Monocular Visual Teach and Repeat Aided by Local Ground Planarity

Lee Clement, Jonathan Kelly, and Timothy D. Barfoot

FSR 2015, Toronto, Canada

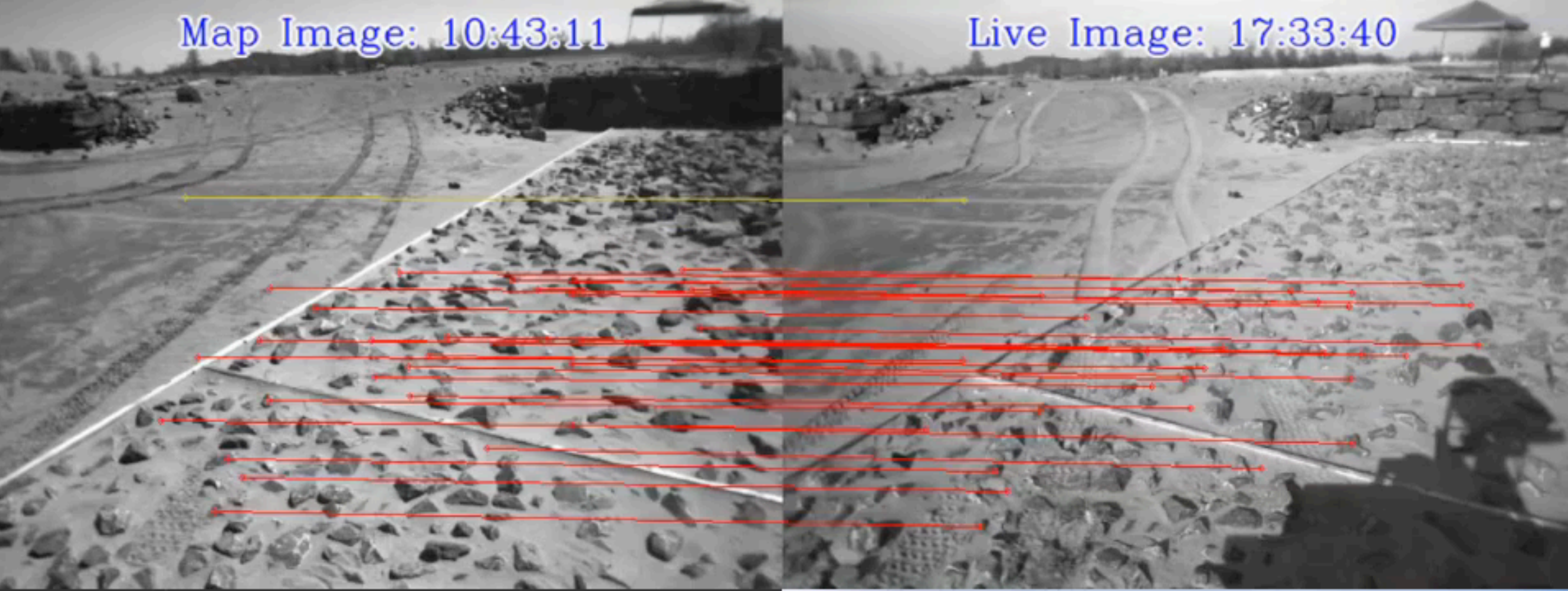
Institute for Aerospace Studies
UNIVERSITY OF TORONTO

