VILENS – The Challenge of Visual Navigation on Legged Robots

UNIVERSITY OF OXFORD

Dr. Maurice Fallon,

Lecturer / Royal Society Research Fellow

Research by: David Wisth and Dr. Marco Camurri

Dynamic Robot Systems Group, Oxford Robotics Institute

State estimation – target applications

EU H2020 Project

Muddy Tunnel – DARPA SubT (x3)

City of Zurich Sewer

Videos: courtesy of DARPA & ETH

State estimation – the challenge

Dynamic Walking – at 0.5m/sec on Flat Ground
Precise Estimate Not Important

Static Stair Climbing
Precise Estimate due to slow motion

Limitations for Dynamic Control e.g. stair climbing:

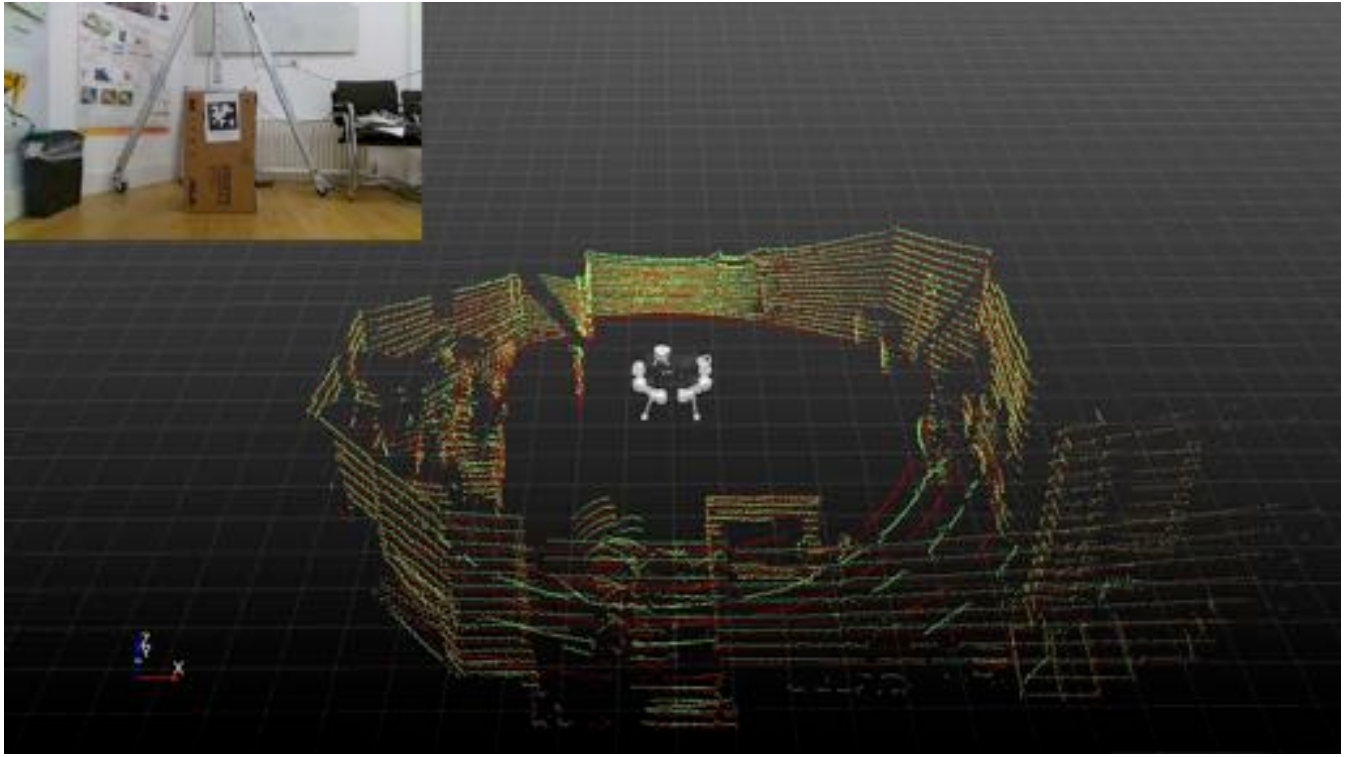
- Stronger leg motors. Lighter robot
- More dynamic control algorithms
- **Poor elevation mapping due to poor state estimation**

State Estimate for Dynamic Gaits has high drift due to:

- Contact Classification Errors. Structural bending
- Rough contact events. **Soft ground**

State estimation – the effect

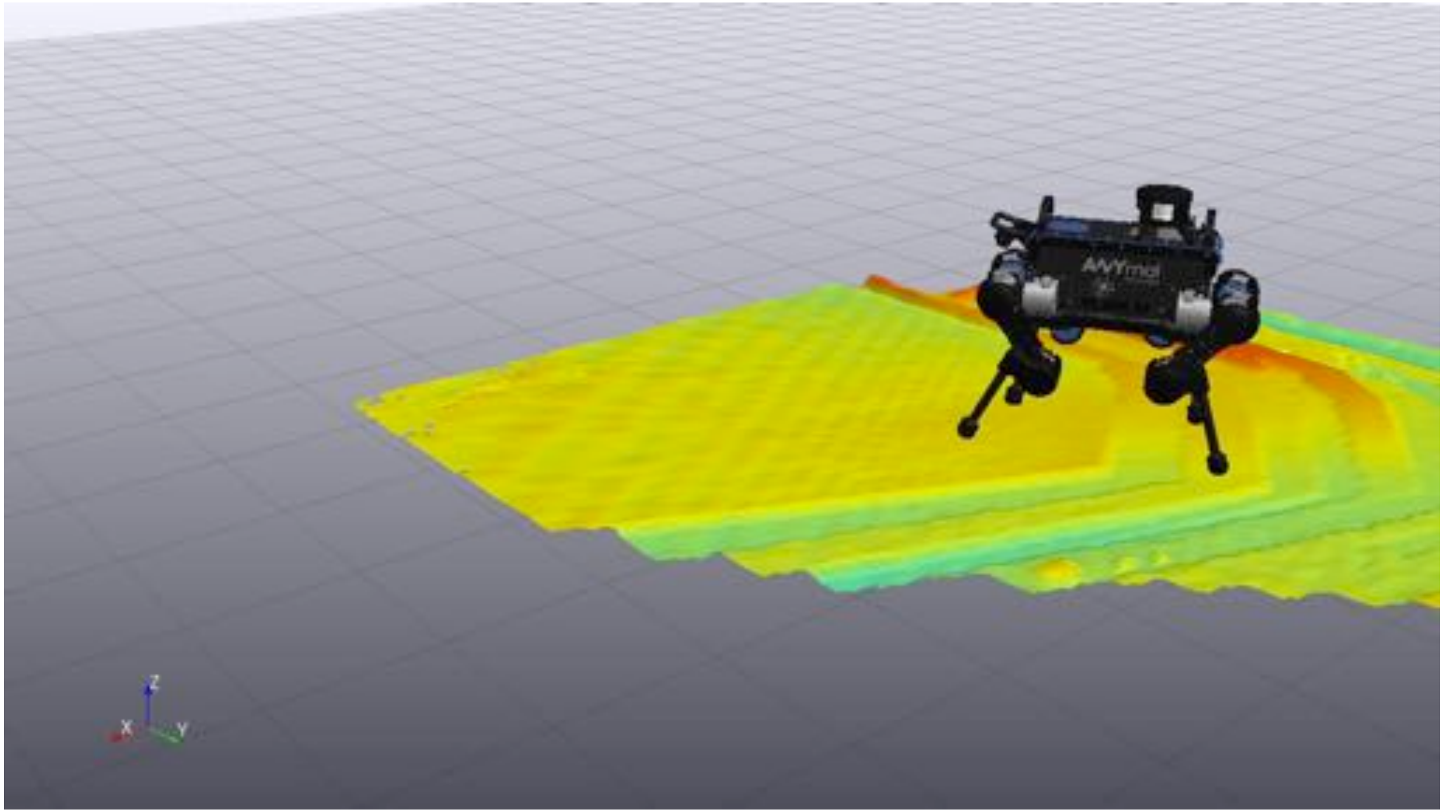
- Essential for control, motion planning, navigation, etc.
- Traditionally, legged robots rely on kinematic-inertial inputs and suffer from drift.



