ROBOTICS & PERCEPTION GROUP

University of Zurich[UZH]

Lecture 13
Visual Inertial Fusion

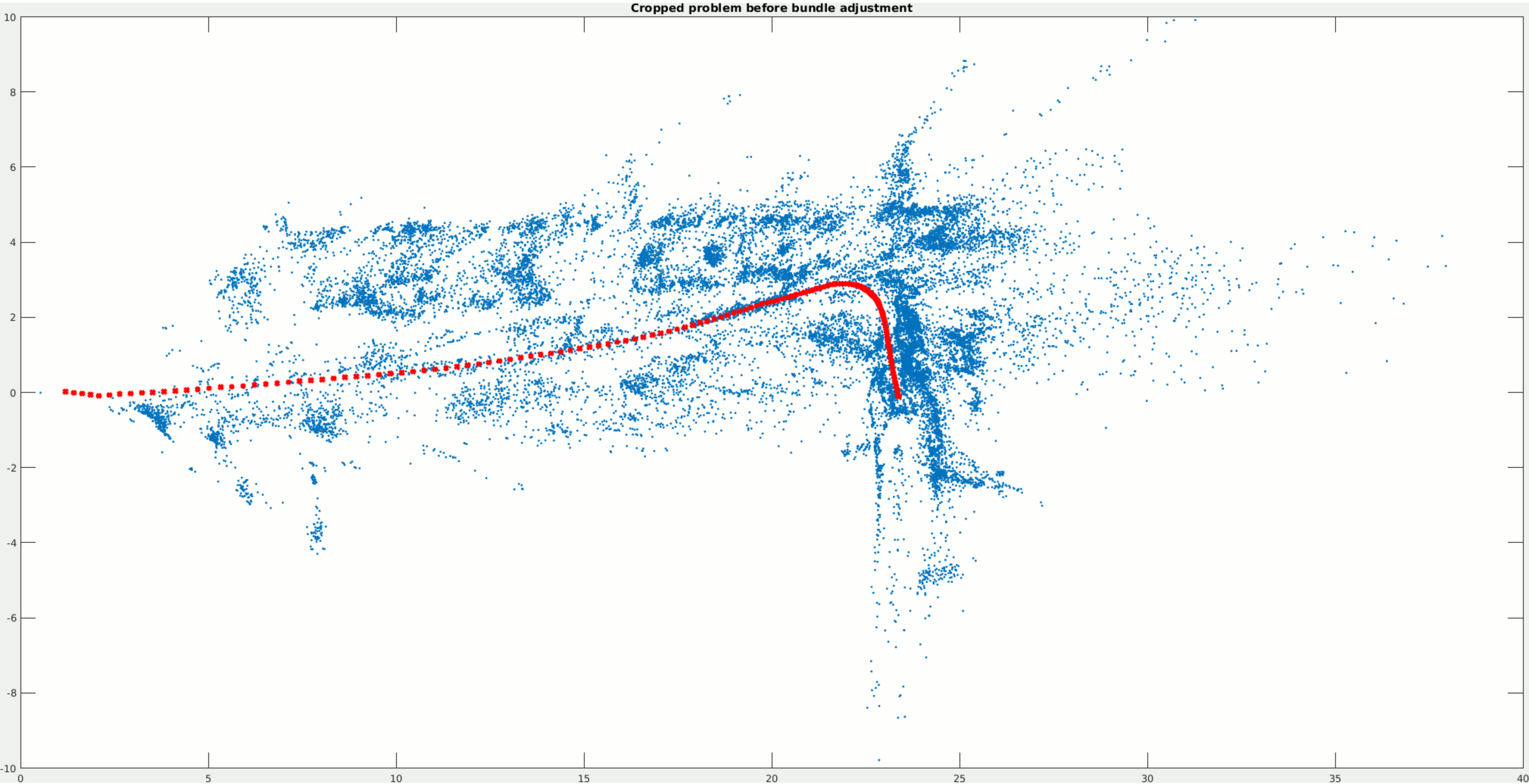
Davide Scaramuzza

Course Evaluation

- Please fill the evaluation form you received by email!
- Provide feedback on
  - Exercises: good and bad
  - Course: good and bad
  - How to improve

Lab Exercise 6 - Today

- Room ETH HG E 33.1 from 14:15 to 16:00
- Work description: Bundle Adjustment

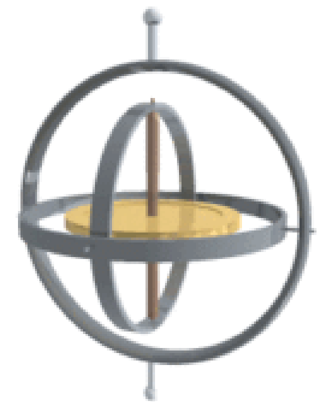


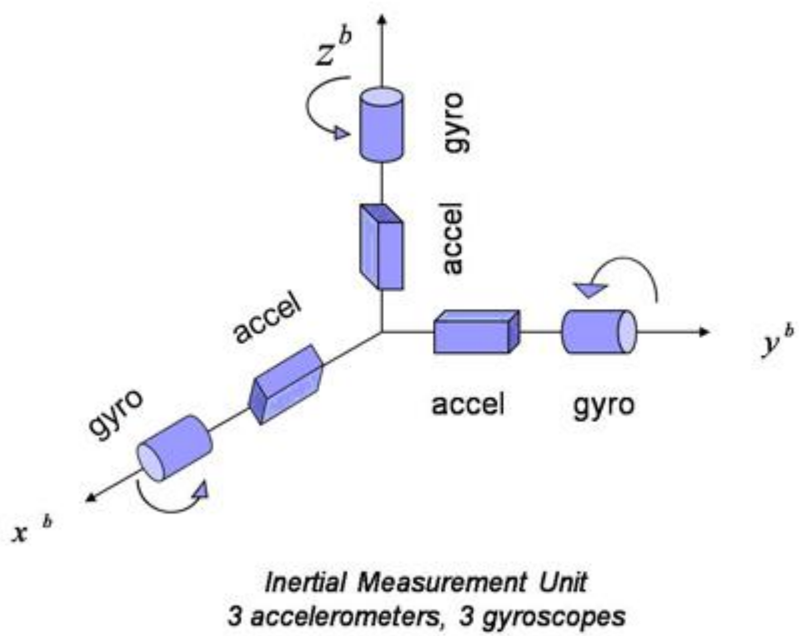
Outline

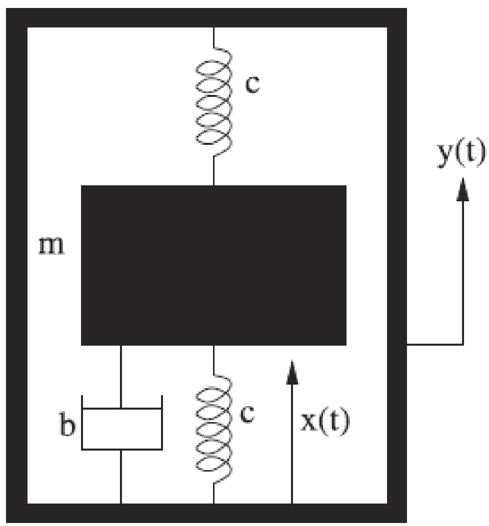
- Introduction
- IMU model and Camera-IMU system
- Different paradigms
  - Filtering
  - Maximum a posteriori estimation
  - Fix-lag smoothing

What is an IMU?