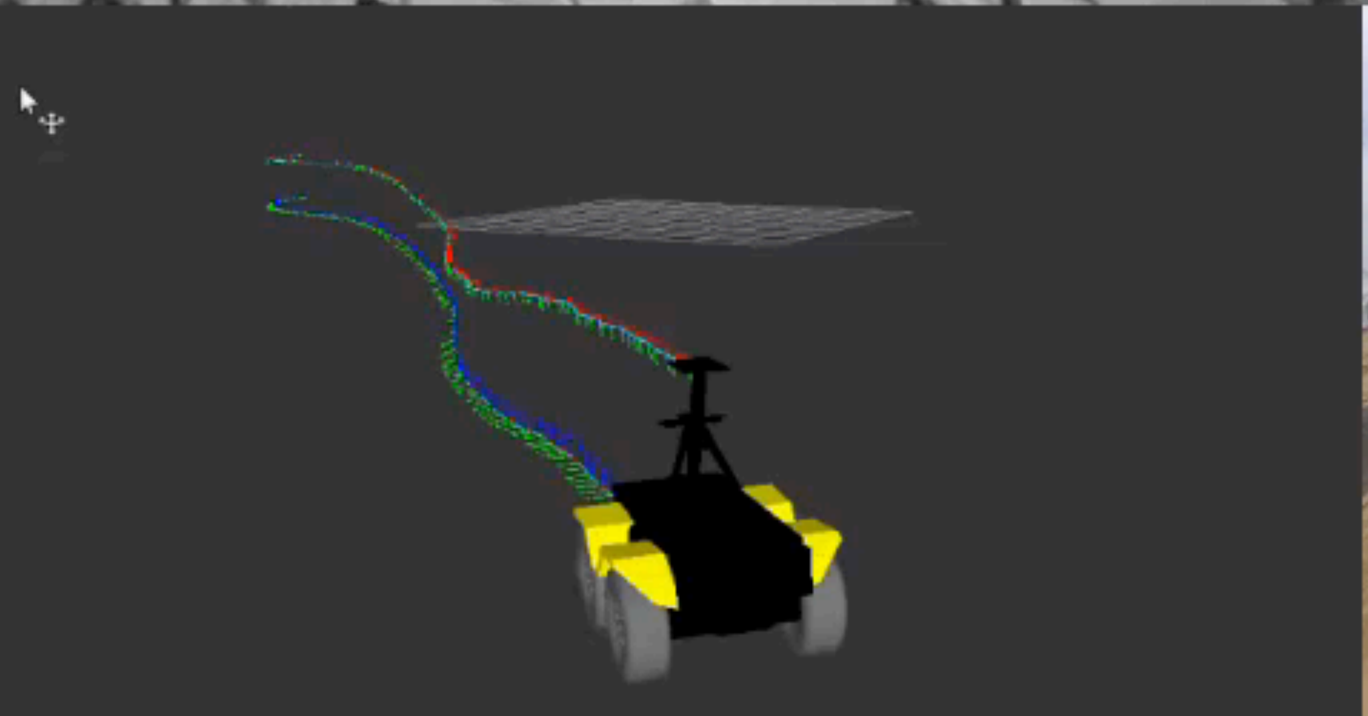
STARS LABORATORY

ASRL

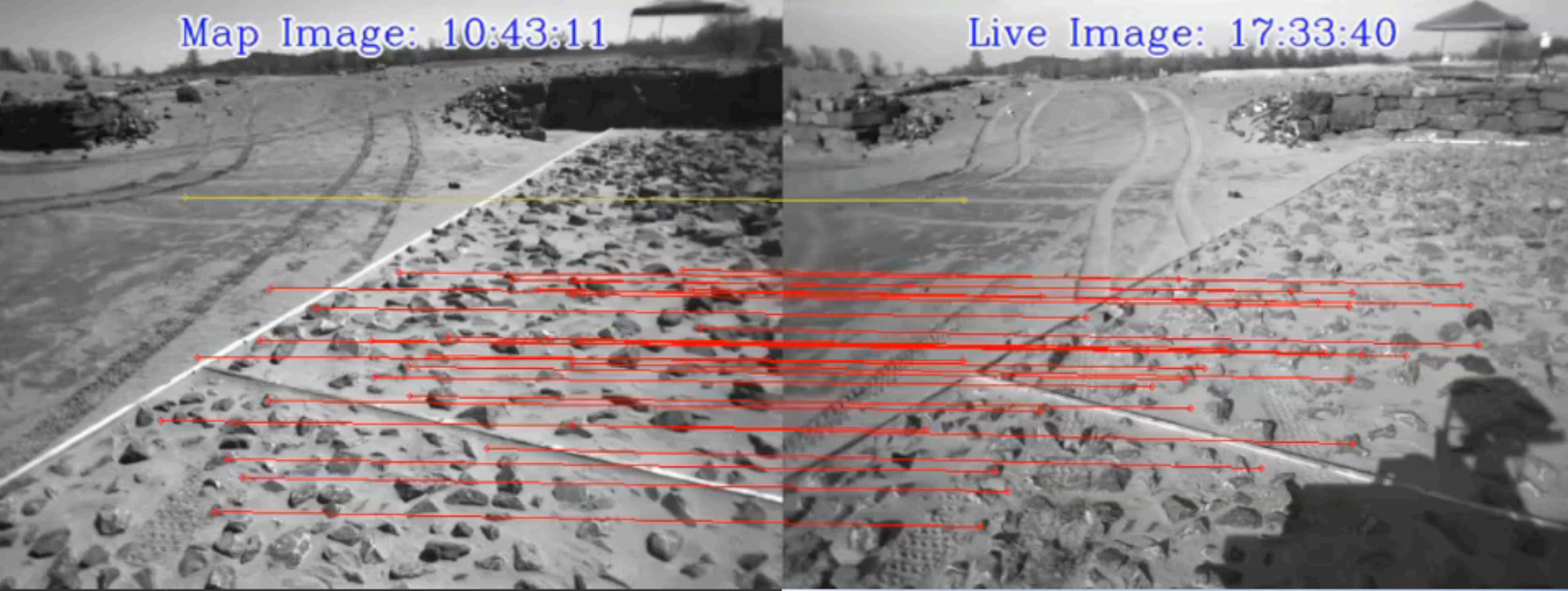
Motivation: Autonomous Navigation for Monocular Robots