Estimate from Kinematic-Inertial Estimator (TSIF)

State estimation – the effect

- This drift can cause issues for other systems (e.g. terrain mapping)



Estimate from Kinematic-Inertial Estimator (TSIF)

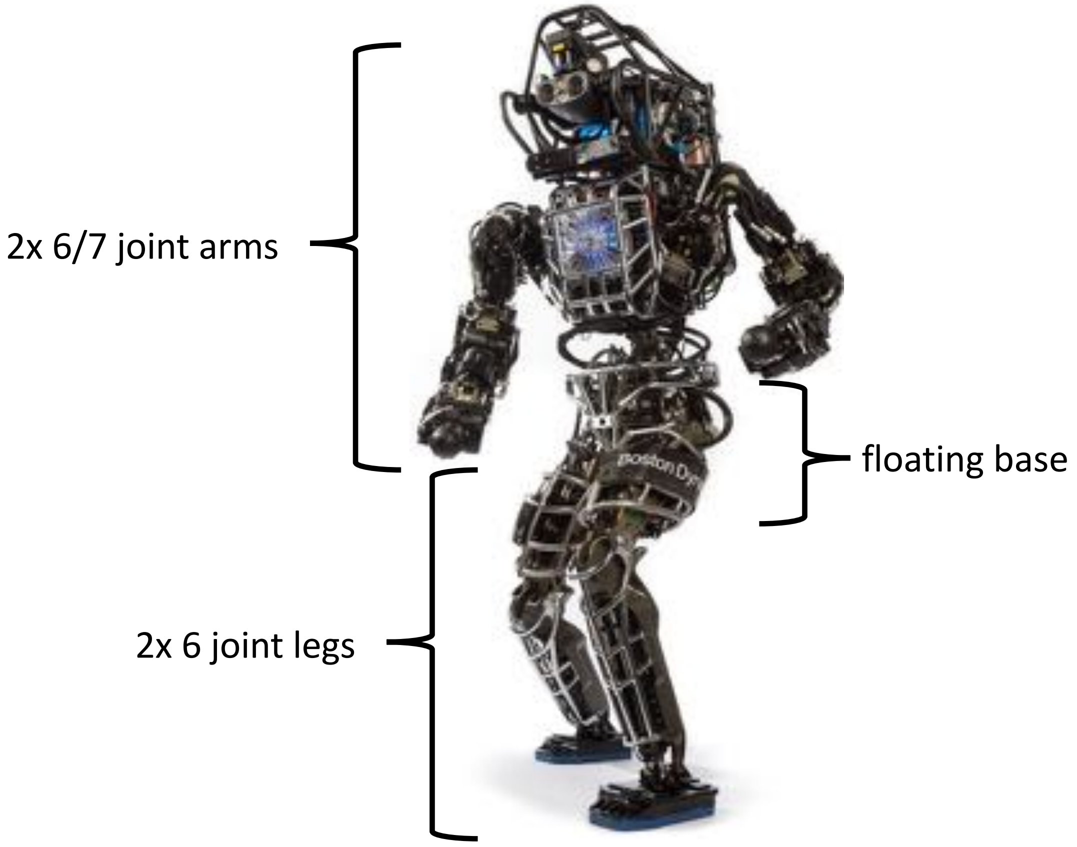
Kinematic State Estimation - Overview

Goal: To estimate the state of the robot:

- Position and orientation of base (6)
- **Linear and Angular Velocities of base (6)**
- Angular Rate and Acceleration biases (6)
- **Joint angles and velocities (N*2)**

For a quadruped:
- 6*2 + 12*2 + 6 biases = 36 states

For a biped:
- 6*2 + 27*2 + 6 biases = 72 states



Kinematic State Estimation - Overview

Goal: To estimate the **state of the robot's base**

- Position, orientation and velocity
- High frequency (250/400 Hz)
- Low latency (2-3 msec)

Core Approach: (Extended) Kalman Filter

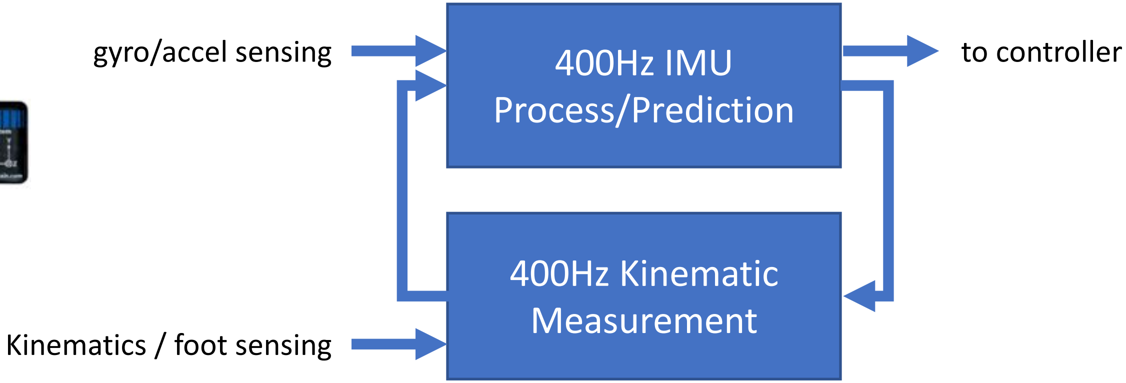
- IMU-driven process model
- Leg kinematics used to measure **linear velocity**
- **No external sensors (cameras, LIDAR)**
- **Crucial:** contact classification using forces

Main Challenges:

- Despite modelling, behaviour in contact is unknown
- **Latency is unacceptable**: estimate used in control

Implementations:

- Anybotics/RSL: TSIF
- Oxford: Pronto



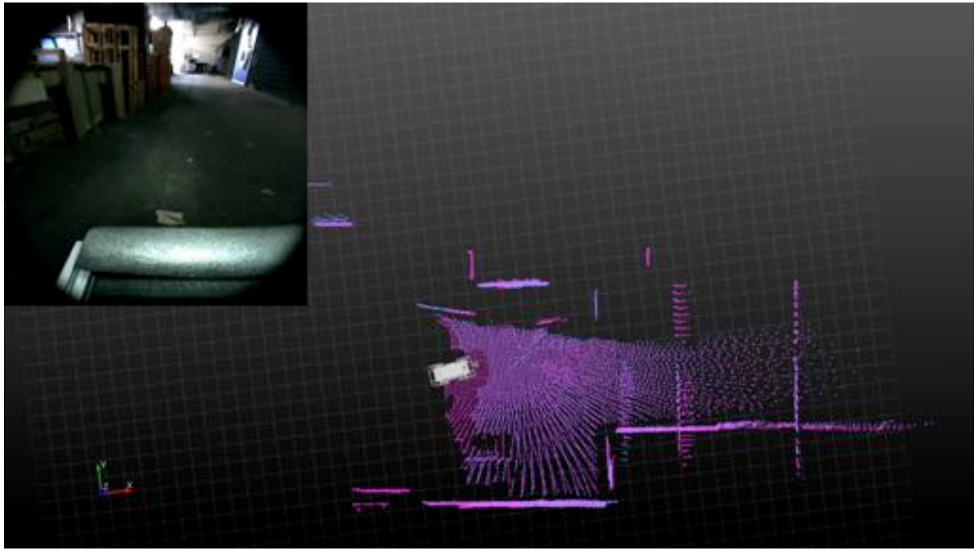
Kinematic State Estimation - Examples

UNIVERSITY OF OXFORD

With Boston Dynamics Atlas Biped

A few cm of drift after 15m travelled

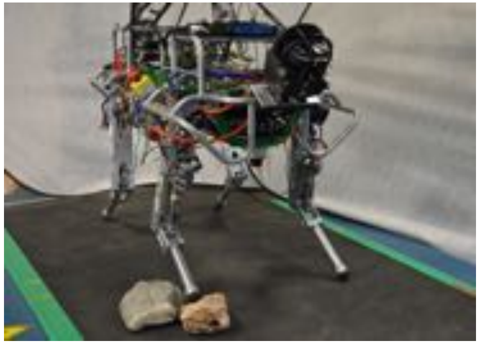
Used in DARPA Robotics Challenge



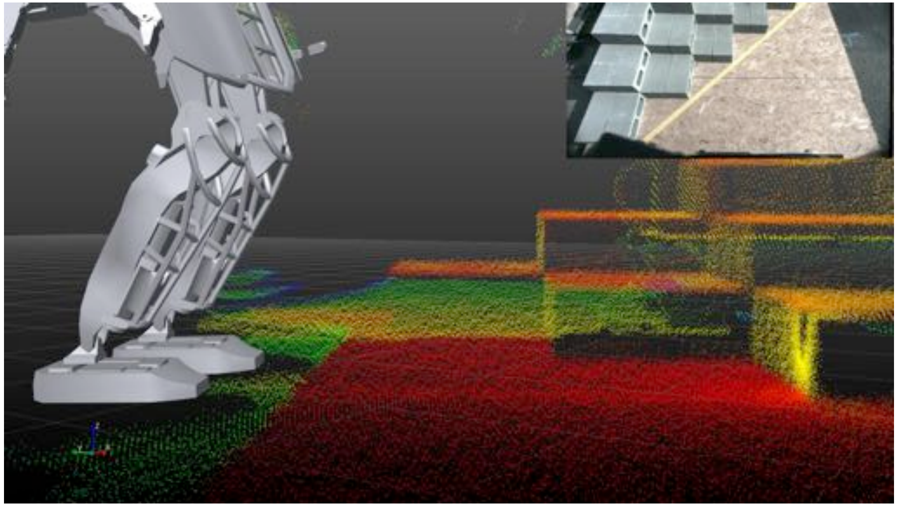
With IIT HyQ Quadruped

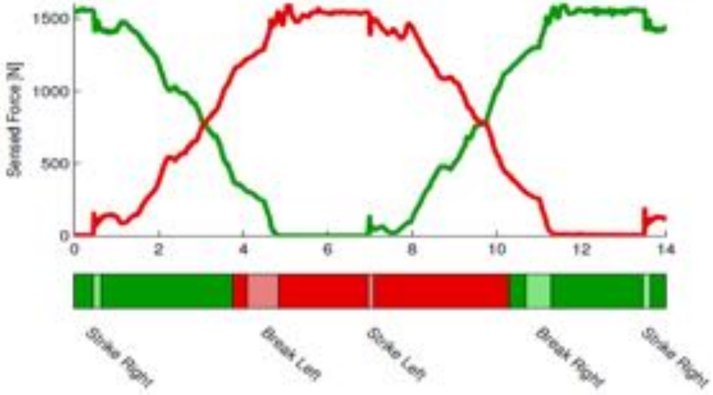
PhD Research of Marco Camurri

ISTITUTO ITALIANO DI TECNOLOGIA

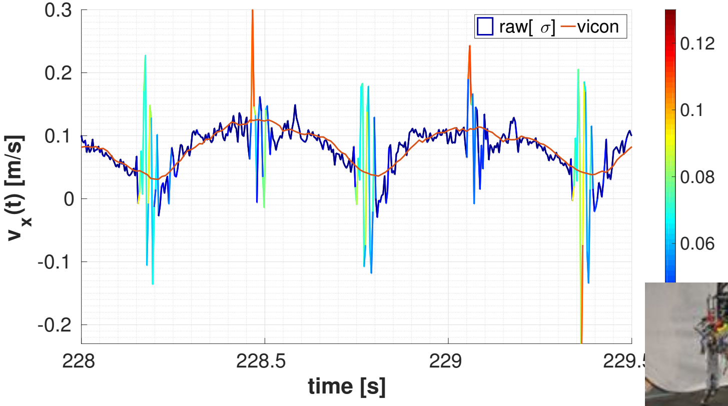


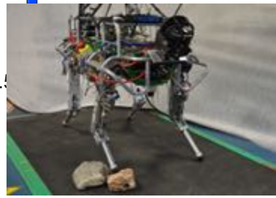
Kinematic State Estimation - Challenges





Biped: foot force sensing state model





Quadruped: spurious raw velocities

Kinematic State Estimation - Overview

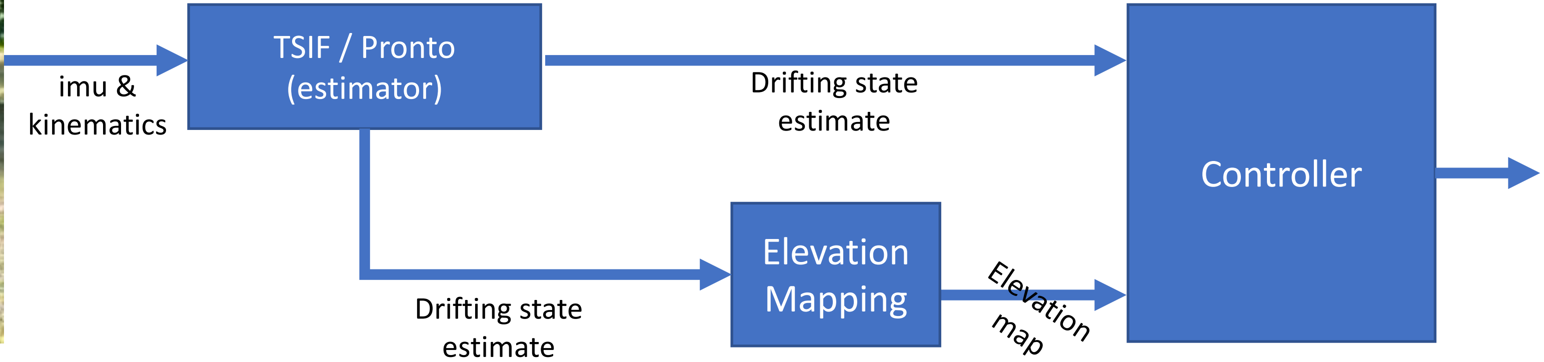
UNIVERSITY OF OXFORD

Goal: To estimate the **state of the robot's base**
- Position, orientation and velocity
- High frequency (250-400 Hz)
- Low latency (2-3 msec)

Main Challenges:
- Despite modelling, behaviour in contact is unknown
- **Latency is unacceptable**: estimate used in control

Core Approach: (Extended) Kalman Filter
- IMU-driven process model
- Leg kinematics used to measure **linear velocity**
- **Crucial:** contact classification using forces
- **No external sensors (cameras, LIDAR)**

Implementations:
- Anybotics/RSL: TSIF
- Oxford: Pronto



Aim to combine Leg Odometry and VIO

Legged Robot State Estimation

- Typically fused with Kalman Filters:
  - TSIF [Bloesch] - RSL / ANYbotics
  - Pronto [Nobili] – MIT / Oxford
- Typically uses high grade sensors.

Visual-Inertial Odometry

- Mature field
- Good results on datasets
- Sliding window optimisation popular (e.g. ROVIO, VINS-Mono)

Bloesch et. al: State Estimation for Legged Robots – Consistent Fusion of Leg Kinematics and IMU. RSS 2013
Nobili et. al: Heterogeneous Sensor Fusion for Accurate State Estimation of Dynamic Legged Robots. RSS 2017

Visual Inertial Legged Navigation System - Contributions

- VILENS - First algorithm to tightly fuse vision, IMU, and leg odometry
- Extensive Testing with 250+ m experiments
- Consumer Grade Cameras (Intel RealSense)

Previous Systems:

- legged odometry centred
  - with optional vision (or LIDAR)

Proposed System:

- vision centred
  - with optional leg odometry

Direction of Progress: COTS depth/stereo cameras

UNIVERSITY OF OXFORD

ANYbotics ANYmal Version C

Boston Dynamics Spot

VILENS (Visual Inertial Legged Navigation System)

- IMU (400Hz) $\mathcal{I}_{\Delta i}$
- Camera (30Hz) $\mathcal{C}_i$
- Kinematics (400Hz) $\mathcal{Q}_{\Delta i}$

VILENS

Factor Graph Optimization

- State estimate

Control,
Mapping,
Navigation,
…

$$\boldsymbol{x}_i \triangleq [\mathbf{R}_i, \mathbf{p}_i, \mathbf{v}_i, \mathbf{b}_i]$$

VILENS (Visual Inertial Legged Navigation System)

UNIVERSITY OF OXFORD

- IMU (400Hz) $\mathcal{I}_{\Delta i}$
- Camera (30Hz) $\mathcal{C}_i$
- Kinematics (400Hz) $\mathcal{Q}_{\Delta i}$

VILENS

Factor Graph Optimization

- State estimate

Control,
Mapping,
Navigation,
…

priors — imu — vision priors — vision — kinematics

$$\mathcal{X}^* = \arg\min_{\mathcal{X}} \|\mathbf{r}_0\|^2_{\Sigma_0} + \sum_{i \in \mathsf{K}_k} \|\mathbf{r}_{\mathcal{I}_{\Delta i}}\|^2_{\Sigma_{\mathcal{I}_{\Delta i}}} + \sum_{j \in M} \|\mathbf{r}_{\mathbf{m}_{j,0}}\|^2_{\Sigma_{\mathbf{m}_{j,0}}} + \sum_{i \in \mathsf{K}_k} \sum_{j \in \mathsf{M}_i} \|\mathbf{r}_{\mathbf{m}_j}\|^2_{\Sigma_{\mathbf{m}_j}} + \sum_{i \in \mathsf{K}_k} \|\mathbf{r}_{\mathcal{Q}_{\Delta i}}\|^2_{\Sigma_{\mathcal{Q}_{\Delta i}}}$$

Factor Definitions

iSAM2 – using GTSAM:

- Sliding window batch optimization

Prior Factors:

- Set initial conditions

$$\mathbf{r}_0(\boldsymbol{x}_0, \mathcal{Z}) = \begin{pmatrix} \Phi(\mathbf{T}_0^{-1}\mathbf{T}_{p_0}) \\ \mathbf{v}_0 - \mathbf{v}_{p_0} \\ \mathbf{b}_0^a - \mathbf{b}_{p_0}^a \\ \mathbf{b}_0^\omega - \mathbf{b}_{p_0}^\omega \end{pmatrix}$$

Forster, Carlone, Dellaert, Scaramuzza. RSS 2015.
IMU Preintegration on Manifold for Efficient Visual-Inertial Maximum-a-Posteriori Estimation.