- **I**nertial **M**easurement **U**nit
  - Angular velocity
  - Linear Accelerations





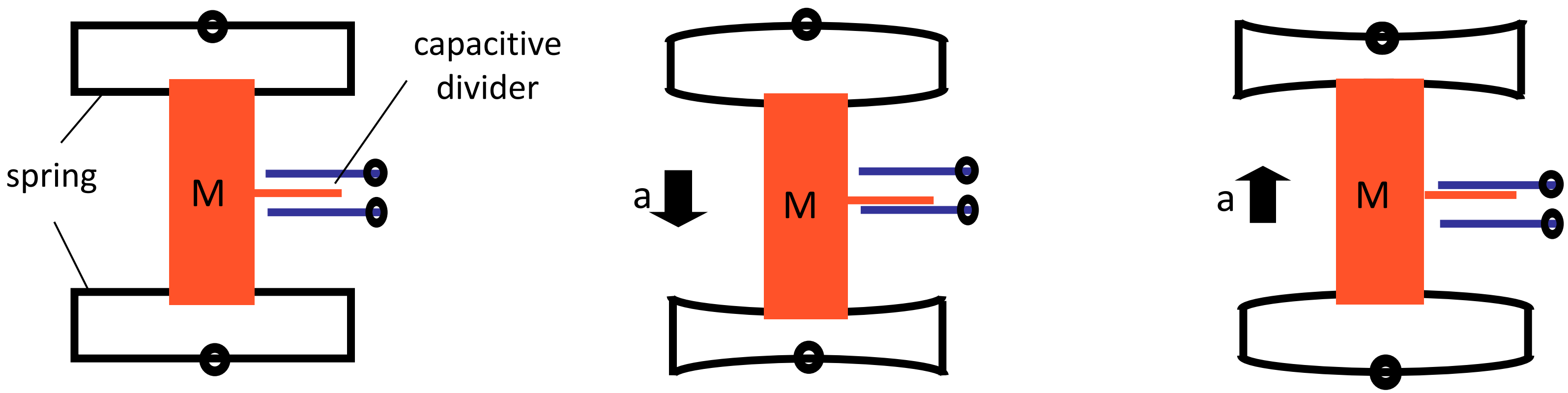


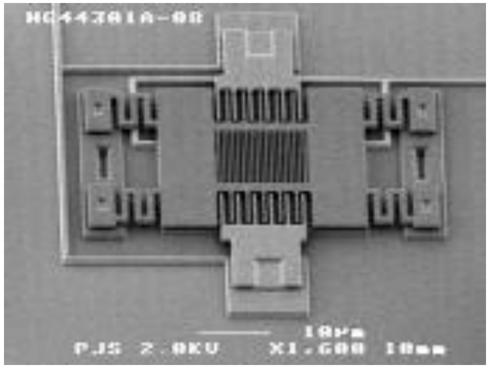
What is an IMU?

- Different categories
  - Mechanical
  - Optical
  - MEMS
  - ....
- For mobile robots: MEMS IMU
  - Cheap
  - Power efficient
  - Light weight and solid state

MEMS Accelerometer

A spring-like structure connects the device to a seismic mass vibrating in a capacity devider. A capacitive divider converts the displacement of the seismic mass into an electric signal. Damping is created by the gas sealed in the device.





MEMS Gyroscopes

- MEMS gyroscopes measure the Coriolis forces acting on MEMS vibrating structures (tuning forks, vibrating wheels, or resonant solids)
- Their working principle is similar to the haltere of a fly
- Haltere are small structures of some two-winged insects, such as flies. They are flapped rapidly and function as gyroscopes, informing the insect about rotation of the body during flight.

Why IMU?

- ➢ Monocular vision is scale ambiguous.
- ➢ Pure vision is not robust enough
  - Low texture
  - High dynamic range
  - High speed motion

Robustness is a critical issue: Tesla accident

“The autopilot sensors on the Model S failed to distinguish a white tractor-trailer crossing the highway against a bright sky. ”

Why vision?

- Pure IMU integration will lead to large drift (especially cheap IMUs)
  - Will see later mathematically
  - Intuition
    - Integration of angular velocity to get orientation: error **proportional to t**
    - Double integration of acceleration to get position: if there is a bias in acceleration, the error of position is **proportional to t^2**
    - Worse, the actually position error also depends on the error of orientation.

	Accelerometer Bias Error		Horizontal Position Error [m]			
Grade	[mg]		1s	10s	60s	1hr
Navigation	0.025		0.13 mm	12 mm	0.44 m	1.6 km
Tactical	0.3		1.5 mm	150 mm	5.3 m	19 km
Industrial	3		15 mm	1.5 m	53 m	190 km
Automotive	125		620 mm	60 m	2.2 km	7900 km

Smartphone accelerometers

http://www.vectornav.com/support/library/imu-and-ins

Why visual inertial fusion?

- Summary: IMU and vision are complementary

Visual sensor	Inertial sensor
✓ Precise in case of non-aggressive motion ✓ Rich information for other purposes x Limited output rate (~100 Hz) x Scale ambiguity in monocular setup. x Lack of robustness	✓ Robust ✓ High output rate (~1,000 Hz) x Large relative uncertainty when at low acceleration/angular velocity x Ambiguity in gravity / acceleration

In common: state estimation based on visual or/and inertial sensor is dead-reckoning, which suffers from drifting over time.
(solution: loop detection and loop closure)

Outline

- Introduction
- IMU model and Camera-IMU system
- Different paradigms
  - Filtering
  - Maximum a posteriori estimation
  - Fix-lag smoothing

IMU model: Measurement Model

➢ Measures angular velocity and acceleration in the body frame:

$$\boxed{{}_{\mathrm{B}}\tilde{\boldsymbol{\omega}}_{\mathrm{WB}}(t)} = {}_{\mathrm{B}}\boldsymbol{\omega}_{\mathrm{WB}}(t) + \boxed{\mathbf{b}^g(t) + \mathbf{n}^g(t)}$$

$$\boxed{{}_{\mathrm{B}}\tilde{\mathbf{a}}_{\mathrm{WB}}(t)} = \mathbf{R}_{\mathrm{BW}}(t)\left({}_{\mathrm{W}}\mathbf{a}_{\mathrm{WB}}(t) - {}_{\mathrm{W}}\mathbf{g}\right) + \boxed{\mathbf{b}^a(t) + \mathbf{n}^a(t)}$$

measurements

noise

where the superscript g stands for Gyroscope and a for Accelerometer

Notations:

- Left subscript: reference frame in which the quantity is expressed
- Right subscript {*Q*}{Frame1}{Frame2}: *Q* of Frame2 with respect to Frame1
- Noises are all in the body frame

IMU model: Noise Property

- Additive Gaussian white noise: $\mathbf{n}^g(t),\ \mathbf{n}^a(t)$

$$E\left[n(t)\right]=0$$

$$E\left[n(t_1)n(t_2)\right]=\sigma^2\delta(t_1-t_2)$$

$$n[k]=\sigma_d w[k]$$

$$w[k]\sim N(0,1)$$

$$\sigma_d=\sigma/\sqrt{\Delta\mathrm{t}}$$

- Bias: $\mathbf{b}^g(t),\ \mathbf{b}^a(t)$

$$\dot{\mathbf{b}}(t)=\sigma_b\mathbf{w}(t)$$

i.e., the derivative of the bias is white Gaussian noise (so-called random walk)

$$\mathbf{b}[k]=\mathbf{b}[k-1]+\sigma_{bd}\mathbf{w}[k]$$

$$\sigma_{bd}=\sigma_b\sqrt{\Delta t}$$

$$w[k]\sim N(0,1)$$

The biases are usually estimated with the other states

- can change every time the IMU is started
- can change due to temperature change, mechanical pressure, etc.