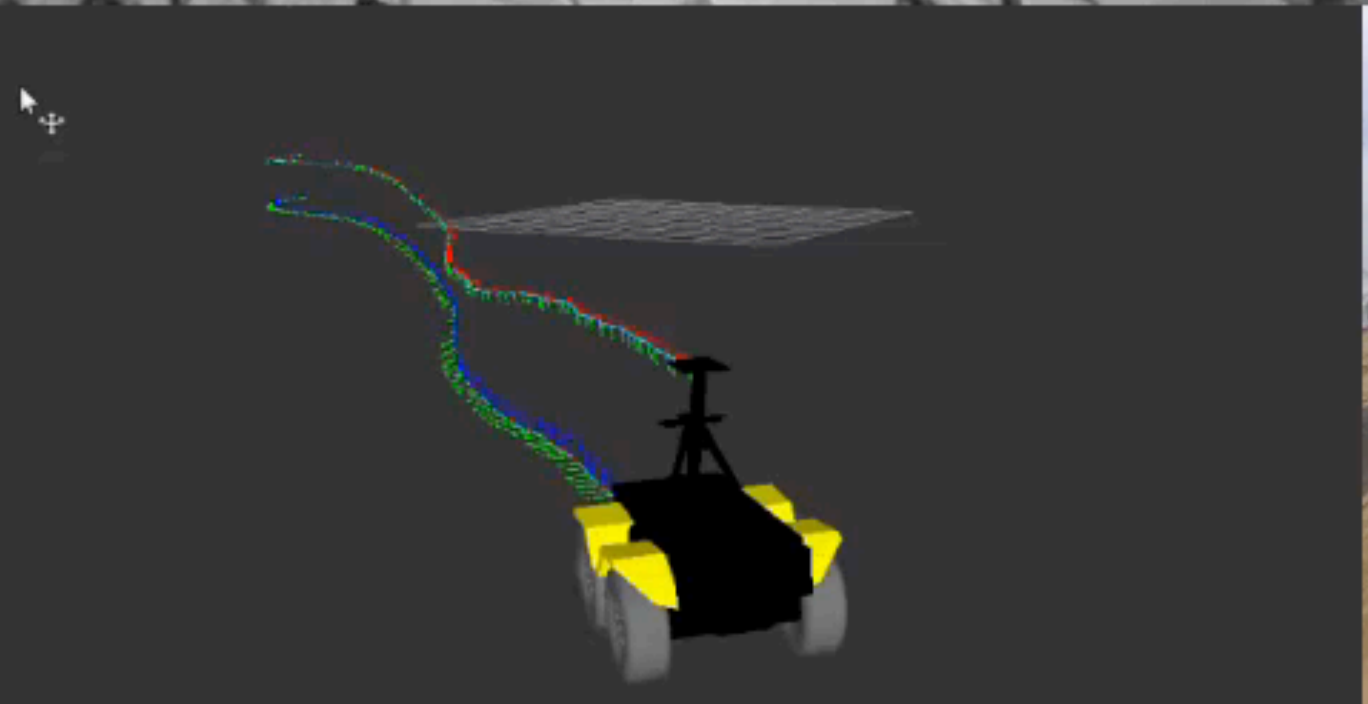
Many robots with **monocular cameras** need to perform **repetitive navigation** tasks — use **Visual Teach & Repeat**!





