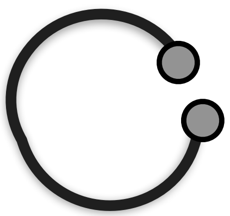
UTIAS Outdoor
600 m in winter environment

<u>Sparse Ground Truth RMSE</u>



UTIAS Indoor
60 m in indoor lab

<u>Loop Closure Error</u>



PROBE: Results

Nominal RANSAC

$$k = \frac{\log(1-p)}{\log(1-w^n)}$$ (w=0.5, p=0.99)

vs.

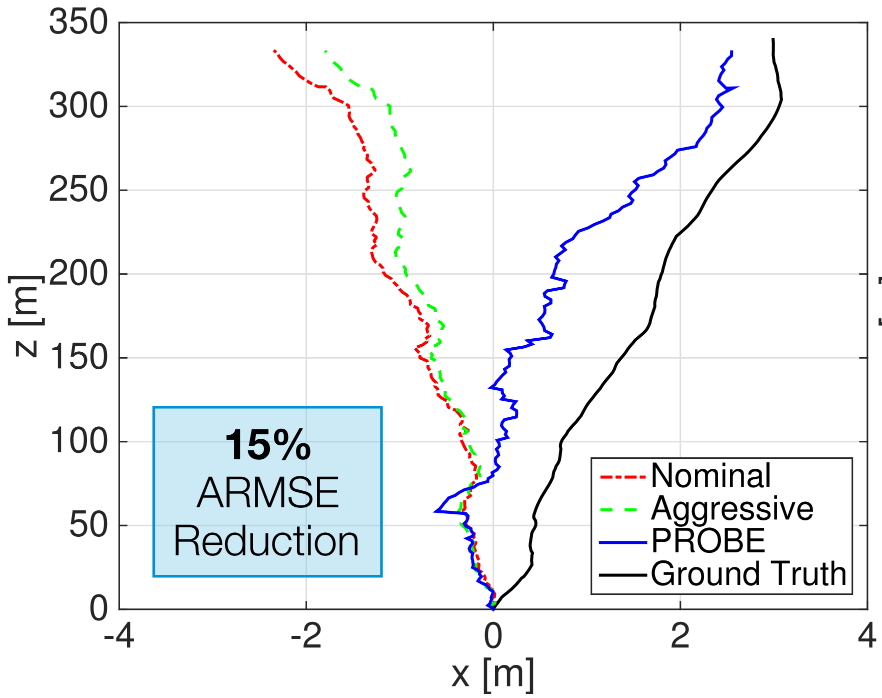
Aggressive RANSAC

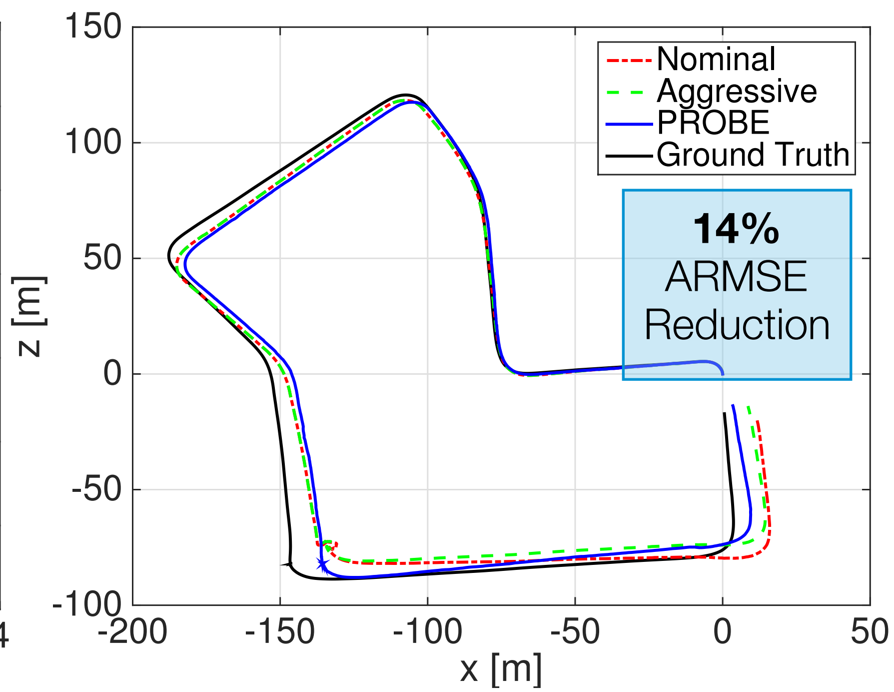
$$k = \frac{\log(1-p)}{\log(1-w^n)}$$ (w=0.5, p=0.9999)

vs.