Trawny, Nikolas, and Stergios I. Roumeliotis. "Indirect Kalman filter for 3D attitude estimation.“
https://github.com/ethz-asl/kalibr/wiki/IMU-Noise-Model

IMU model: Integration

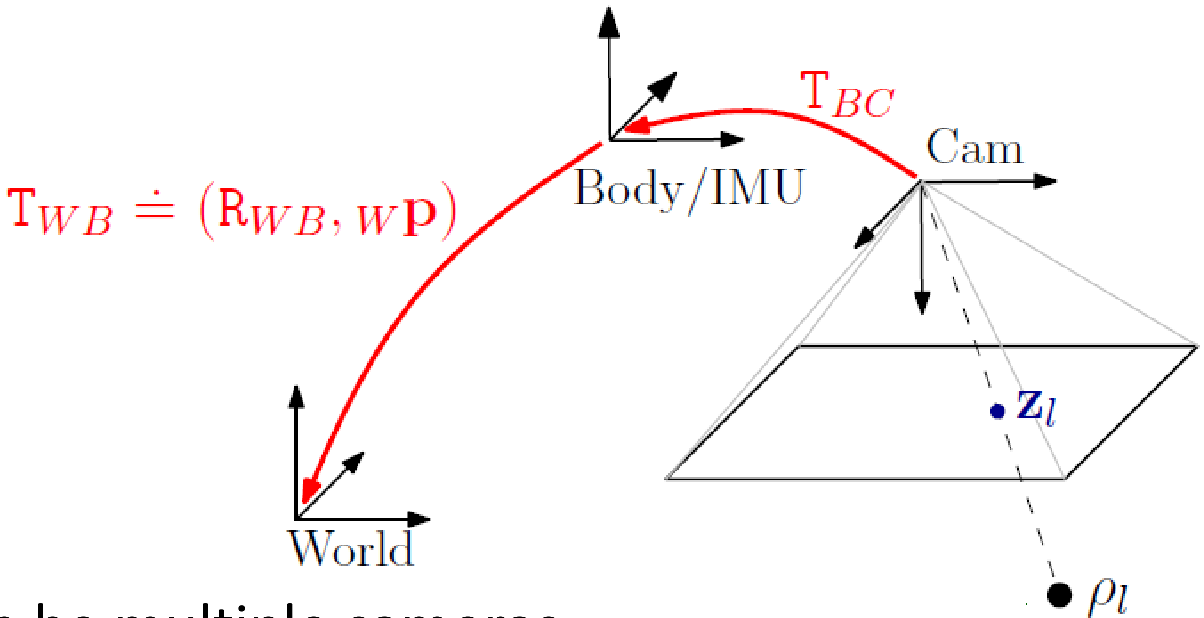
- Per component: {*t*} stands for {B}ody frame at time *t*

$$\mathbf{p}_{\mathrm{Wt}_2} = \boxed{\mathbf{p}_{\mathrm{Wt}_1}} + (t_2 - t_1)\boxed{\mathbf{v}_{\mathrm{Wt}_1}} + \int_{t_1}^{t_2}\!\!\int \left(\boxed{\mathbf{R}_{\mathrm{Wt}}}(t)\left(\tilde{\mathbf{a}}(t) - \mathbf{b}^a(t)\right) + {}_{\mathrm{w}}\mathbf{g}\right)\mathrm{d}t^2$$

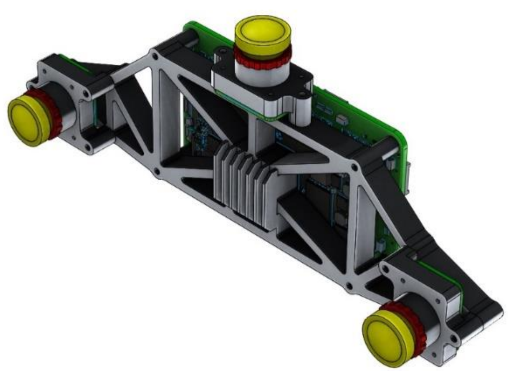
- Depends on initial position and velocity
- The rotation R(t) is computed from the gyroscope

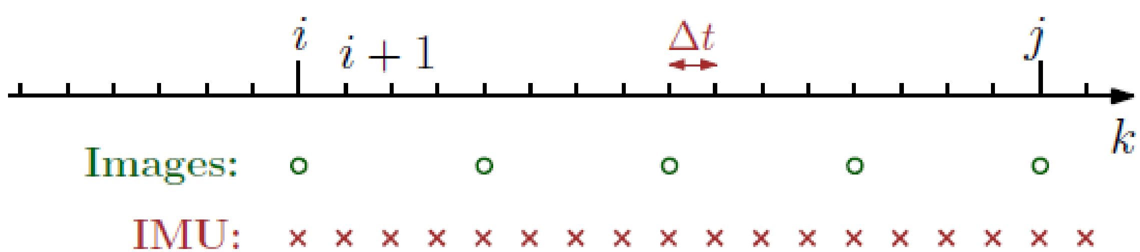
Trawny, Nikolas, and Stergios I. Roumeliotis. "Indirect Kalman filter for 3D attitude estimation."

Camera-IMU System



There can be multiple cameras.





Outline

- Introduction
- IMU model and Camera-IMU system
- Different paradigms
  - Closed-form solution
  - Filtering
  - Fix-lag smoothing
  - Maximum a posteriori estimation

Closed-form Solution (intuitive idea)

- The absolute pose x is known up to a scale s, thus

$$x = s\tilde{x}$$

- From the IMU

$$x = x_0 + v_0(t_1 - t_0) + \iint_{t_0}^{t_1} a(t)dt$$

- By equating them

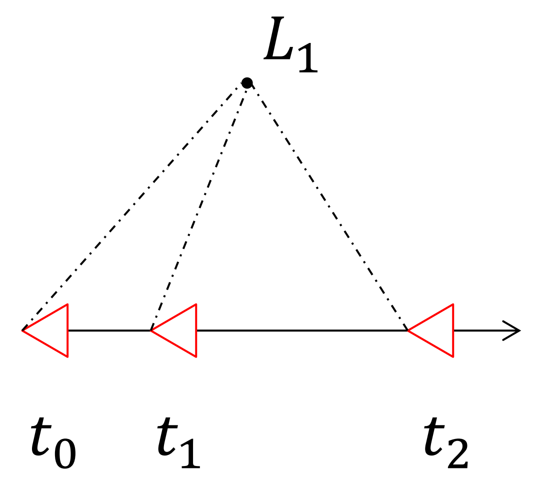
$$s\tilde{x} = x_0 + v_0(t_1 - t_0) + \iint_{t_0}^{t_1} a(t)dt$$

- As shown in [Martinelli'14], for 6DOF, both s and v_0 can be determined in closed form from a **single feature observation and 3 views**. x_0 can be set to 0.