VT&R: Autonomous Vision-Based Route Following





VT&R: Autonomous Vision-Based Route Following

VT&R: The Basics

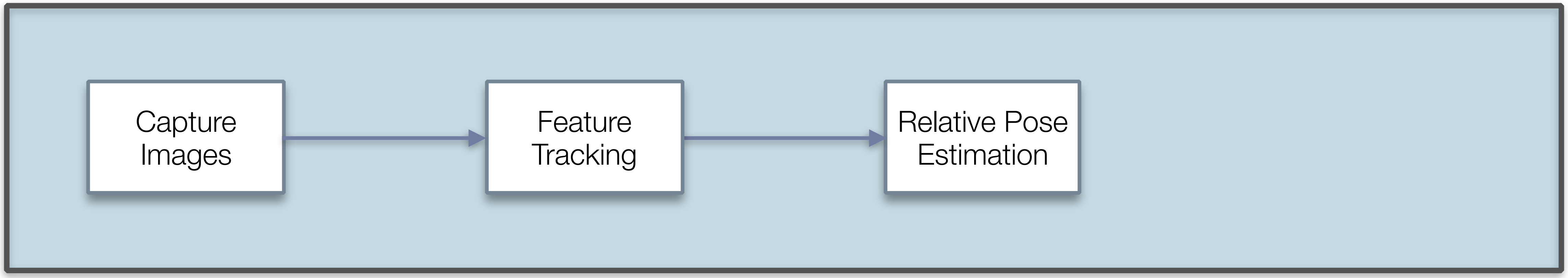
VT&R: The Basics

VT&R: The Basics

Teach Pass