IMU Factors (Forster et al):

- Difference between IMU preintegration & estimate, w/ biases.

$$\mathbf{r}_{\mathcal{I}_{\Delta i}} = \left[\mathbf{r}^{\mathsf{T}}_{\Delta \mathbf{R}_{\Delta i}}, \mathbf{r}^{\mathsf{T}}_{\Delta \mathbf{p}_{\Delta i}}, \mathbf{r}^{\mathsf{T}}_{\Delta \mathbf{v}_{\Delta i}}, \mathbf{r}^{\mathsf{T}}_{\Delta \mathbf{b}_{\Delta i}}\right]$$

Leg Odometry Factors:

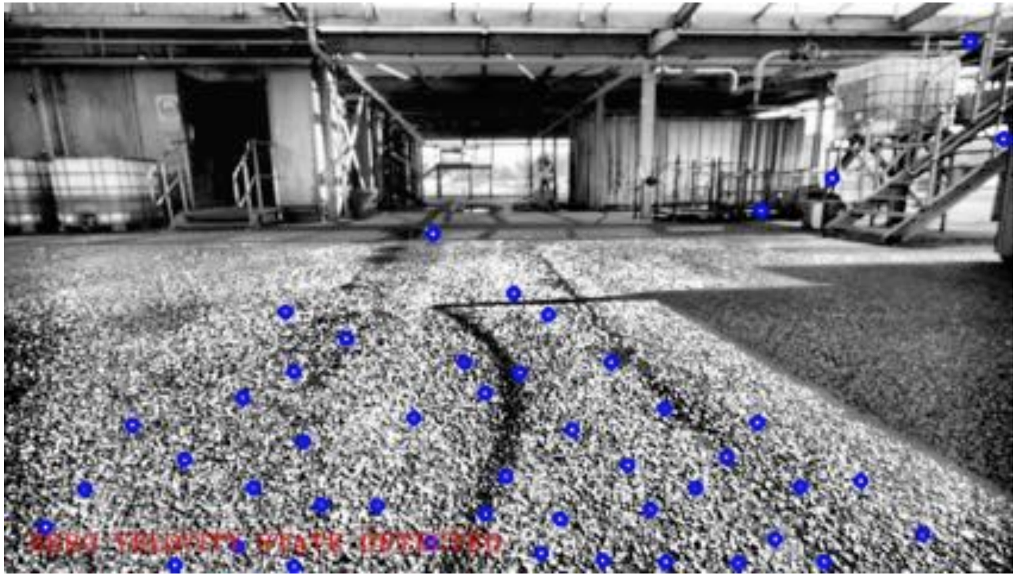
- Assuming the contact points are fixed, estimate the relative motion.
- Formulate output of existing kinematic-inertial TSIF estimator (Bloesch et al.) as a relative pose constraint.

$$\mathbf{r}_{\mathcal{Q}_{\Delta i}} = \Phi\left((\mathbf{T}_{i-1}^{-1}\mathbf{T}_i)^{-1}\widetilde{\mathbf{T}}_{i-1}^{-1}\widetilde{\mathbf{T}}_i\right)$$

Vision Cost Functions

Feature Tracking Front-End:

- Track feature through successive frames (KLT feature tracker).



Vision Factors:

- Estimate 3D location of landmarks.
- Minimise reprojection error between estimate and measured:

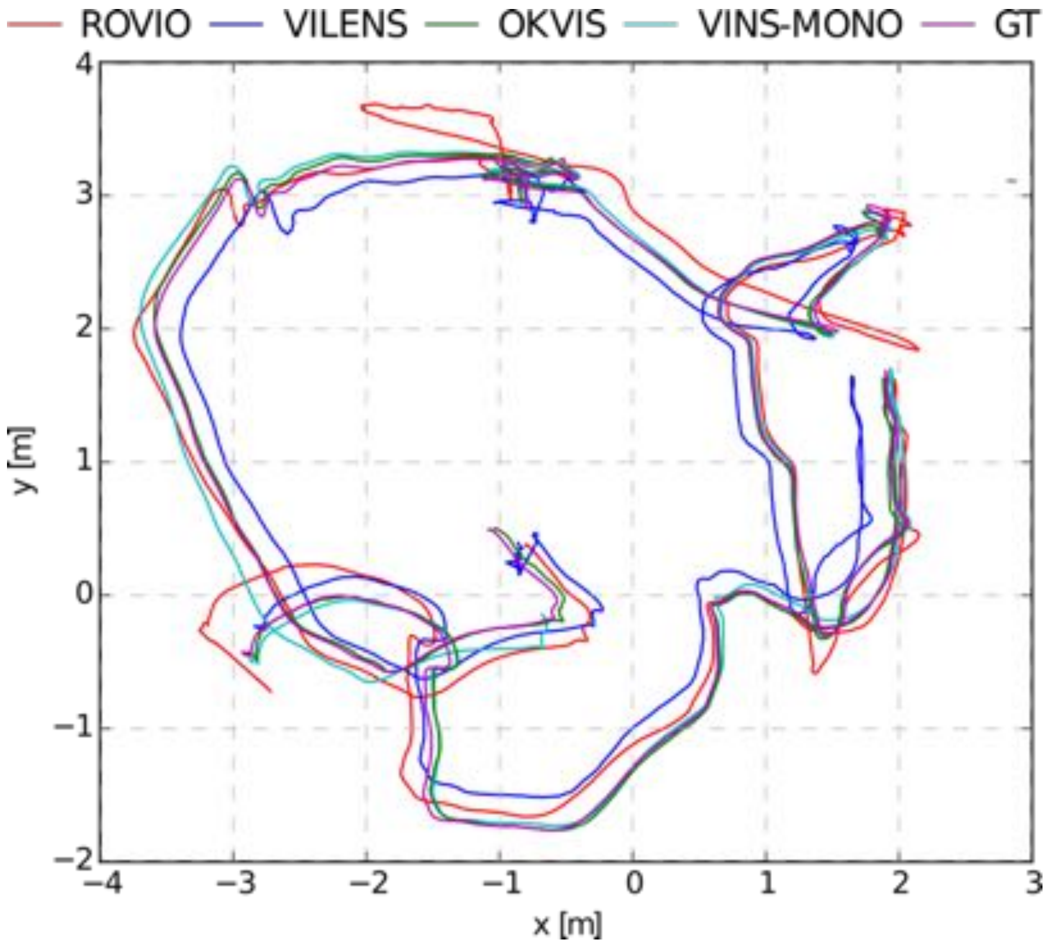
$$\mathbf{r}_{\mathbf{m}_j} = \begin{pmatrix} \pi_u(\mathbf{R}_i, \mathbf{p}_i, \mathbf{m}_j) - u_{i,j} \\ \pi_v(\mathbf{R}_i, \mathbf{p}_i, \mathbf{m}_j) - v_{i,j} \end{pmatrix}$$

- Add prior to help under-constrained landmarks:

$$\mathbf{r}_{\mathbf{m}_{j,0}} = \mathbf{m}_j - \mathbf{m}_{j,0}$$

Experimental Results - EUROC

- Standard visual-inertial odometry dataset.
- Qualitative results, demonstrating the system can function as a stand-alone VINS system.
- Comparable performance to state of the art VINS systems.