Martinelli, Closed-form solution of visual-inertial structure from motion, International Journal of Computer Vision, 2014

Closed-form Solution

$$\begin{cases} s\widetilde{x_1} = v_0(t_1 - t_0) + \iint_{t_0}^{t_1} a(t)dt \\ s\widetilde{x_2} = v_0(t_2 - t_0) + \iint_{t_0}^{t_2} a(t)dt \end{cases}$$



$$\begin{bmatrix} \widetilde{x_1} & (t_0 - t_1) \\ \widetilde{x_2} & (t_0 - t_2) \end{bmatrix} \begin{bmatrix} s \\ v_0 \end{bmatrix} = \begin{bmatrix} \iint_{t_0}^{t_1} a(t)dt \\ \iint_{t_0}^{2} a(t)dt \end{bmatrix}$$

Martinelli, Closed-form solution of visual-inertial structure from motion, International Journal of Computer Vision, 2014

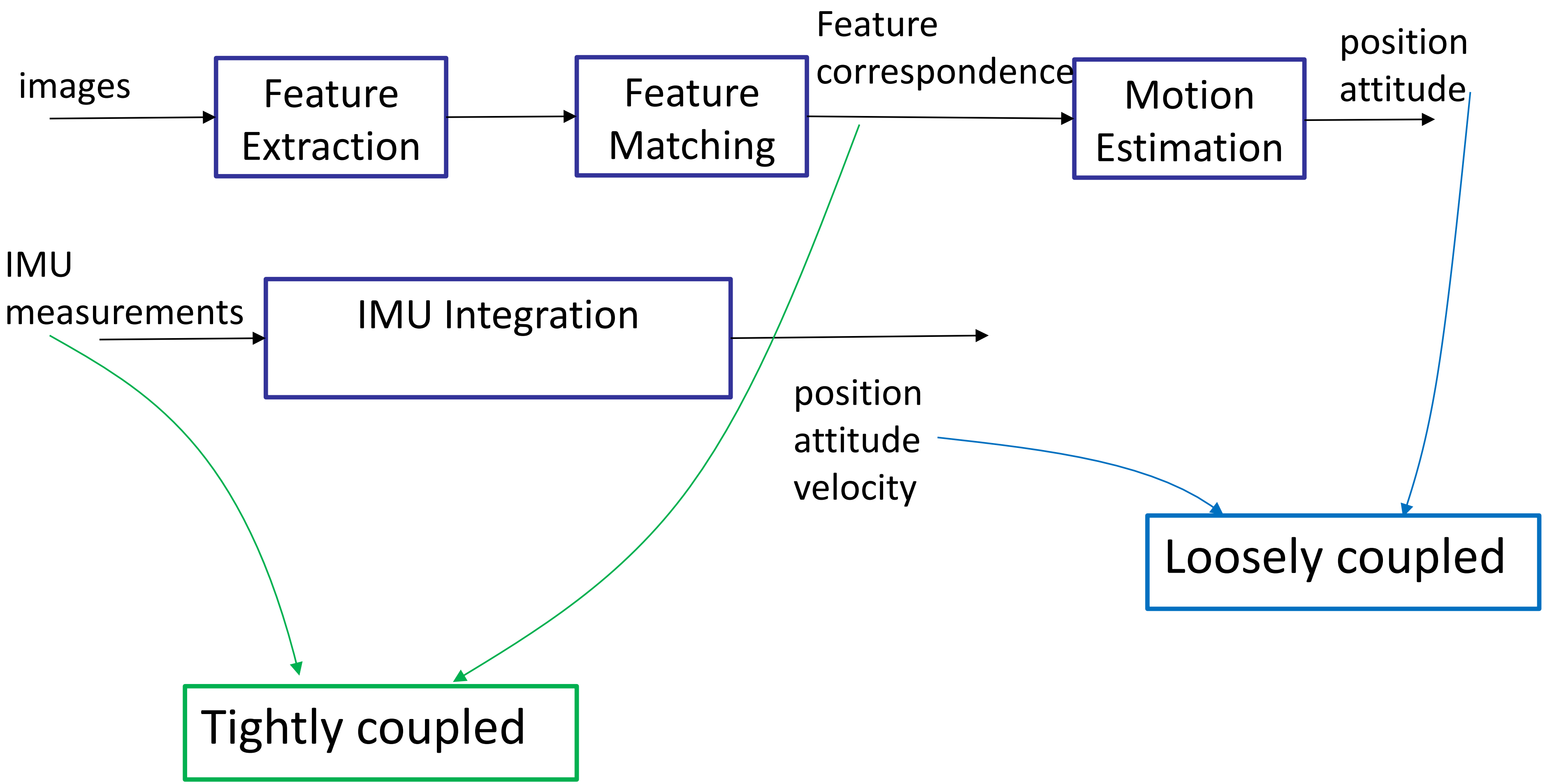
Different paradigms

- Loosely coupled: use the output of individual system
  - Estimate the states individually from visual and inertial data
  - Combine the separate states estimations
- Tightly coupled: use the internal states
  - Make use of the raw measurements
    - Feature positions
    - IMU readings
    - …
  - Example:
    - Use IMU for guided feature matching
    - Minimizing reprojection error and IMU error together
    - …

More accurate

More implementation effort.

Different paradigms



More accurate

More implementation effort.

Different paradigms

Filtering	Fix-lag Smoothing	Maximum-A-Posteriori (MAP) Estimation
Filtering the most recent states • (e.g., extended Kalman filter)	Optimize window of states • Marginalization • Nonlinear least squares optimization	Optimize all states • Nonlinear Least squares optimization
×1 Linearization ×Accumulation of linearization errors ×Gaussian approximation of marginalized states ✓Faster	✓Re-Linearize ×Accumulation of linearization errors ×Gaussian approximation of marginalized states ✓Fast	✓Re-Linearize ✓Sparse Matrices ✓Highest Accuracy ×Slow

Filtering: Visual Inertial Formulation

System states:

Tightly coupled: $\mathbf{X} = \left[\boxed{{}_{\mathrm{W}}\mathbf{p}(t);\ \mathbf{q}_{\mathrm{WB}}(t);\ {}_{\mathrm{W}}\mathbf{v}(t); \mathbf{b}^{a}(t); \mathbf{b}^{g}(t);}\ {}_{\mathrm{W}}\mathbf{L}_1;\ {}_{\mathrm{W}}\mathbf{L}_2; \ldots; {}_{\mathrm{W}}\mathbf{L}_K \right]$