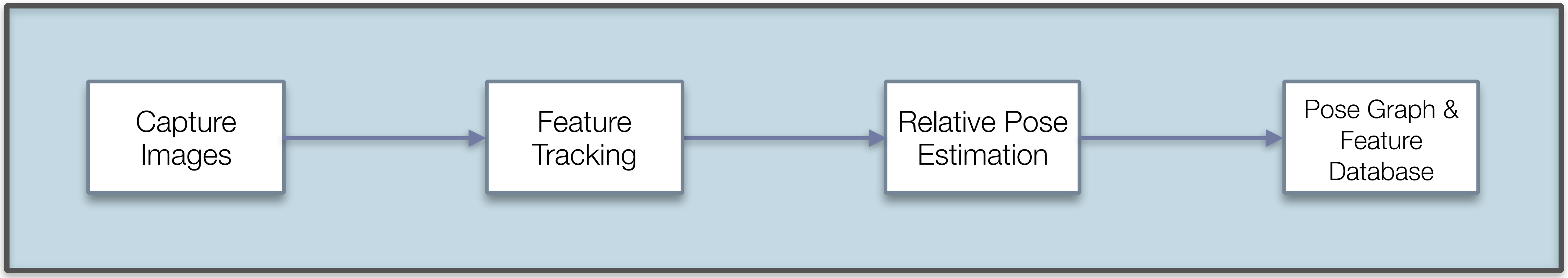
VT&R: The Basics

Teach Pass



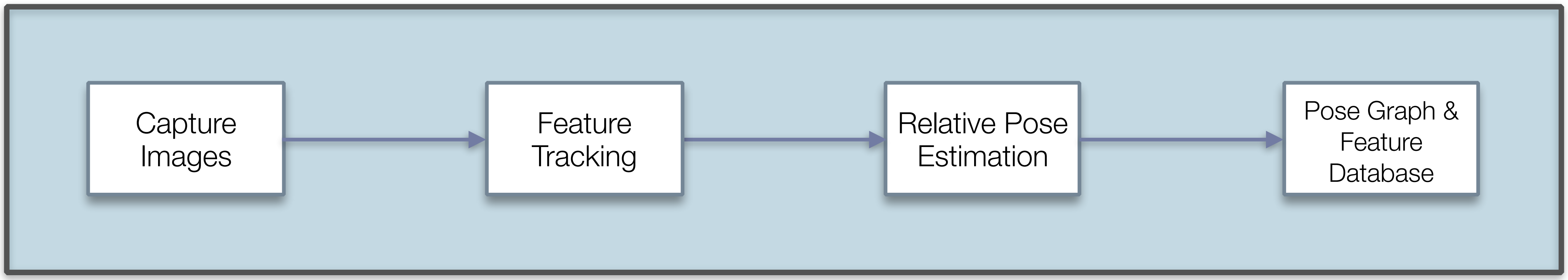
VT&R: The Basics

Teach Pass



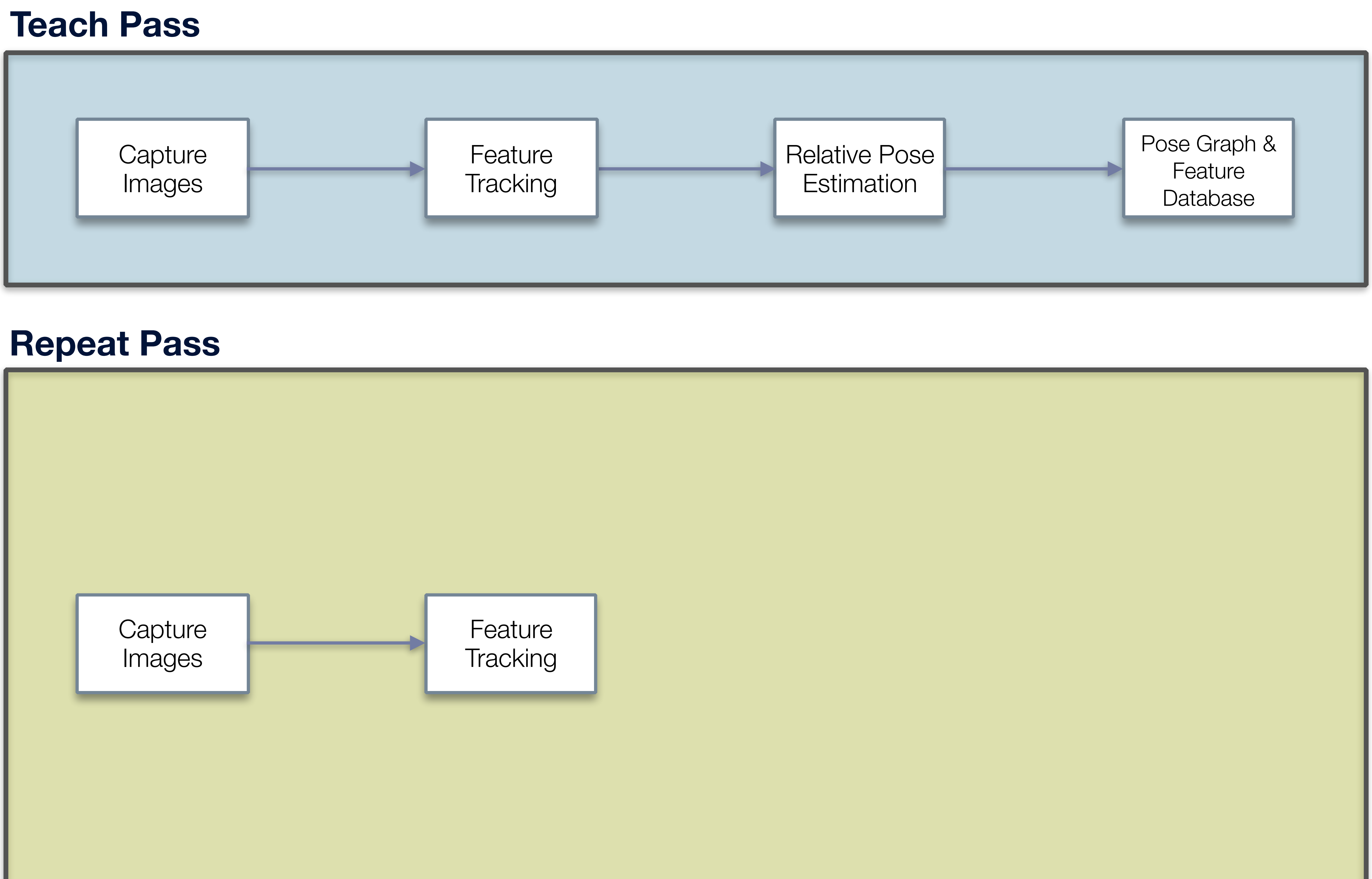
VT&R: The Basics

Teach Pass