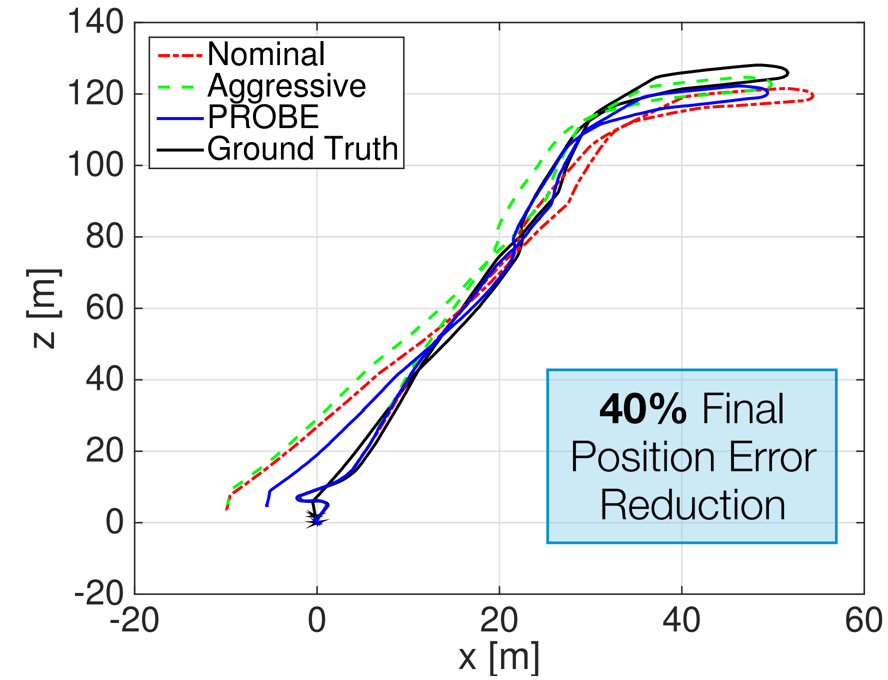
PROBE

KITTI Dataset

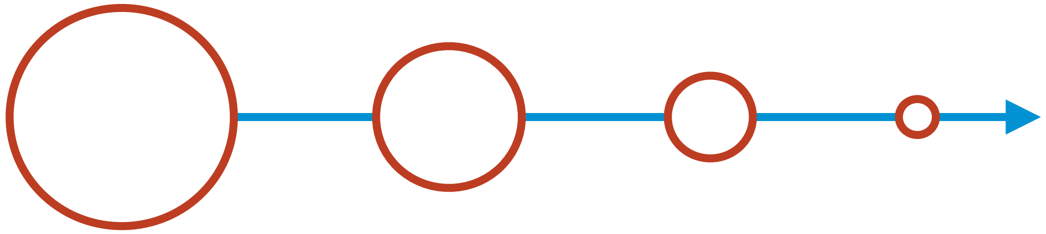




UTIAS Outdoor

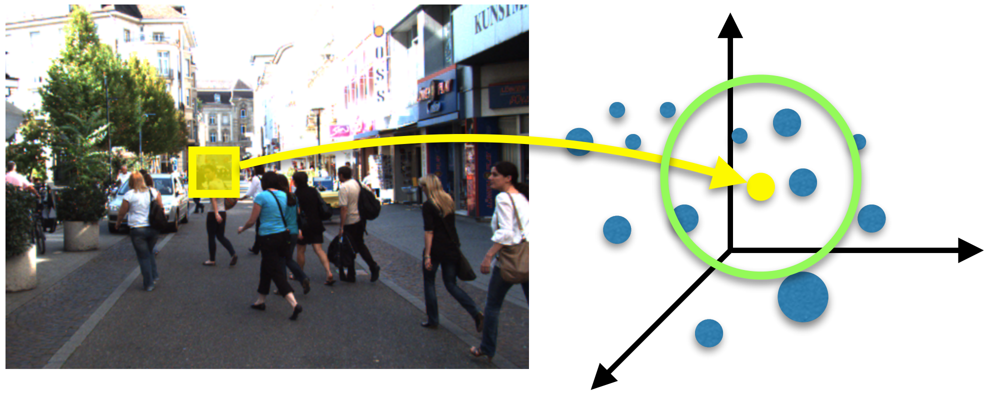


PROBE: Putting it all together



High Uncertainty   Informative

Summary



- ✓ **PROBE** predicts the informativeness of visual features, scaling image covariances
- ✓ Shown to improve VINS on KITTI & experimental data relative to RANSAC
- ✓ Training done with sparse ground truth

Future Work

- New version (PROBE-GK, submitted to ICRA 2016)
  - ☑ Derivation from first principles: Full covariance learning using Bayesian framework with a predictive estimator
  - ☑ Removed necessity for ground truth.
  - ☑ Comparison to M-estimation.
- Further questions:
  - Is **online learning** possible?
  - How can we select **informative predictors**?