Experimental Setup - ANYmal

Oil Rig Training Site: Realistic Industrial Mock-up

250m of continuous walking. Brightness variation. Climbing, Trotting

Experimental Setup – Ground Truth

UNIVERSITY OF OXFORD

Leica TS-16 - Ground Truth Tracking

Oil Rig dataset results

Robust Legged Robot State Estimation Using Factor Graph Optimization

David Wisth, Marco Camurri, Maurice Fallon

Oxford Robotics Institute – University of Oxford

RA-L SUBMISSION WITH IROS OPTION

Analysis + Discussion

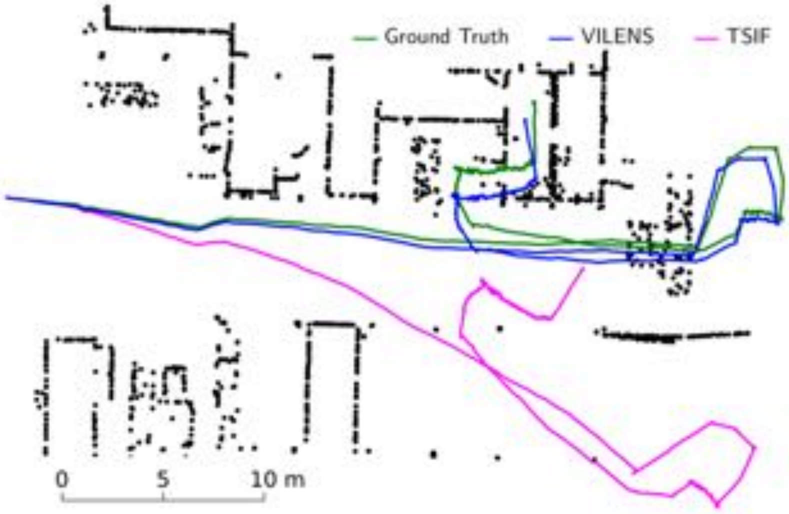
UNIVERSITY OF OXFORD

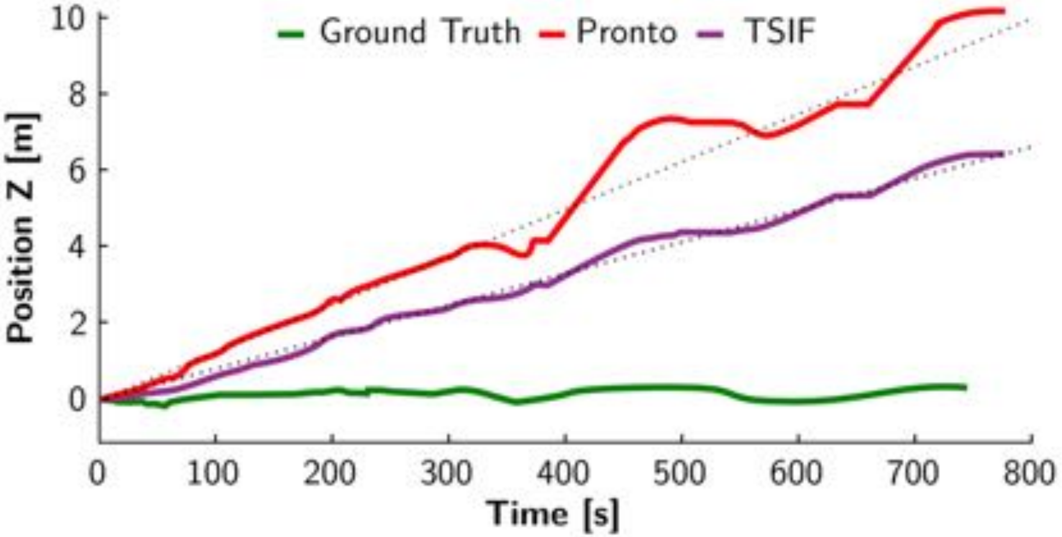
- We outperform the baseline kinematic-inertial estimator (TSIF)
  - 55% in RPE and 76% in ATE.
- Our algorithm operates even when off-the-shelf VIO fails.

TABLE II
MEAN (AND STANDARD DEVIATION) PERFORMANCE ON THE KEBLE COLLEGE AND OIL RIG DATASETS.

	RPE $\mu(\sigma)$ [m]		Yaw Error $\mu(\sigma)$ [deg]	
Dataset	**TSIF** [3]	**VILENS**	**TSIF** [3]	**VILENS**
Keble 1	0.53 (0.21)	**0.30** (0.12)	6.64 (2.23)	**0.99** (0.80)
Keble 2	0.51 (0.10)	**0.23** (0.10)	5.72 (0.94)	**1.47** (1.07)
Keble 3	0.67 (0.10)	**0.52** (0.15)	6.68 (0.80)	**3.86** (1.90)
Keble 4	0.47 (0.11)	**0.40** (0.10)	3.32 (1.15)	**1.13** (1.46)
Oil Rig	0.44 (0.37)	**0.41** (0.18)	4.89 (3.38)	**3.68** (4.10)

Leg Odometry Biases





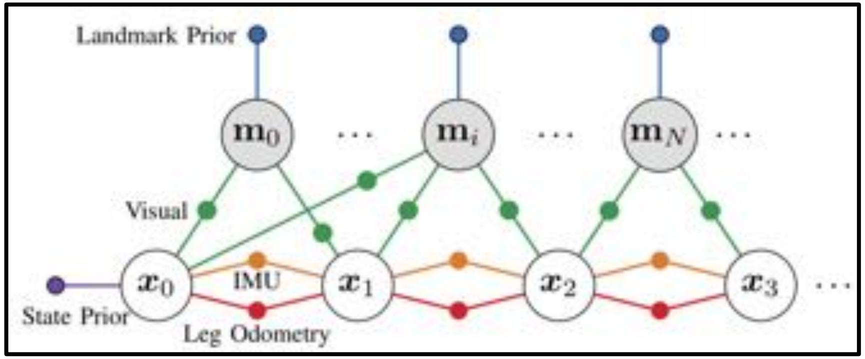
TSIF: bad yaw bias estimation
VILENS: good yaw bias estimation

TSIF: drift in height. Poorly observed

Visual Inertial Legged Navigation (VILENS) - 2020

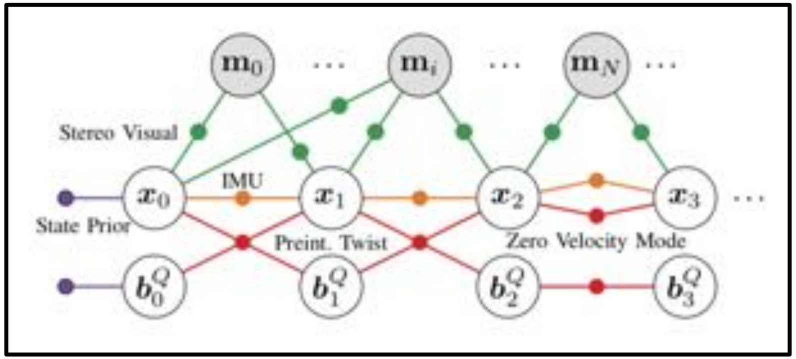
UNIVERSITY OF OXFORD

Previous Factor Graph (IROS 2019)



Visual features Leg position Inertial sensing

Current Factor Graph (Under Review)



$$\mathcal{X}_k^* = \arg\min_{\mathcal{X}_k} \|\mathbf{r}_0\|^2_{\Sigma_0} + \sum_{i\in \mathsf{K}_k} \|\mathbf{r}_{\mathcal{I}_{ij}}\|^2_{\Sigma_{\mathcal{I}_{ij}}} + \sum_{i\in \mathsf{K}_k}\|\mathbf{r}_{\mathcal{V}_{ij}}\|^2_{\Sigma_{\mathcal{V}_{ij}}}$$
$$+ \sum_{i\in \mathsf{K}_k}\|\mathbf{r}_{\mathbf{b}_{ij}}\|^2_{\Sigma_\mathbf{b}} + \sum_{i\in \mathsf{K}_k}\sum_{\ell\in \mathsf{M}_i}\|\mathbf{r}_{i,\mathbf{m}_\ell}\|^2_{\Sigma_\mathbf{m}}$$

$$\|\mathbf{r}_{\mathbf{b}_{ij}}\|^2_{\Sigma_\mathbf{b}} \triangleq \|\mathbf{b}_i^g - \mathbf{b}_{i-1}^g\|^2_{\Sigma_{\mathbf{b}^g}} + \|\mathbf{b}_i^a - \mathbf{b}_{i-1}^a\|^2_{\Sigma_{\mathbf{b}^a}} +$$
$$+ \|\mathbf{b}_i^\omega - \mathbf{b}_{i-1}^\omega\|^2_{\Sigma_{\mathbf{b}^\omega}} + \|\mathbf{b}_i^v - \mathbf{b}_{i-1}^v\|^2_{\Sigma_{\mathbf{b}^v}}$$