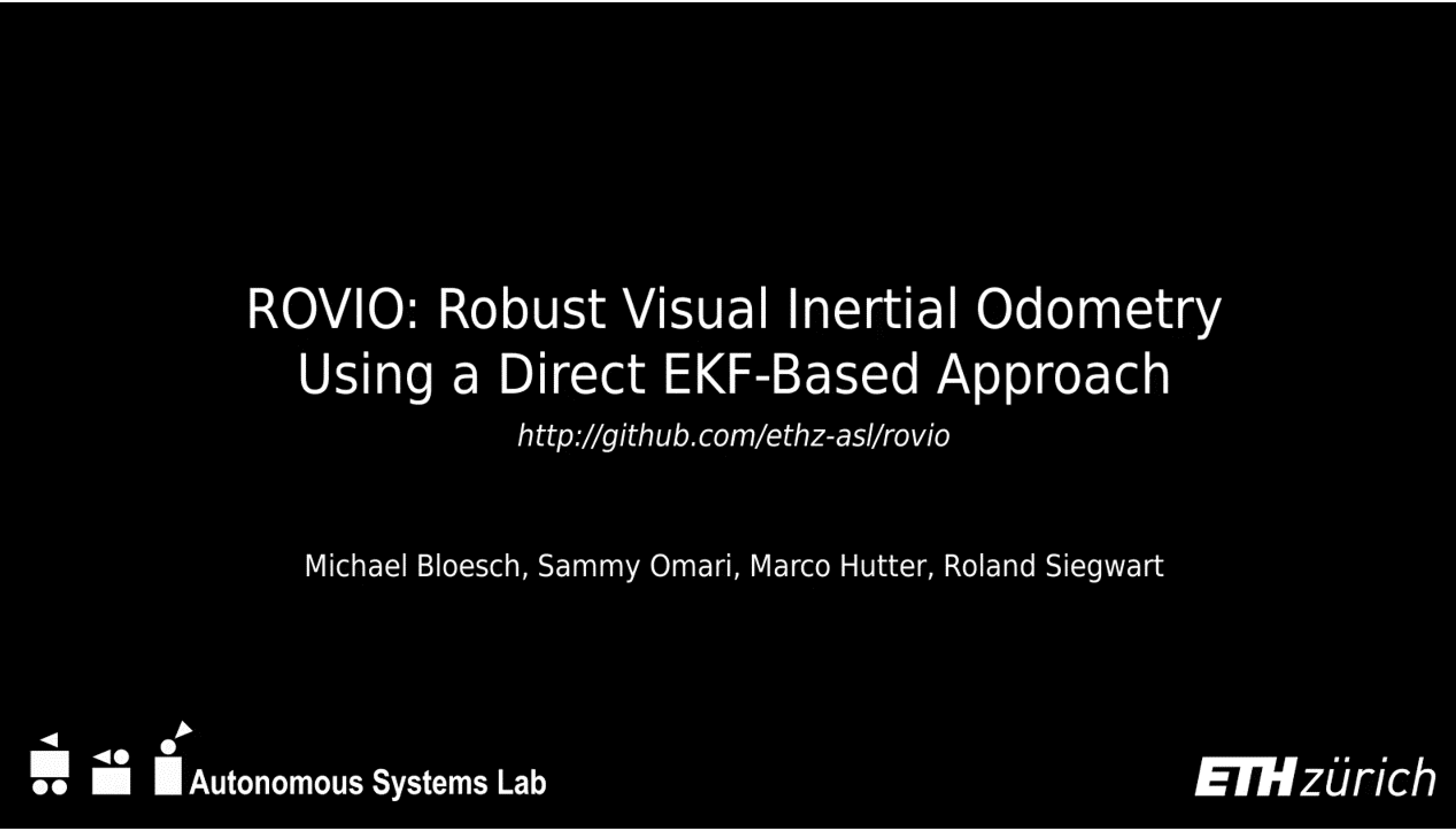
Loosely coupled: $\mathbf{X} = \left[{}_{\mathrm{W}}\mathbf{p}(t);\ \mathbf{q}_{\mathrm{WB}}(t);\ {}_{\mathrm{W}}\mathbf{v}(t); \mathbf{b}^{a}(t); \mathbf{b}^{g}(t) \right]$

Process Model: from IMU

- Integration of IMU states (rotation, position, velocity)
- Propagation of IMU noise
  - needed for calculating the Kalman Filter gain

Filtering: ROVIO

Use pixel intensities as measurements.



Bloesch, Michael, et al. "Robust visual inertial odometry using a direct EKF-based approach."

Filtering: Potential Problems

- Wrong linearization point
  - Linearization depends on the current estimates of states, which may be erroneous
  - Linearization around different values of the same variable leads to estimator inconsistency (wrong observability/covariance estimation)
- Wrong covariance/initial states
  - Intuitively, wrong weights for measurements and prediction
  - May be overconfident/underconfident
- Explosion of number of states
  - Each 3D point: 3 variables

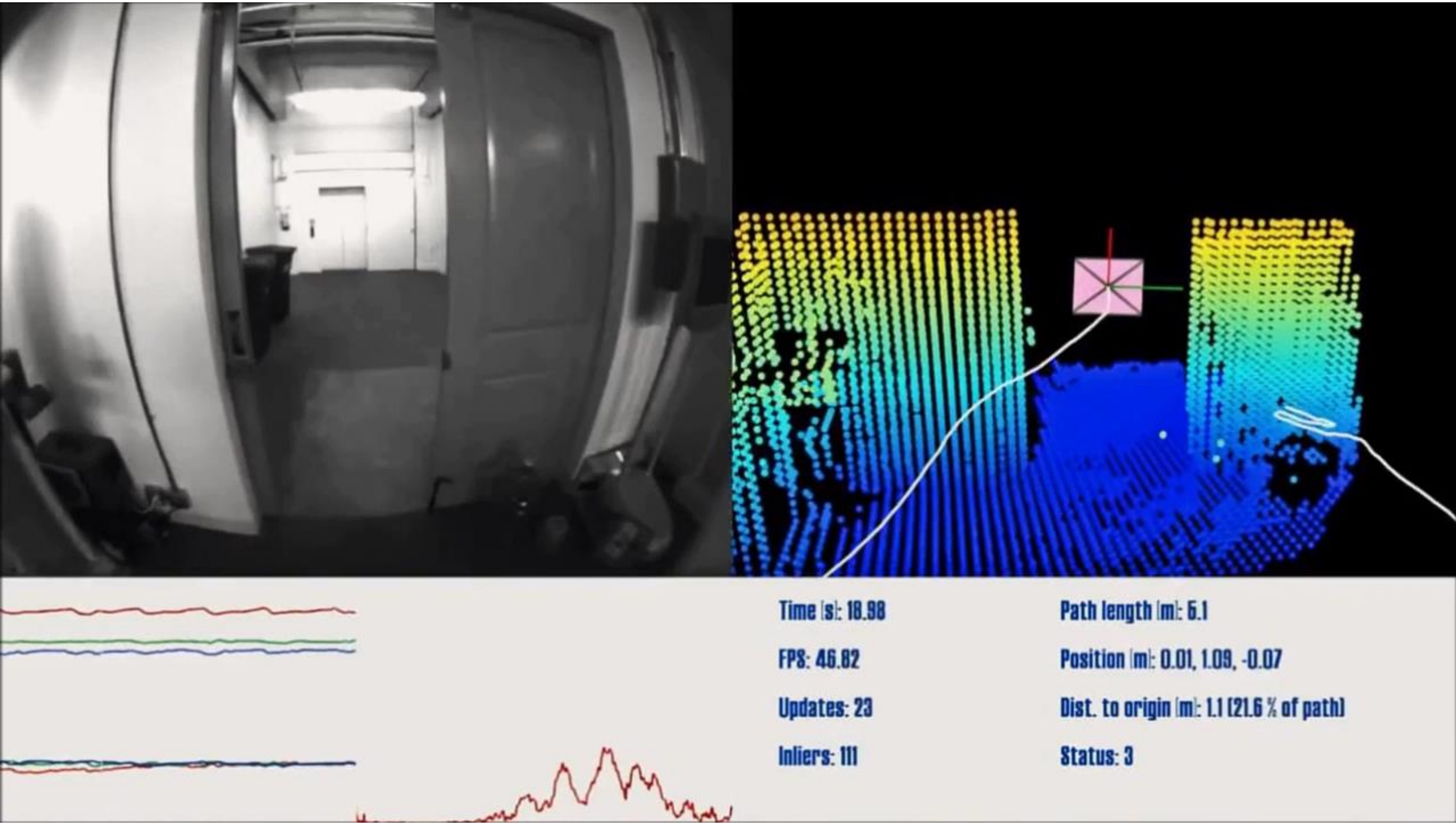
Filtering: MSCKF (Multi-State Constraint Kalman Filter): used in Google Tango