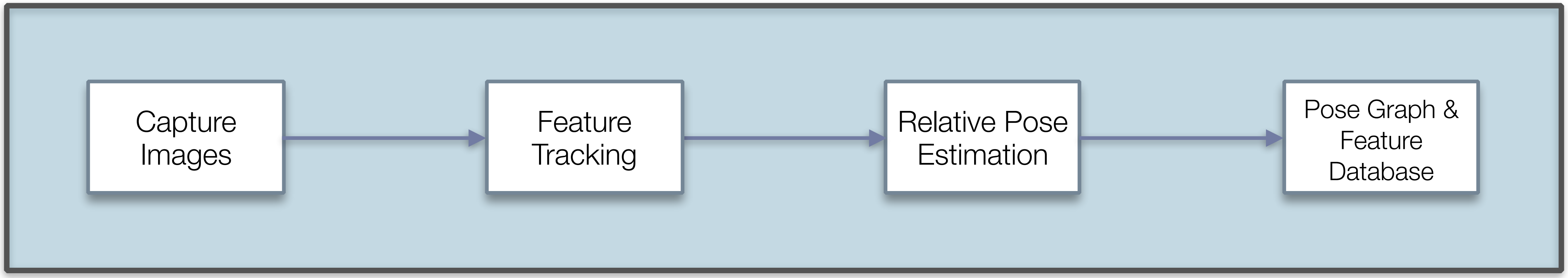
Repeat Pass

VT&R: The Basics

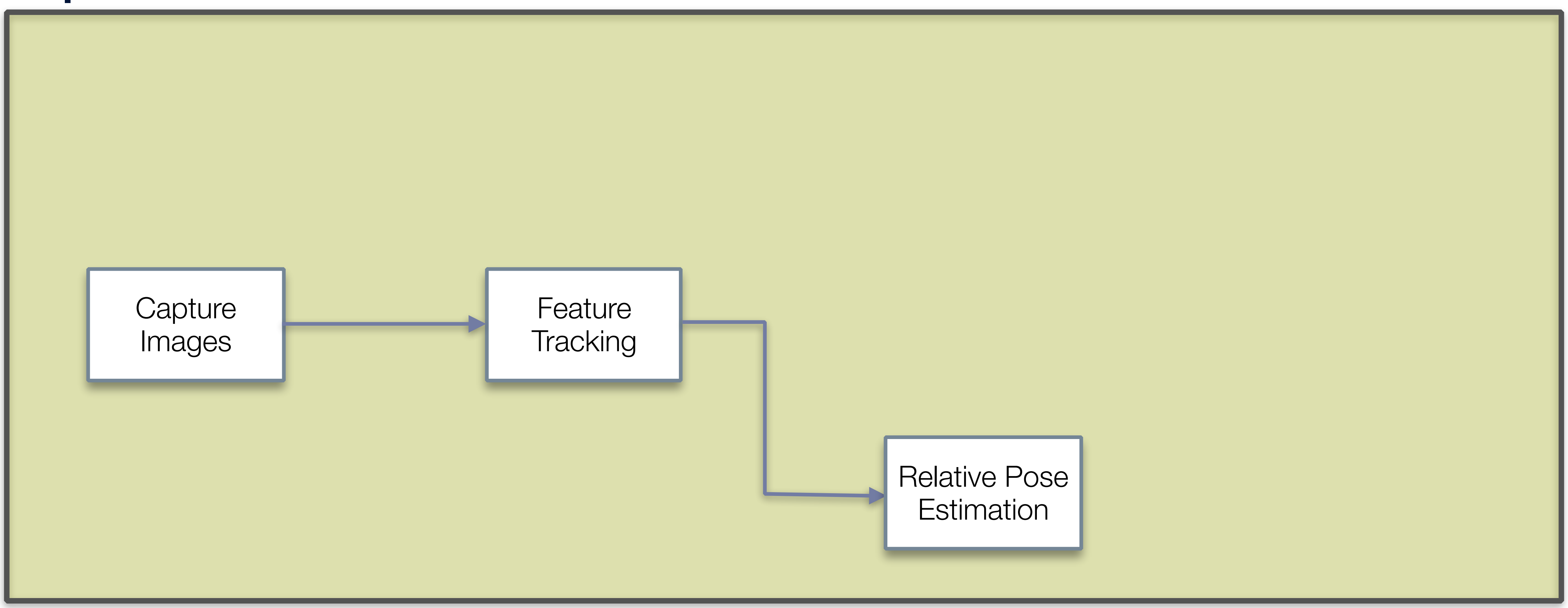


VT&R: The Basics

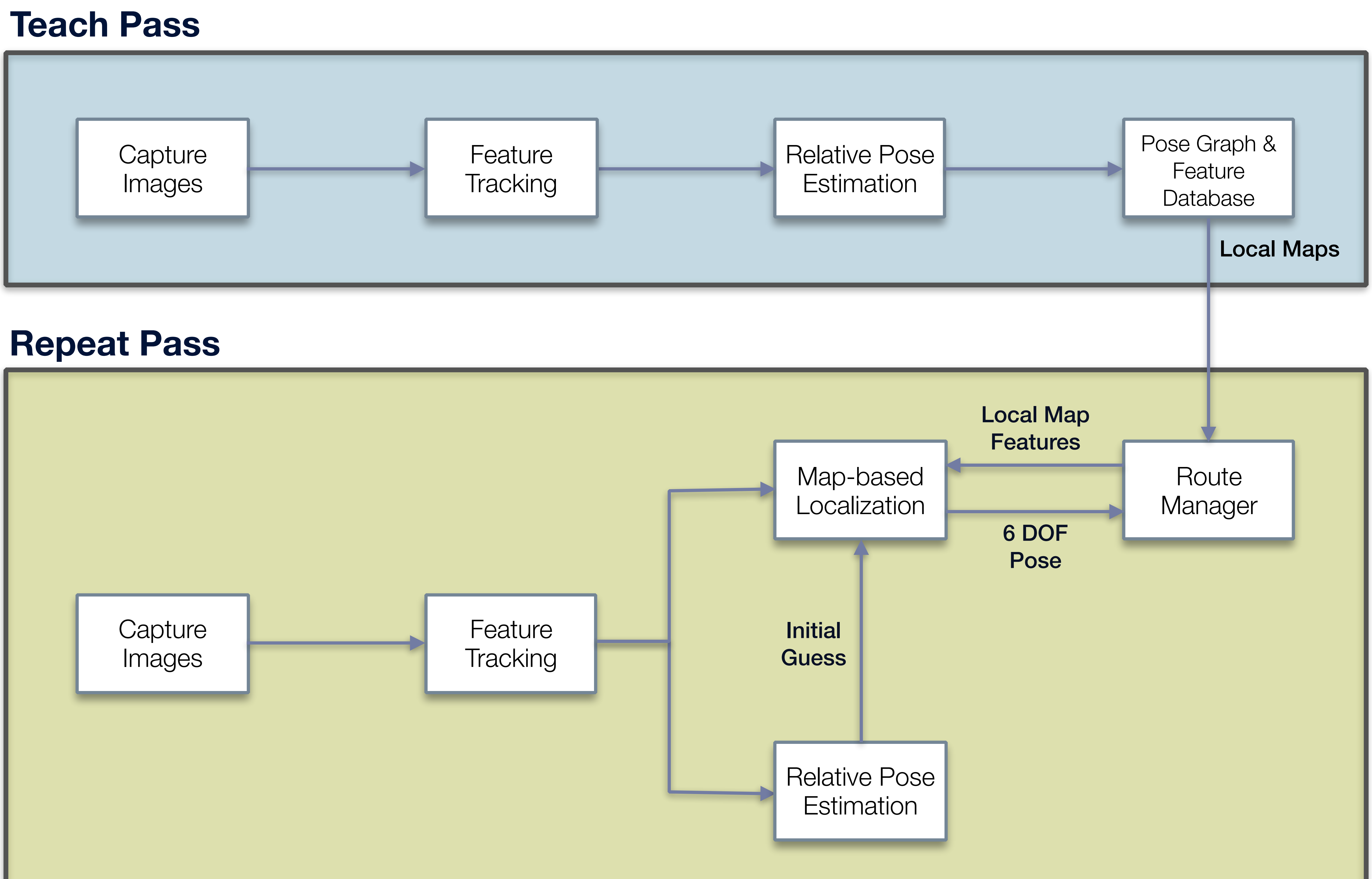
Teach Pass



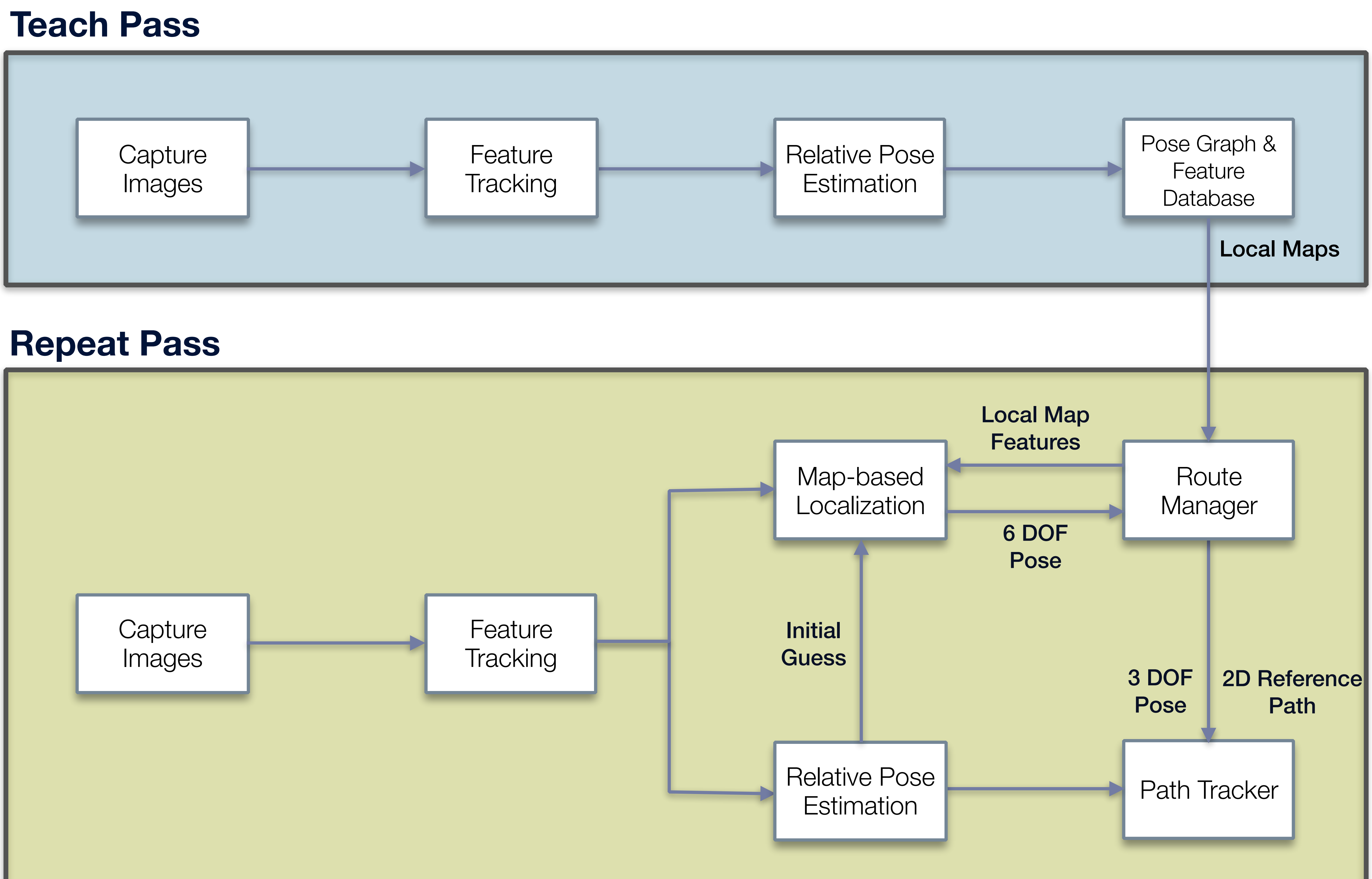