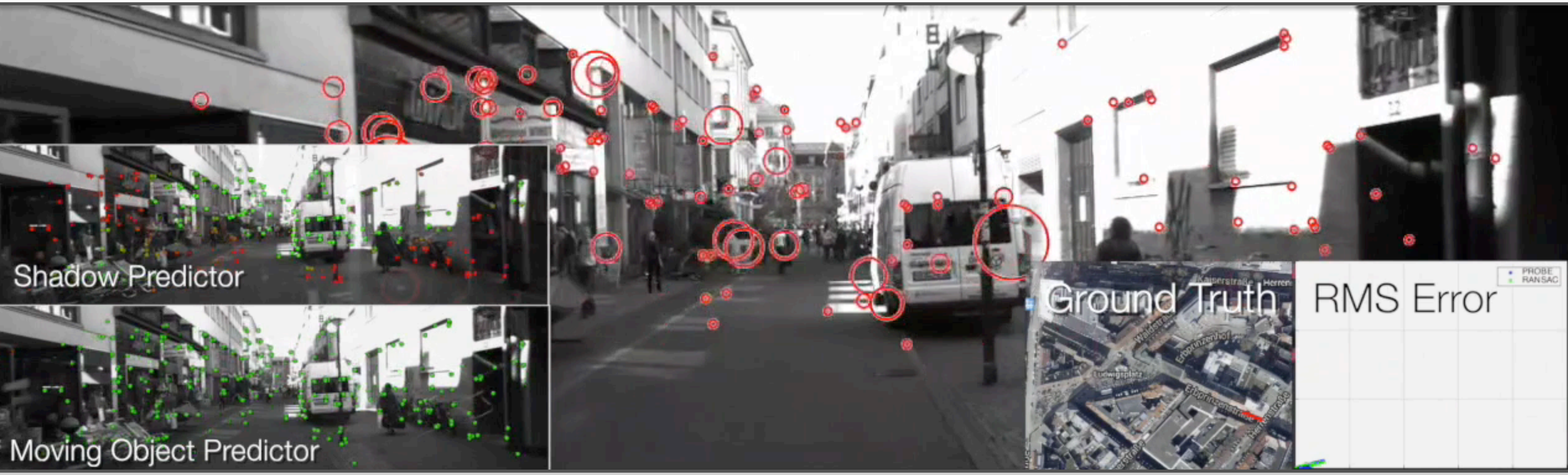
Thanks! Questions?

Email: v.peretroukhin@utoronto.ca
Web: http://starslab.ca

Institute for Aerospace Studies
UNIVERSITY OF TORONTO

PROBE: Predictive Robust Estimation for Visual Inertial Navigation

PROBE: Predictive Robust Estimation for Visual-Inertial Navigation



Valentin Peretroukhin, Lee Clement, Matthew Giamou, and Jonathan Kelly
starslab.ca