- Key idea:
  - Keep a window of recent states
  - incorporate visual observations **without** including **point positions** into the states

Li, Mingyang, and Anastasios I. Mourikis. "High-precision, consistent EKF-based visual–inertial odometry."

Filtering: Google Tango

Project Tango



Optimization-based Approaches (MAP, Fix-lag smoothing)

- Fusion solved as a *non-linear optimization problem*
- Increased accuracy over filtering methods

$$x_k = f(x_{k-1})$$
$$z_i = h(x_{i_k}, l_{i_j})$$

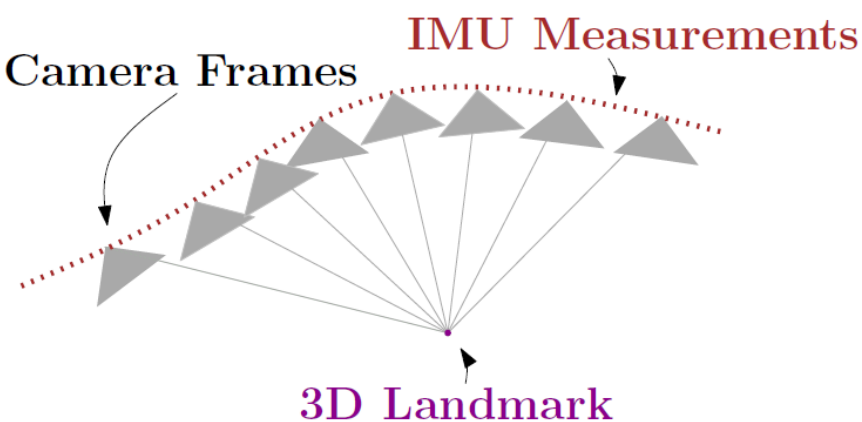
$X = \{x_1, \ldots x_N\}$: robot states
$L = \{l_1, \ldots\}$: 3D points
$Z = \{z_i, \ldots z_M\}$: feature positions

$$\{X^*, L^*\} = \underset{\{X,L\}}{\operatorname{argmax}} P(X, L \mid Z)$$
$$= \underset{\{X,L\}}{\operatorname{arg\,min}} \left\{ \sum_{k=1}^{N} \left\| f(x_{k-1}) - x_k \right\|_{\Lambda_k}^{2} + \sum_{i=1}^{M} \left\| h(x_{i_k}, l_{i_j}) - z_i \right\|_{\Sigma_i}^{2} \right\}$$

Forster, Carlone, Dellaert, Scaramuzza, IMU Preintegration on Manifold for efficient Visual-Inertial Maximum-a-Posteriori Estimation, *Robotics Science and Systens*'15, **Best Paper Award Finalist**

Optimization-based Approaches (MAP, Fix-lag smoothing)

- Fusion solved as a *non-linear optimization problem*
- Increased accuracy over filtering methods

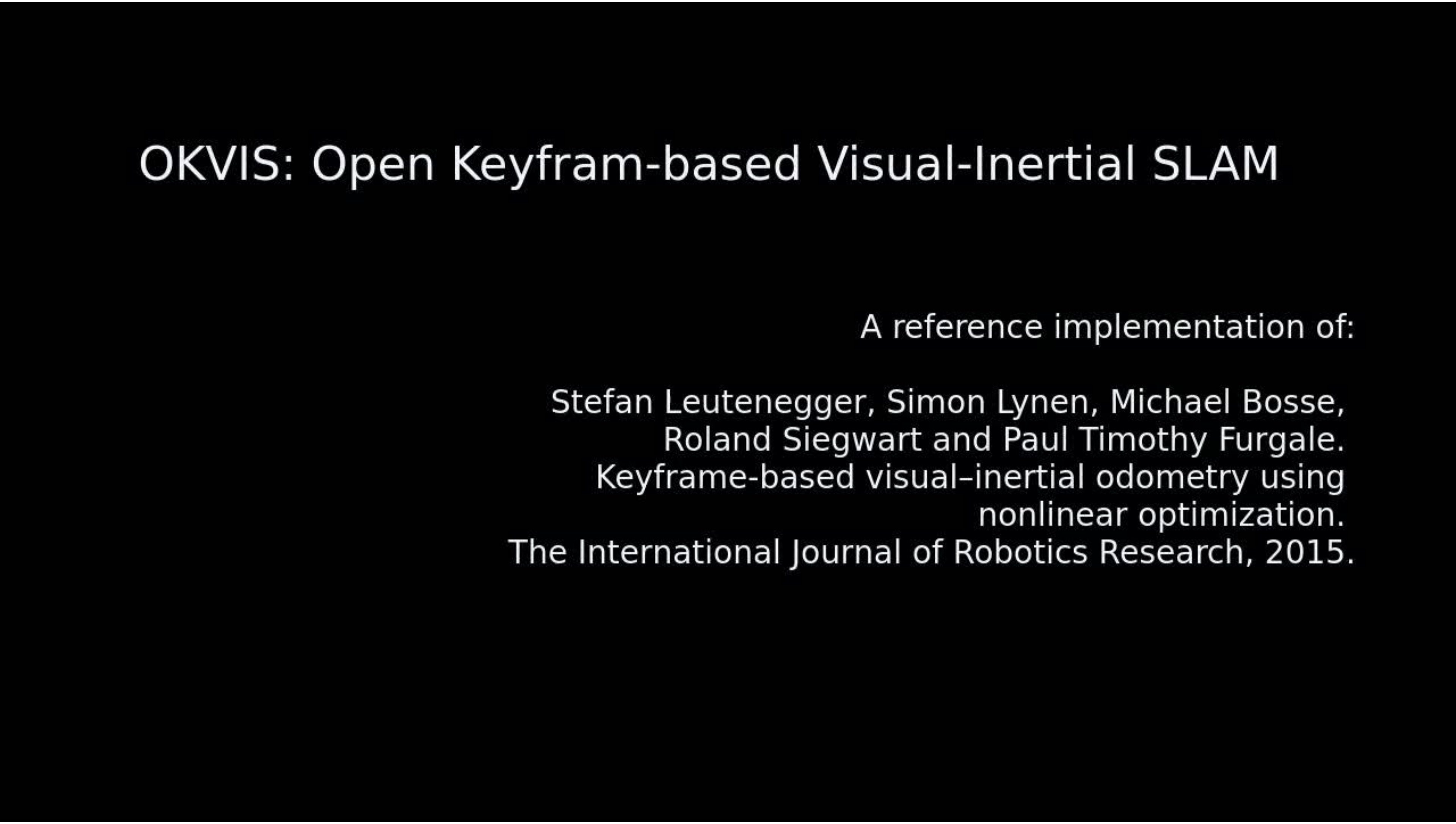


$$\{X^*, L^*\} = \underset{\{X, L\}}{\operatorname{argmax}} P(X, L \mid Z)$$

$$= \underset{\{X, L\}}{\arg\min} \left\{ \sum_{k=1}^{N} \left\| f(x_{k-1}) - x_k \right\|_{\Lambda_k}^{2} + \sum_{i=1}^{M} \left\| h(x_{i_k}, l_{i_j}) - z_i \right\|_{\Sigma_i}^{2} \right\}$$