Repeat Pass



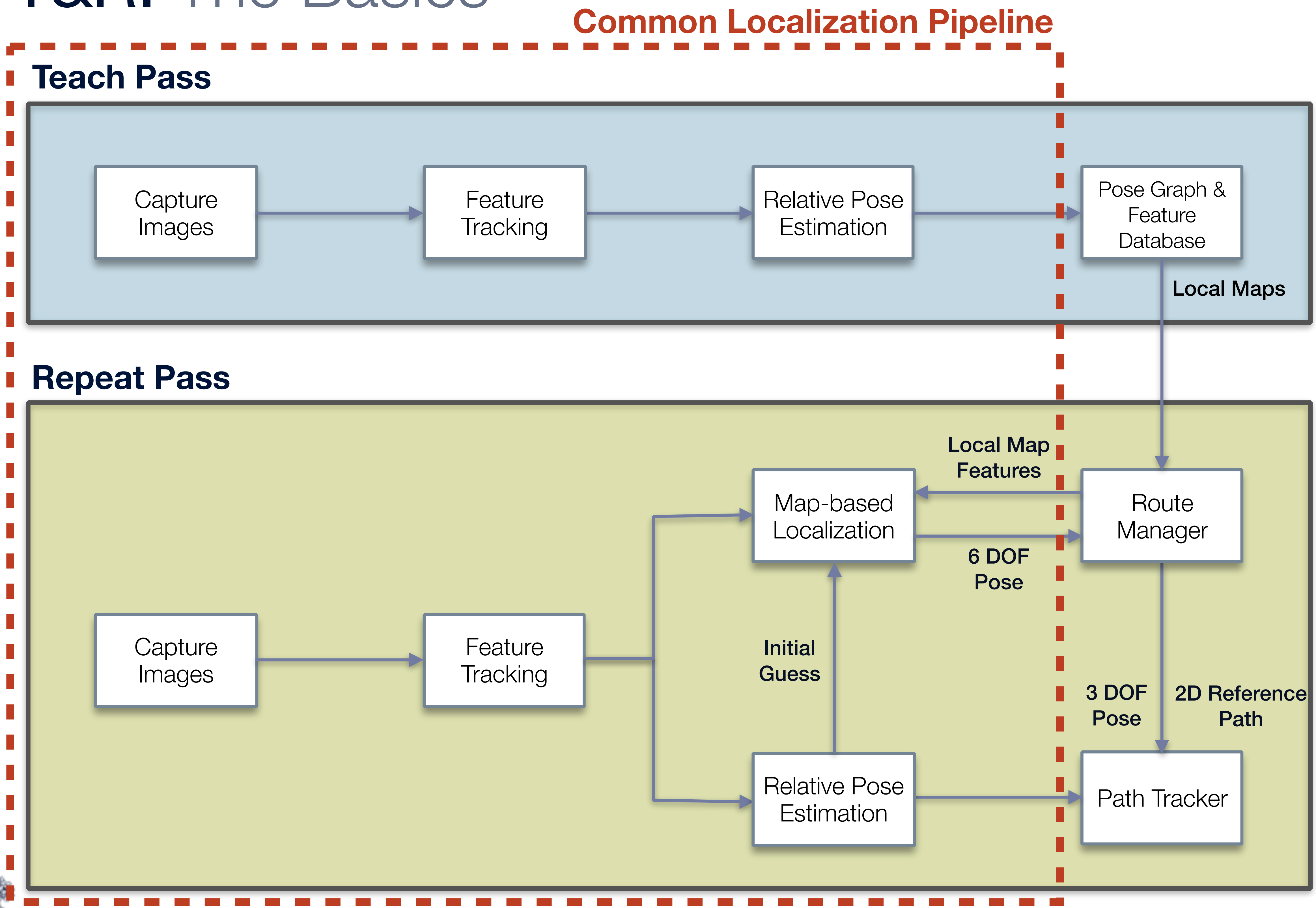
VT&R: The Basics



VT&R: The Basics



VT&R: The Basics



VT&R: Localization Pipeline

VT&R: Localization Pipeline

Stereo Pipeline
(Furgale & Barfoot, 2010)