Institute for Aerospace Studies
UNIVERSITY OF TORONTO

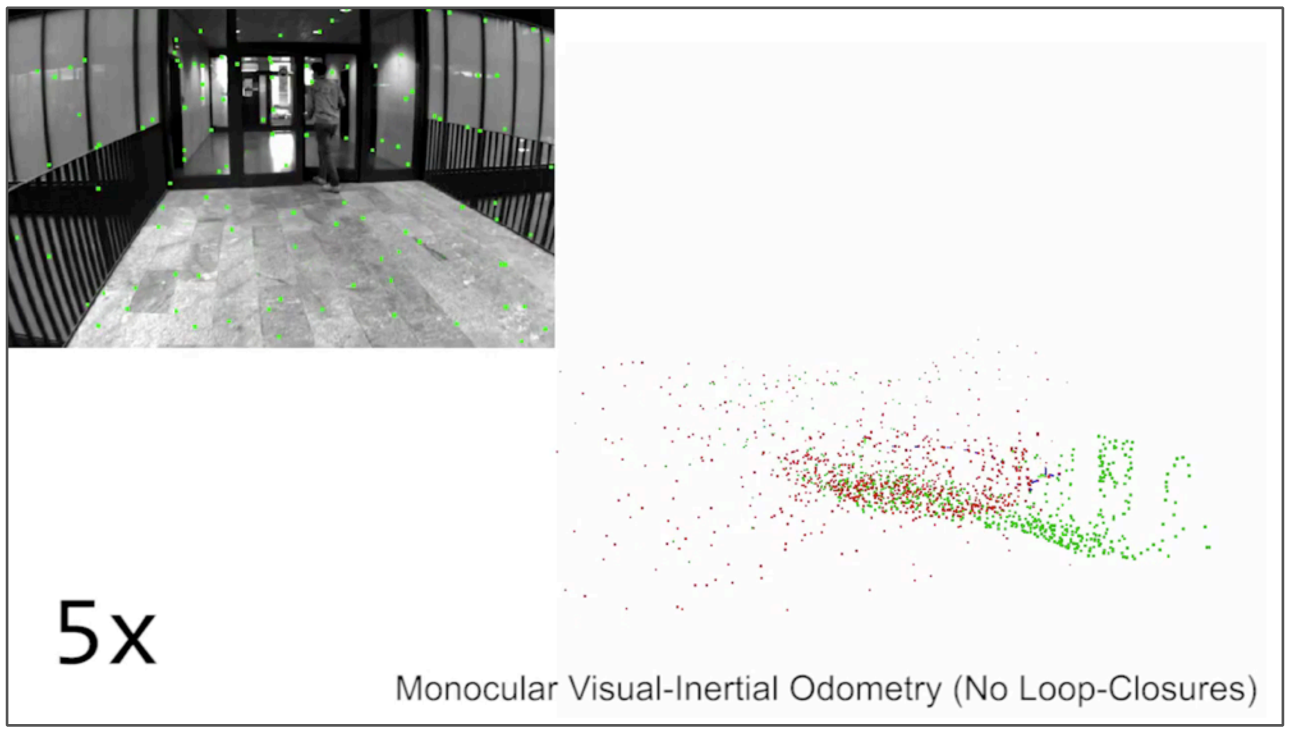
STARS LABORATORY

Additional Slides

Visual Inertial Navigation

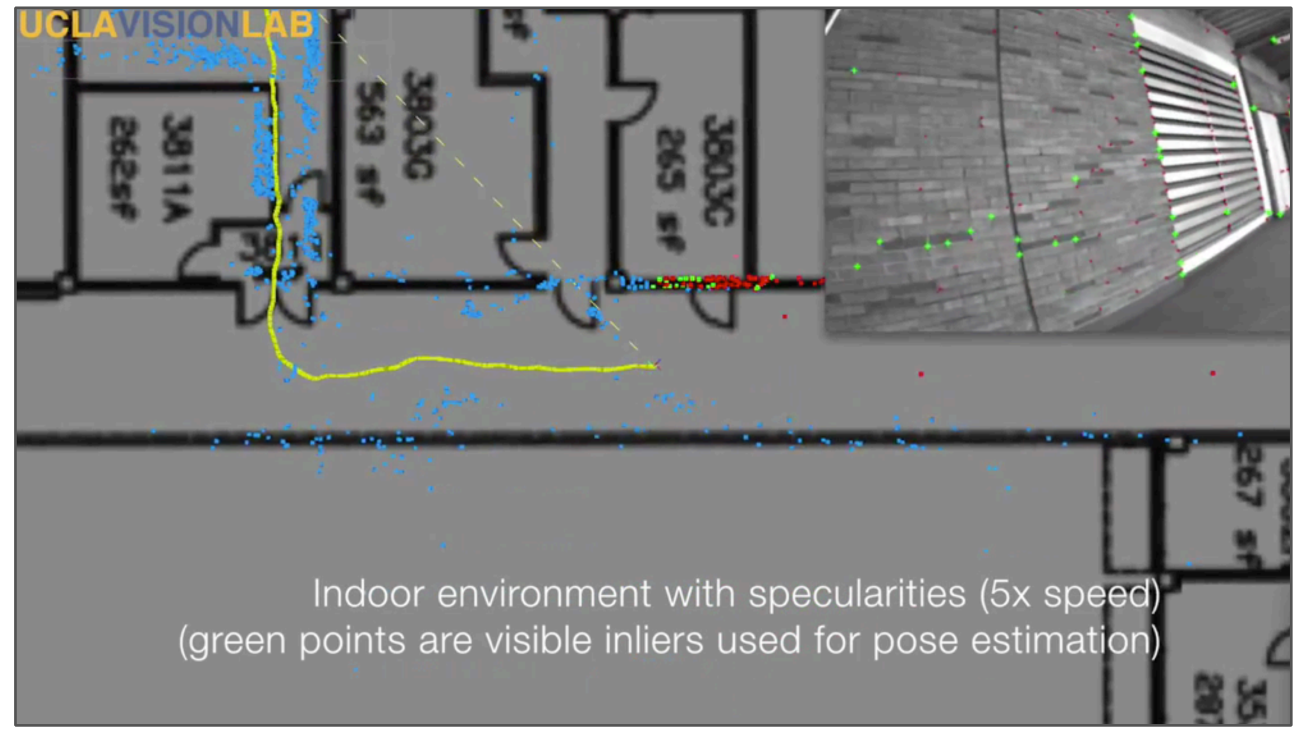
Modern Approaches in the Literature

Forster et al.



Christian Forster et al., "**IMU Pre-Integration on Manifold for Efficient Visual-Inertial Maximum-a-Posteriori Estimation**", Robotics: Science and Systems (RSS) 2015.

Tsotsos et al.