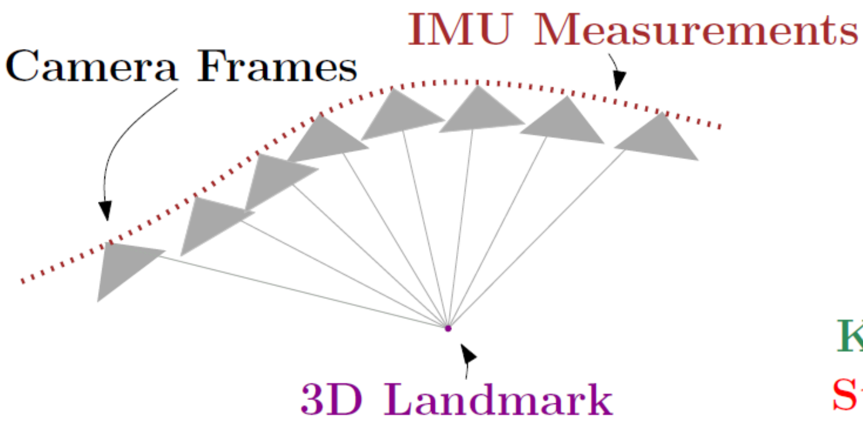
Forster, Carlone, Dellaert, Scaramuzza, IMU Preintegration on Manifold for efficient Visual-Inertial Maximum-a-Posteriori Estimation, *Robotics Science and Systens*'15, **Best Paper Award Finalist**

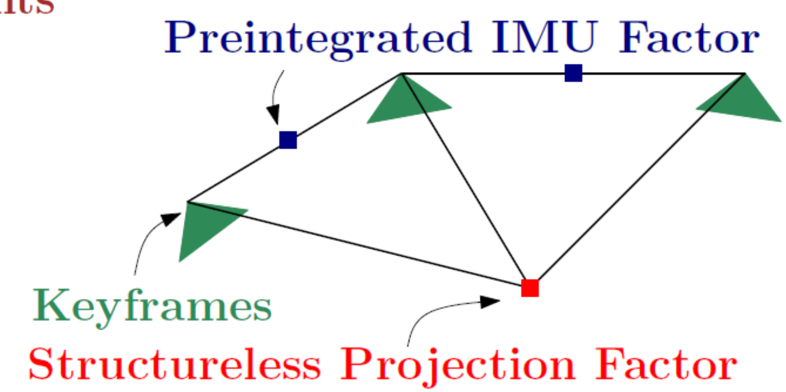
Fix-lag smoothing: OKVIS



Optimization-based Approaches (MAP, Fix-lag smoothing)

- Fusion solved as a *non-linear optimization problem*
- Increased accuracy over filtering methods



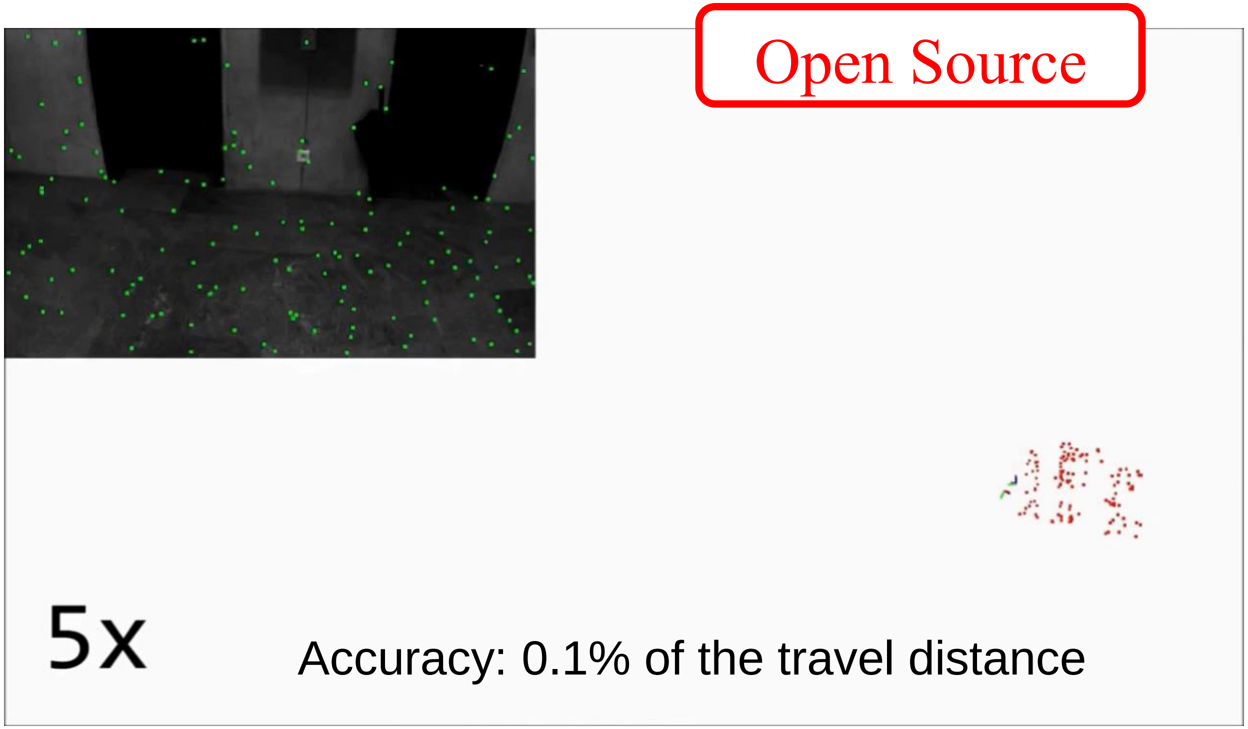


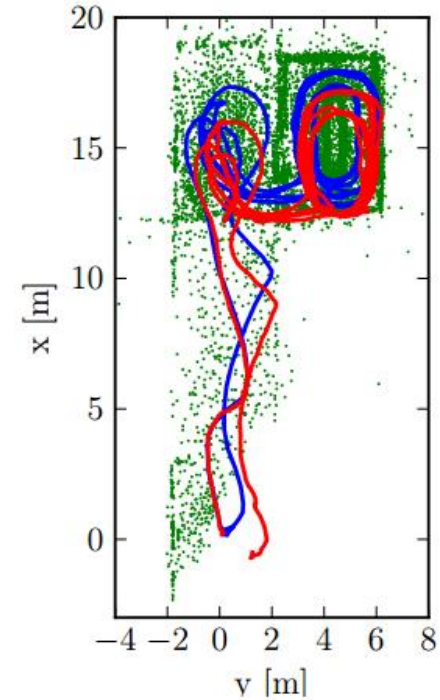
$$\sum_{(i,j)\in\mathcal{K}_k} \|\mathbf{r}_{\mathcal{I}_{ij}}\|^2_{\Sigma_{ij}} + \sum_{i\in\mathcal{K}_k} \sum_{l\in\mathcal{C}_i} \|\mathbf{r}_{\mathcal{C}_{il}}\|^2_{\Sigma_{\mathcal{C}}}$$