Gyro ang. velocity Accelerometer

Leg ang. velocity Leg linear velocity

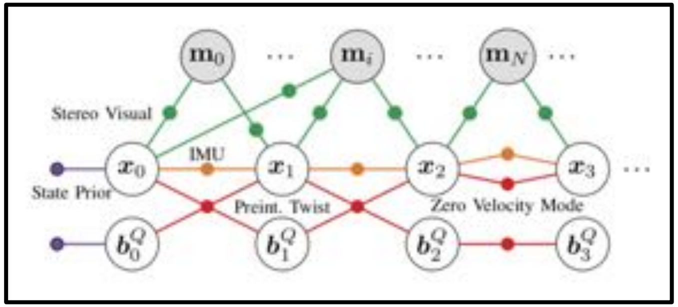
Visual Inertial Legged Navigation (VILENS) - 2020

Preintegrated Velocity Bias Estimation to Overcome Contact Nonlinearities in Legged Robot Odometry

David Wisth, Marco Camurri, Maurice Fallon

Oxford Robotics Institute - University of Oxford

Factor Graph Optimization



Visual features

Leg position

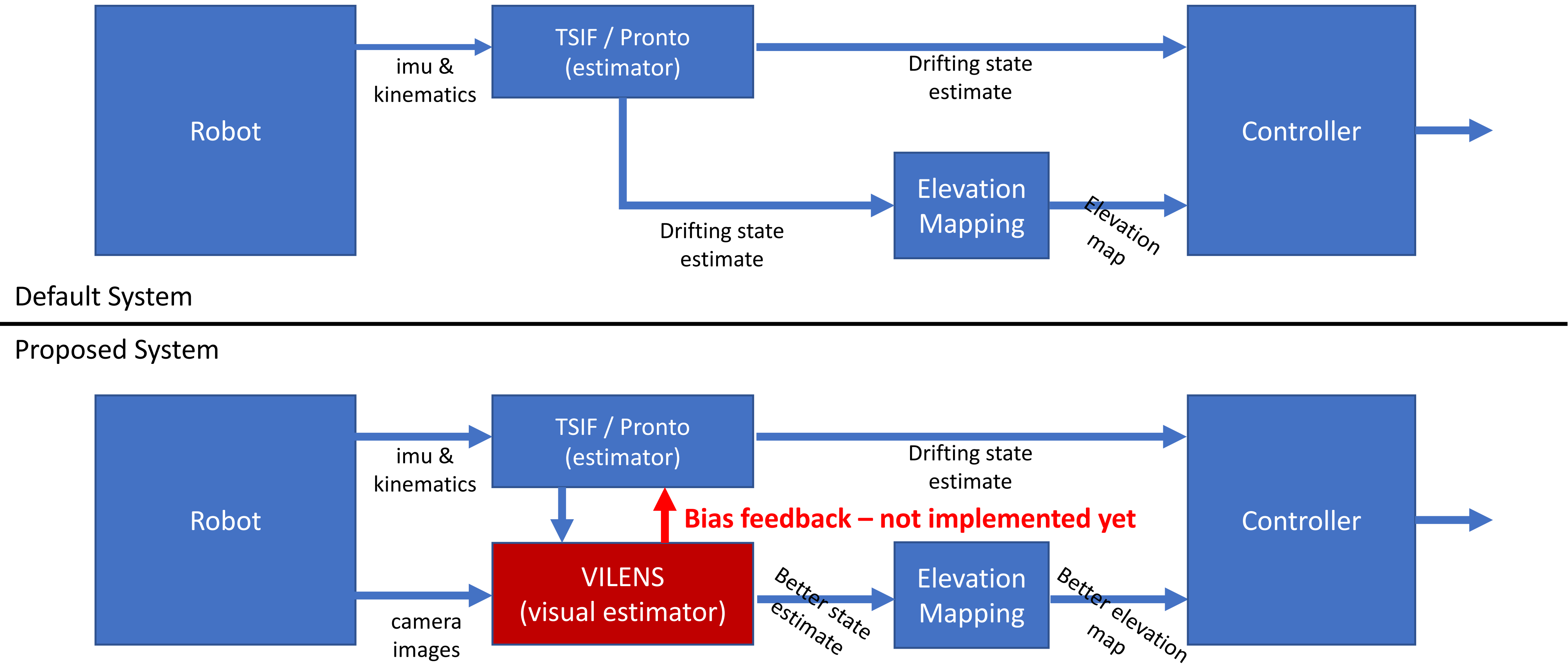
Inertial sensing

Tracker
(Ground Truth)

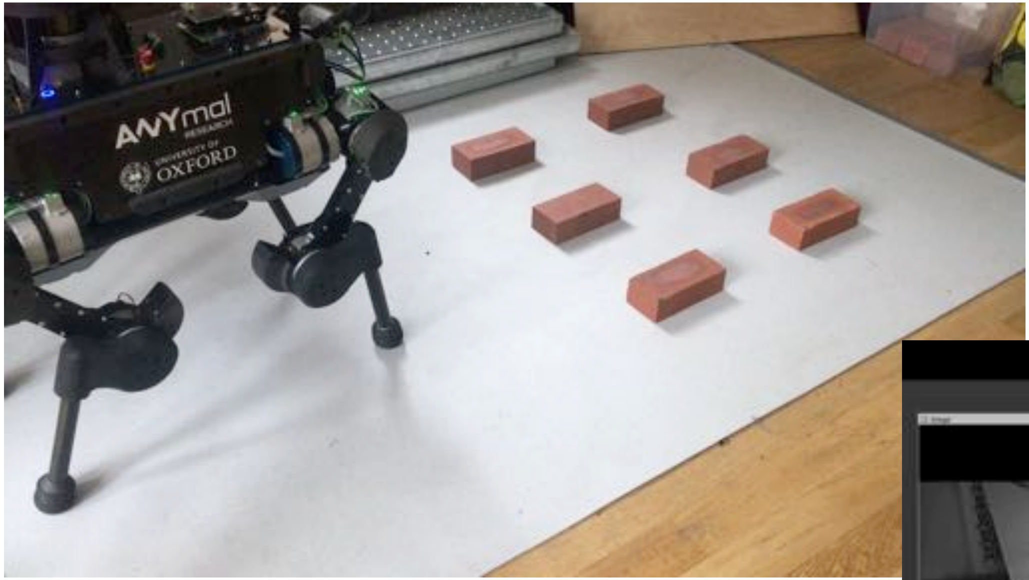
Intel RealSense Depth Camera
D435i Active Depth Camera + IMU

Preintegrated Velocity Bias Estimation to Overcome Contact Nonlinearities in Legged Robot Odometry
D. Wisth, M. Camurri, M. Fallon. Under Review. Available on Arxiv