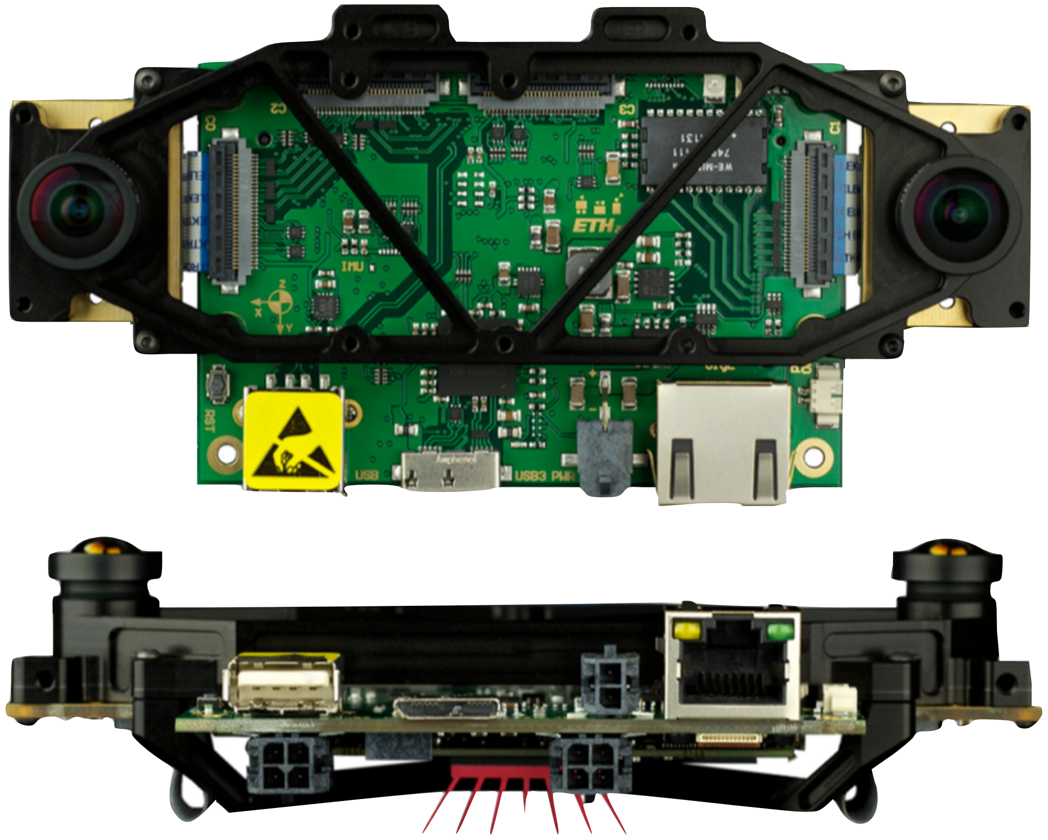


Konstantine Tsotsos et al., "**Robust Inference for Visual-Inertial Sensor Fusion**", International Conference for Robotics and Automation (ICRA) 2015.

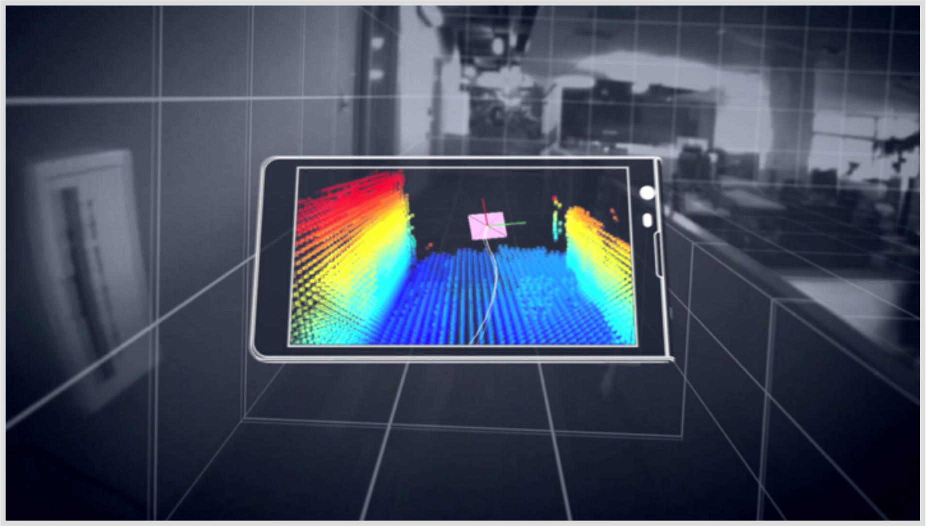
Introduction

VINS and Robotics

- Visual-Inertial Navigation Systems (VINS) use cameras and inertial measurement units to estimate motion.
- Sensors are complementary, relatively cheap, and light-weight.
- VINS can be applied to many robotics applications (ground, air, human).
- Difficulty: data is high resolution and high rate.