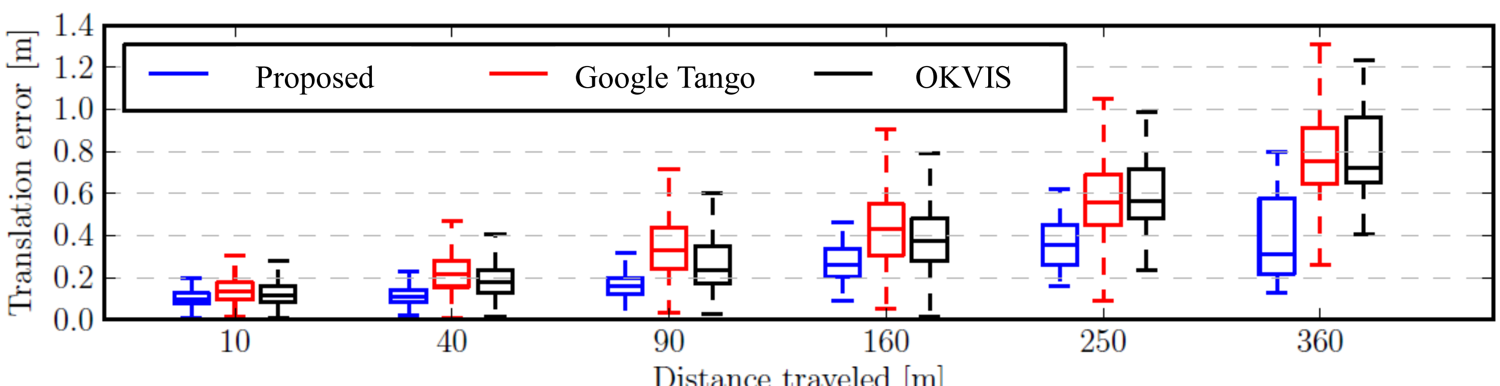
IMU residuals *Reprojection residuals*

Forster, Carlone, Dellaert, Scaramuzza, IMU Preintegration on Manifold for efficient Visual-Inertial Maximum-a-Posteriori Estimation, *Robotics Science and Systens'*15, **Best Paper Award Finalist**

MAP: SVO + IMU Preintegration

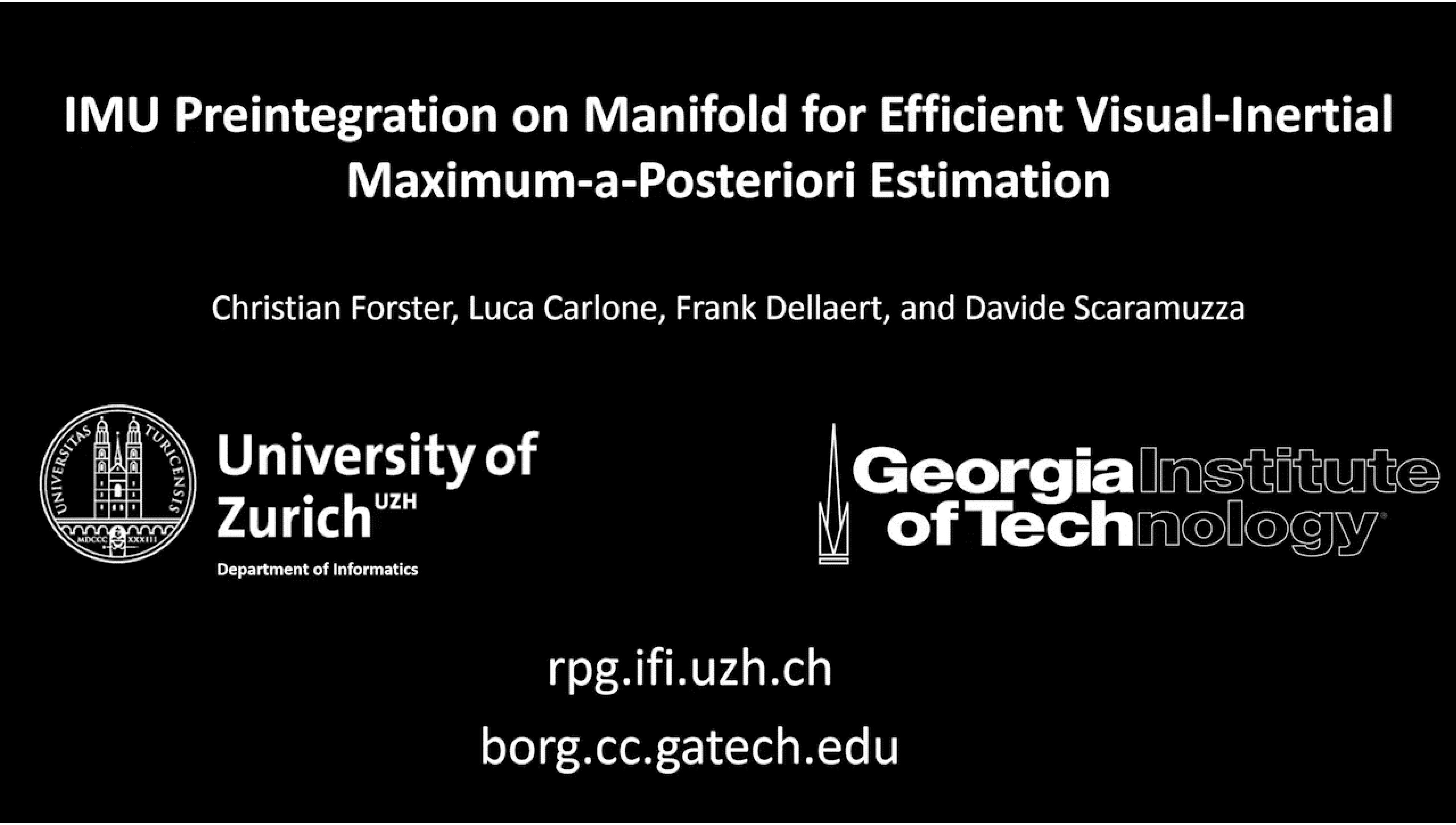






Forster, Carlone, Dellaert, Scaramuzza, IMU Preintegration on Manifold for efficient Visual-Inertial Maximum-a-Posteriori Estimation, *Robotics Science and Systens*’15 and IEEE TRO’16, **Best Paper Award Finalist**

MAP: SVO + IMU Preintegration



Further topics

- Rotation parameterization
  - Rotation is tricky to deal with…
  - Euler angle / Rotation matrix / Quaternion / SO(3)
- Consistency: filtering and fix-lag smoothing
  - Linearization around different values of the same variable may lead to error