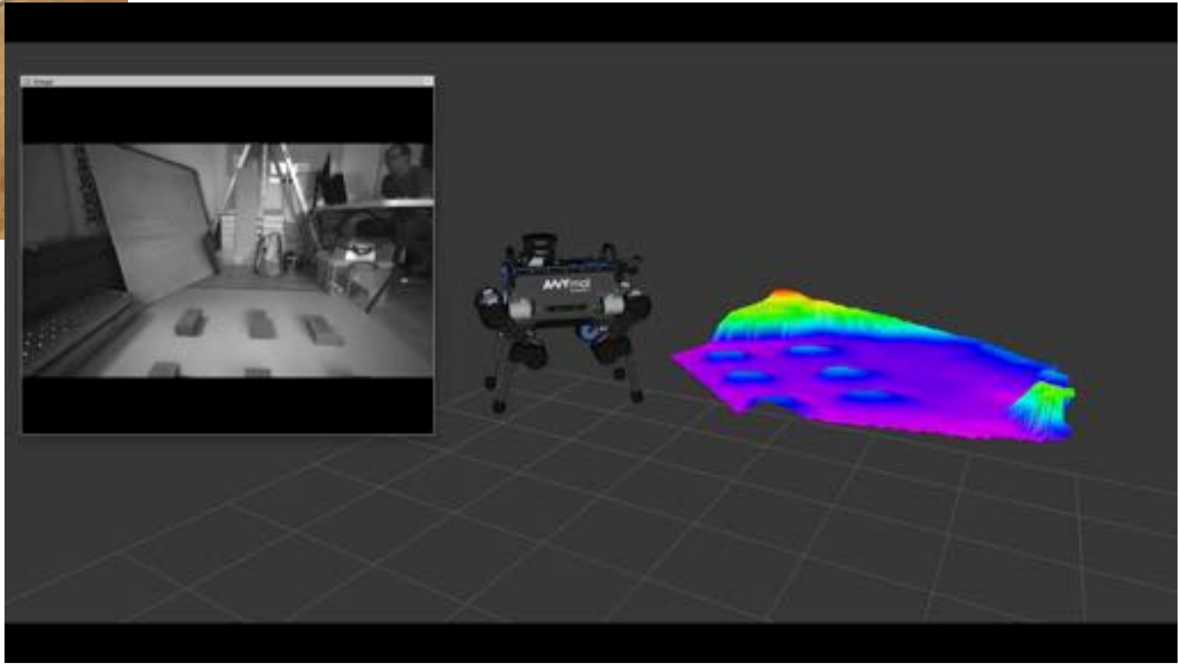
Integration into real-time system



Local Terrain Map Improvement



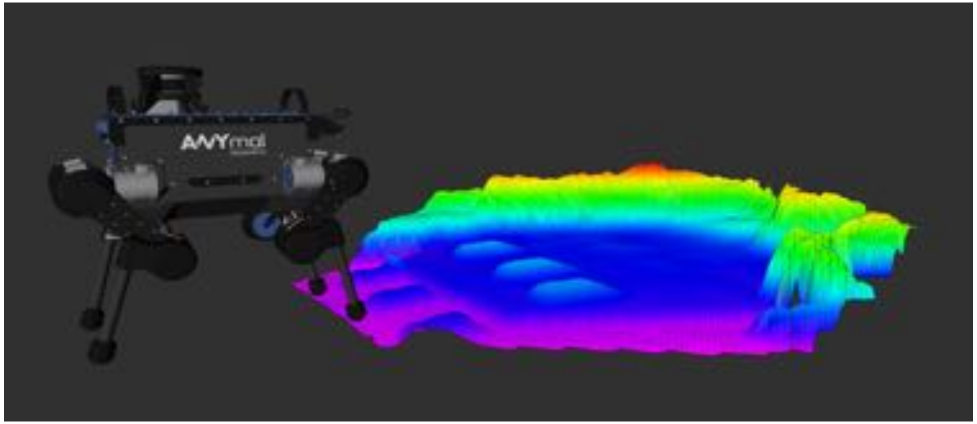
Using vision in estimation improves terrain map