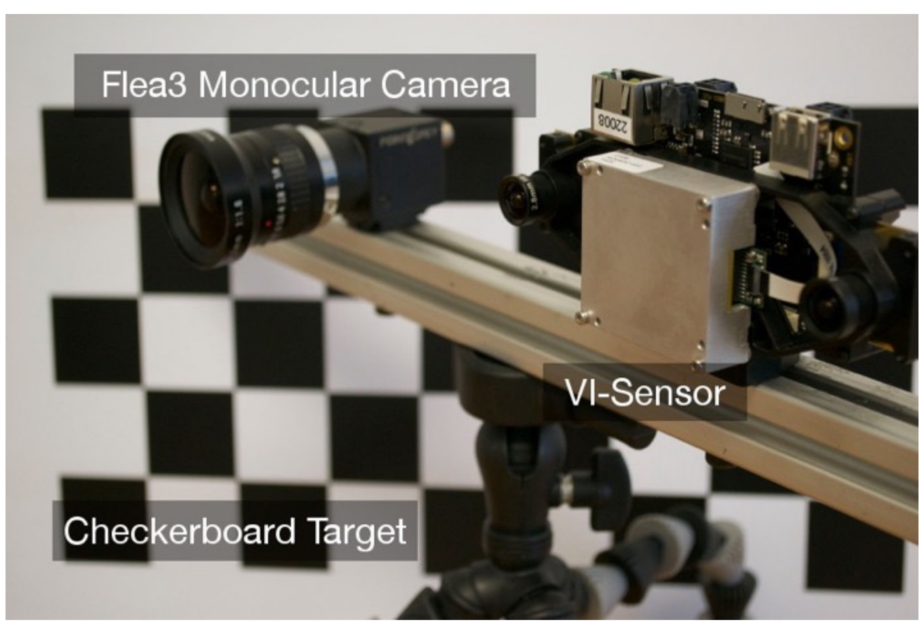
Skybotix VI Sensor



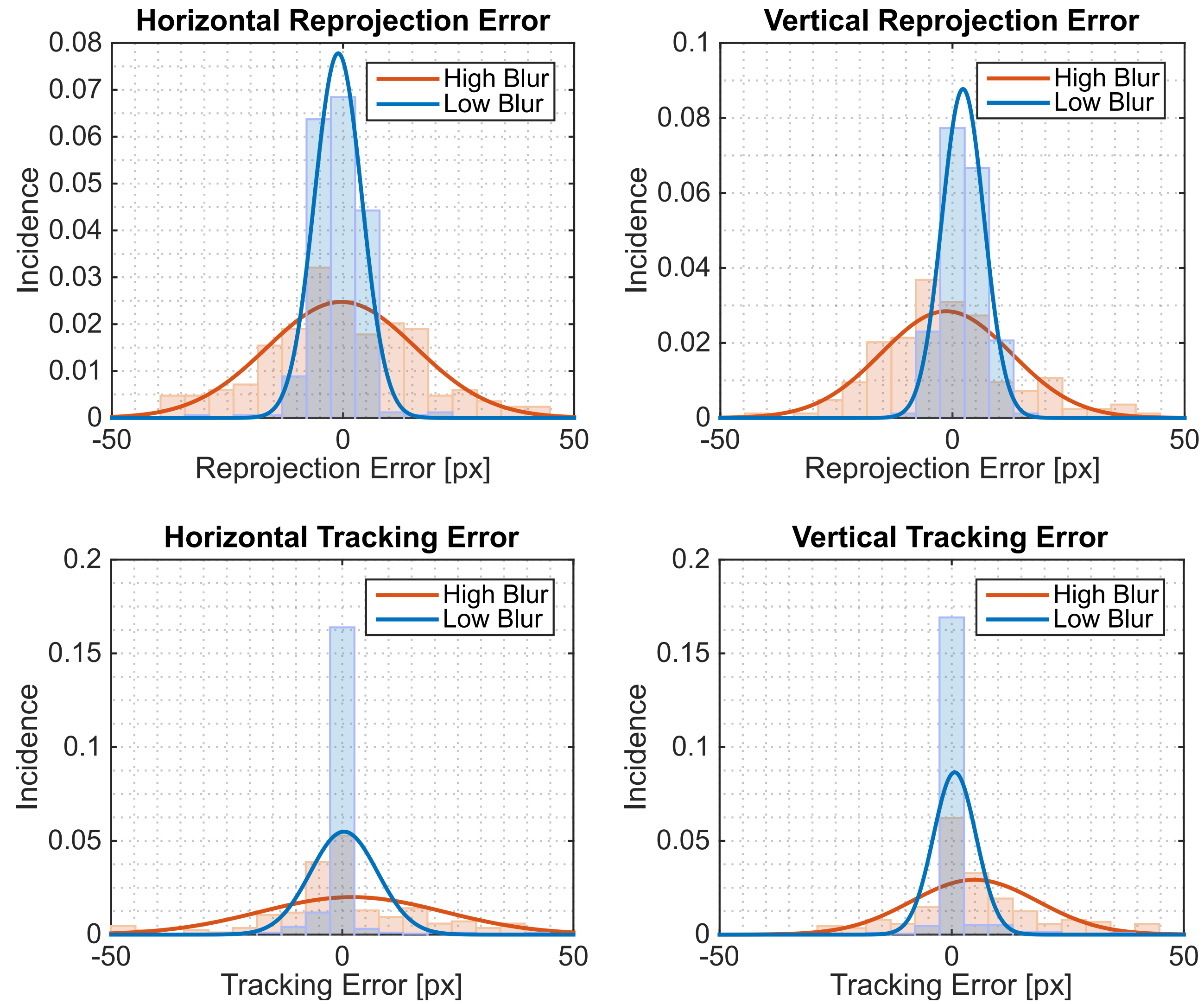
Google Project Tango

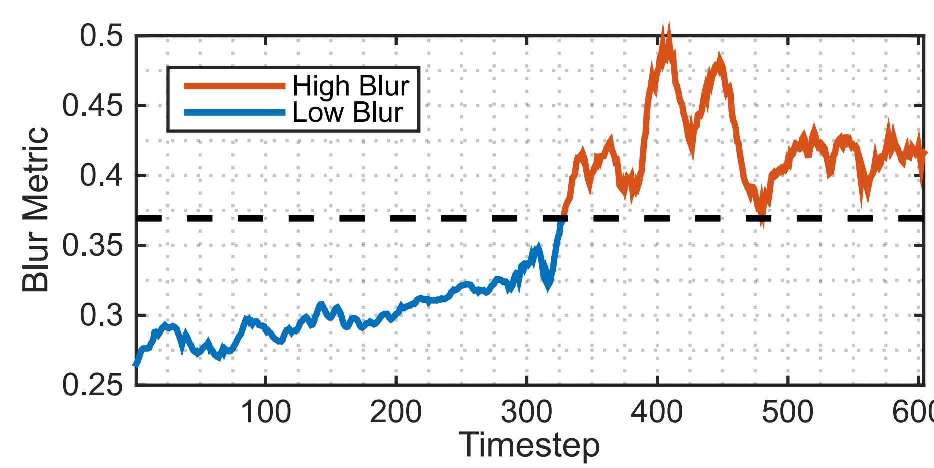
PROBE: Motion Blur

- We quantified reprojection and tracking error with & without motion blur.
- Checkerboard corners are extracted and tracked (using KLT).
- Result: tracking and reprojection error can both be represented by larger variances in additive noise.