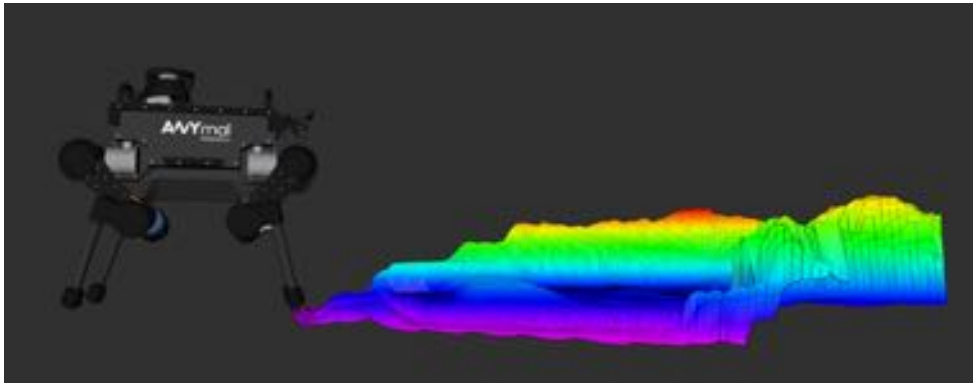


VILENS running live on ANYmal

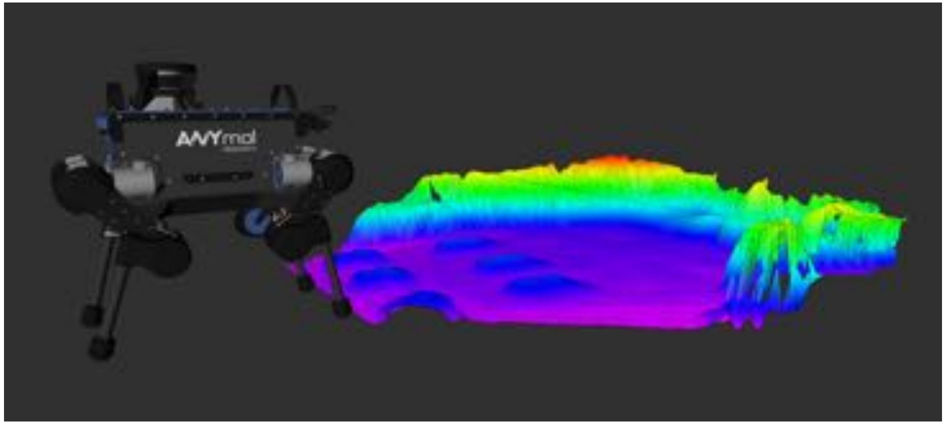
Local Terrain Map Improvement

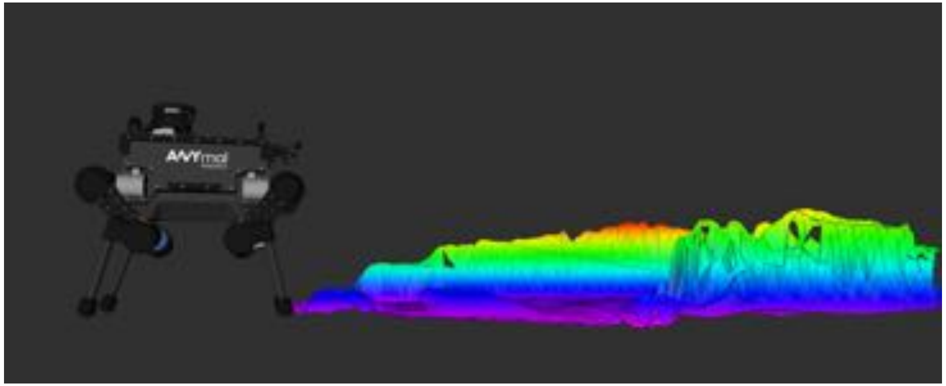
TSIF map is very inaccurate



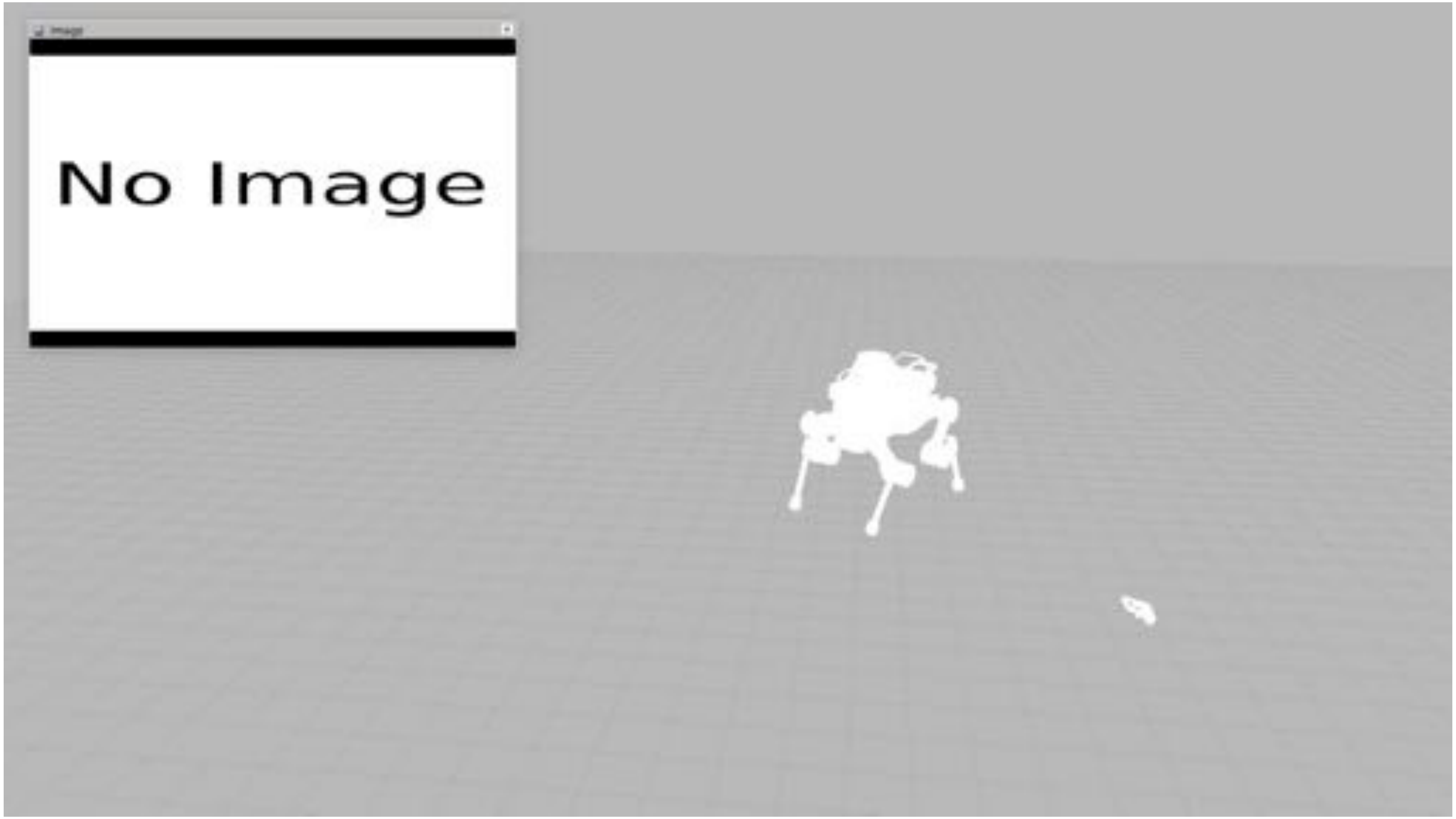


VILENS accurate reconstruction





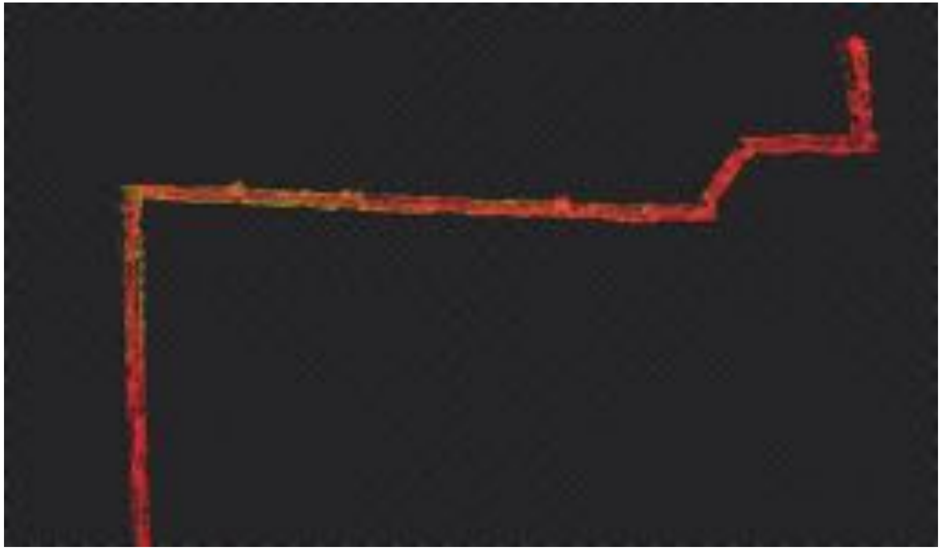
Local Terrain Map Improvement - Live

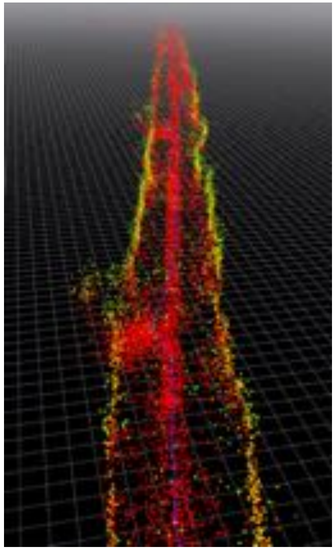


Work with RSL (ETH)

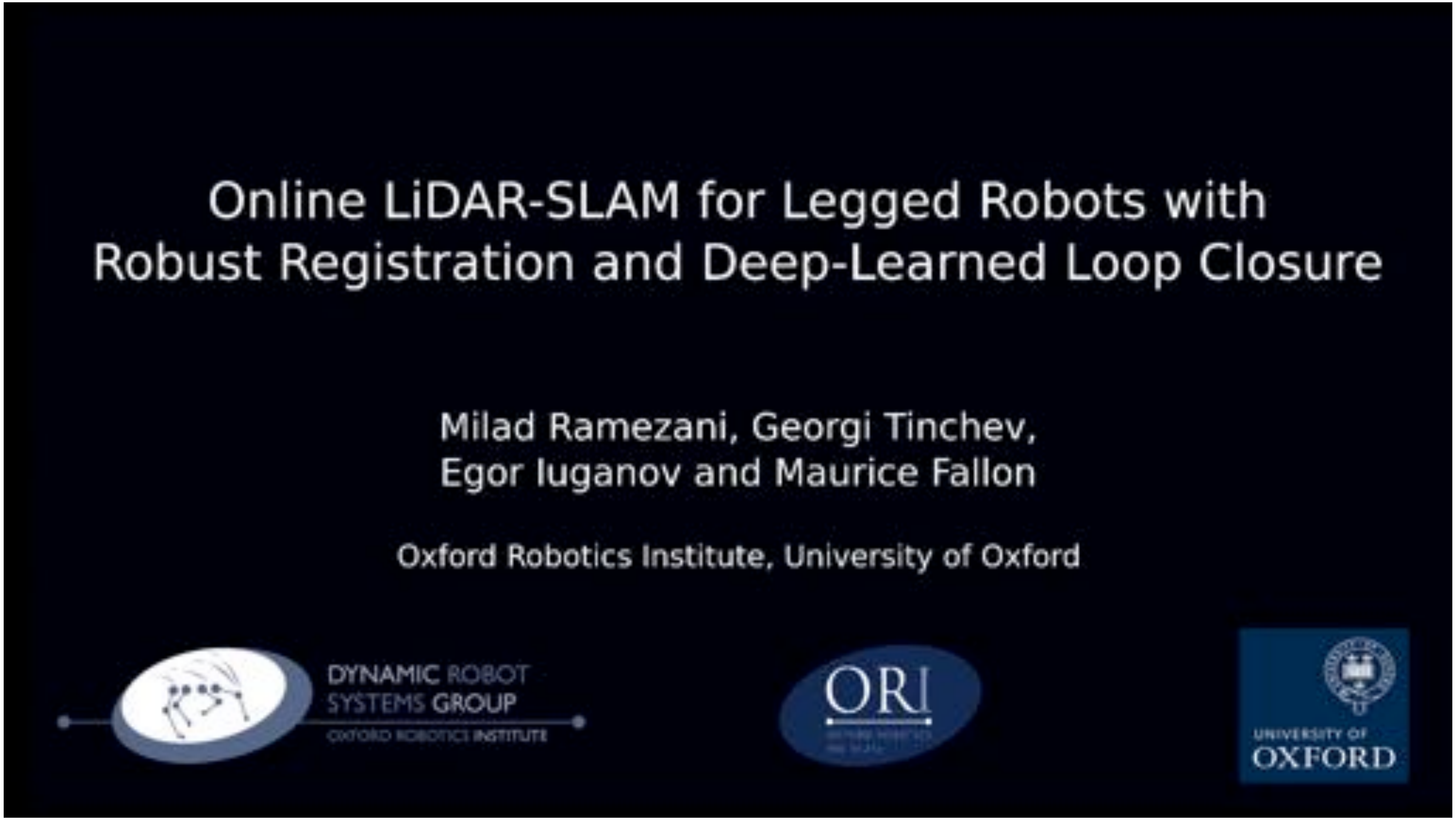
Implementation

- iSAM2 incremental optimizer (part of GTSAM library).
- Zero velocity states based on average feature movement.
- Works reliably with D435 on the robot at 15Hz
  - Exploring using RealSense T265
  - Considering mono-fisheye
- Multiple Sensor and Processing Message Threads
- Tested on three different copies of ANYmal:
  - Oxford
  - RSL (ETH)
  - ANYbotics
  - Contributing to ANYmal's SLAM in DARPA SubT Cerberus





SLAM or Loop Closure?



LIDAR Pose-Graph SLAM with Deep Learning Loop Closure

Thank You

UNIVERSITY OF OXFORD

For further information:

David Wisth

davidw@robots.ox.ac.uk

Data with ground truth:

ori.ox.ac.uk/vilens

Thanks to:

RSL (ETH) & ANYbotics

DYNAMIC ROBOT SYSTEMS GROUP
OXFORD ROBOTICS INSTITUTE

ori.ox.ac.uk/drs