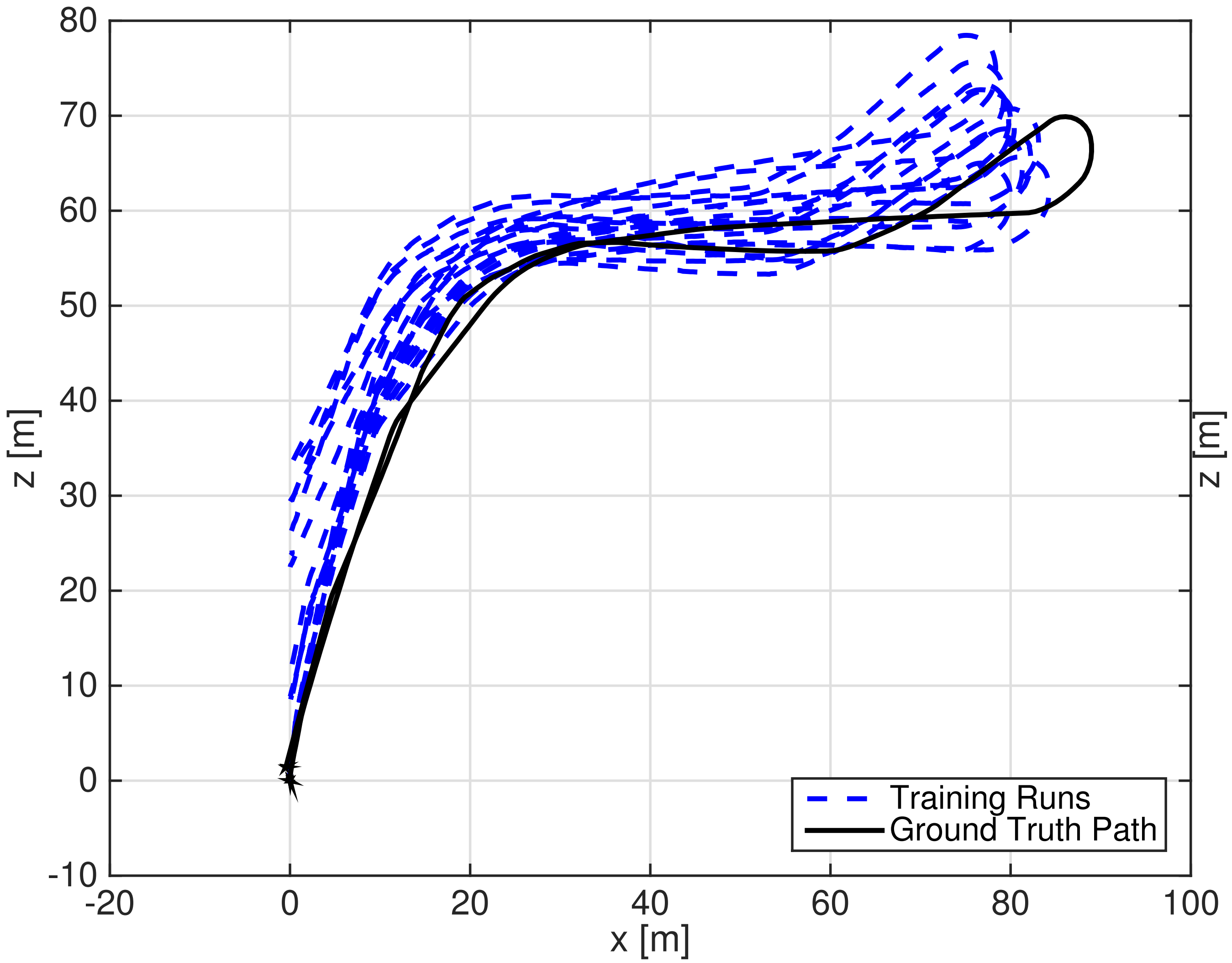


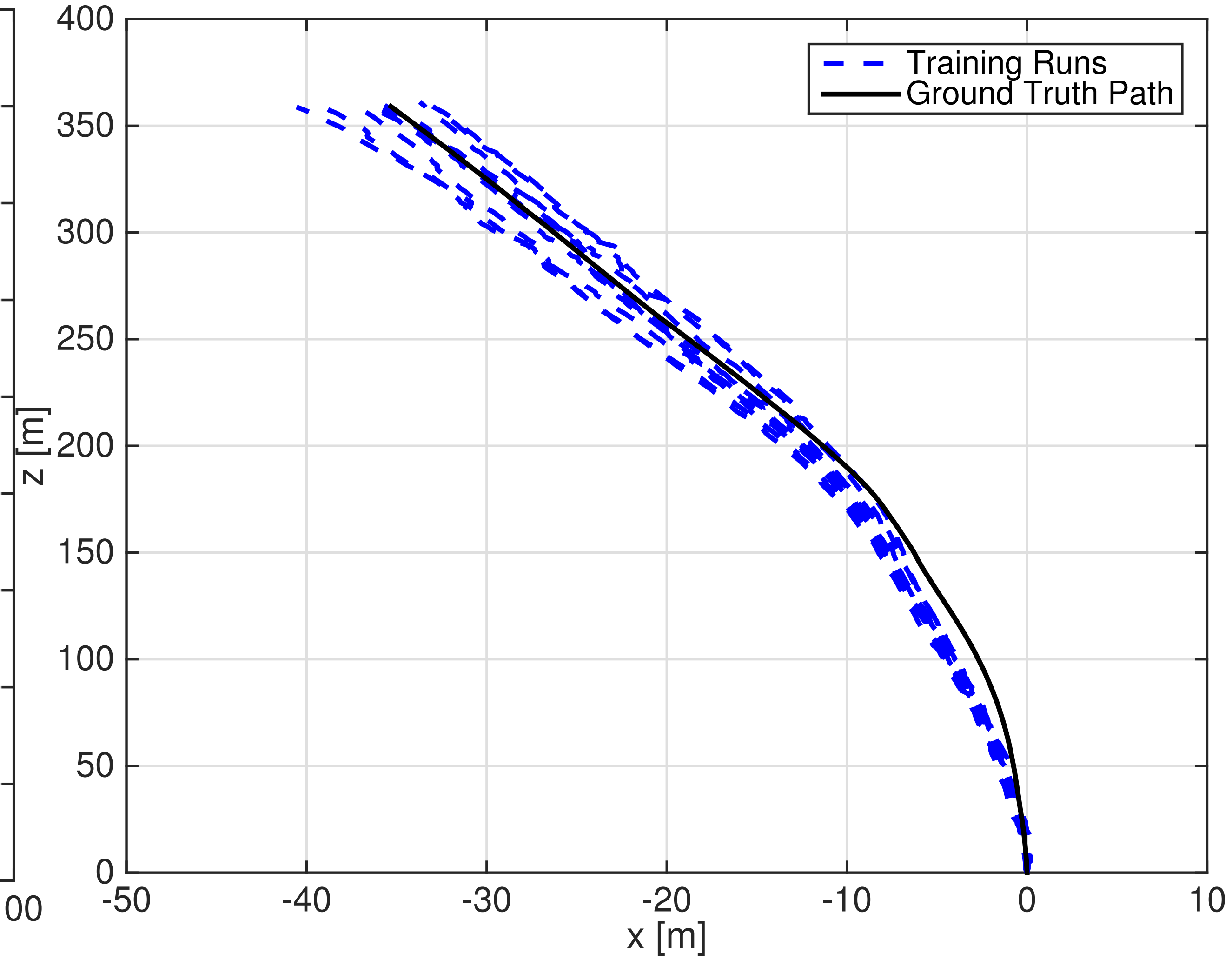
Flea3 @ 125Hz, VI Sensor @ 20 Hz.





PROBE: Training Procedure





1. Perform a training run, selecting a subset of all available features at each step.
2. Compute navigation estimate using feature subset, compare to ground truth, and record RMS error.