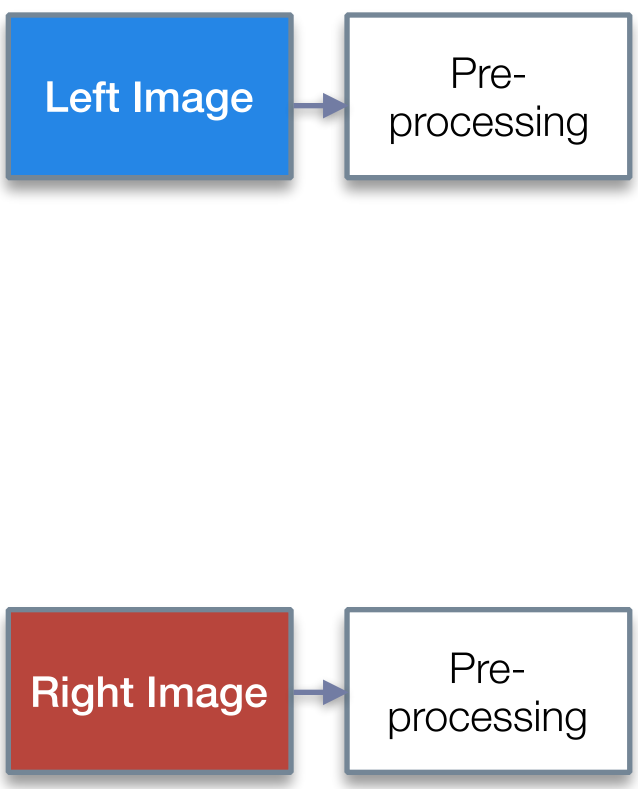
Left Image

Right Image

VT&R: Localization Pipeline

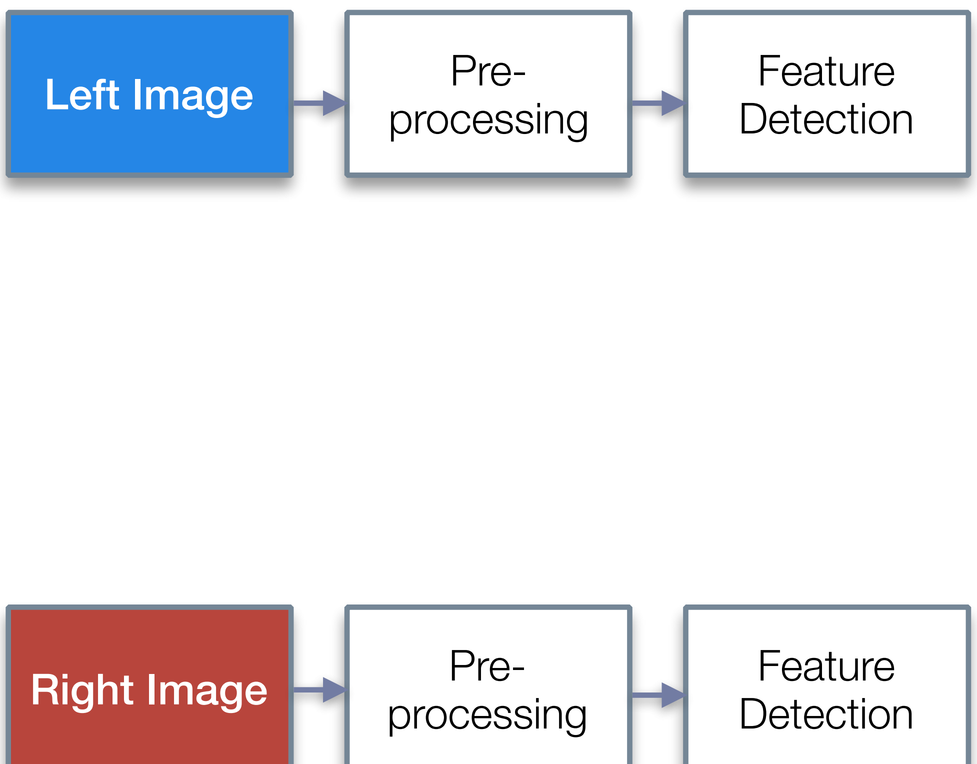
Stereo Pipeline

(Furgale & Barfoot, 2010)



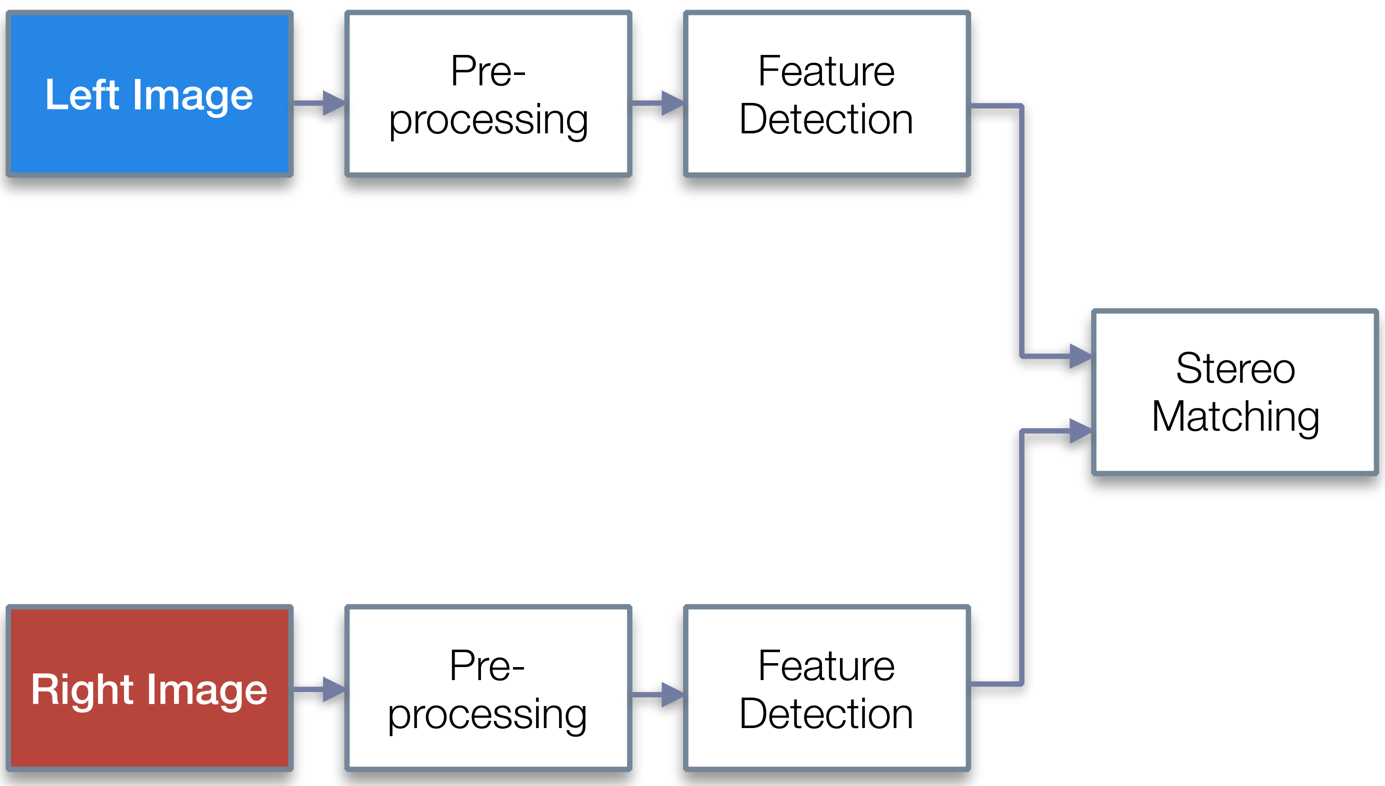
VT&R: Localization Pipeline

Stereo Pipeline
(Furgale & Barfoot, 2010)



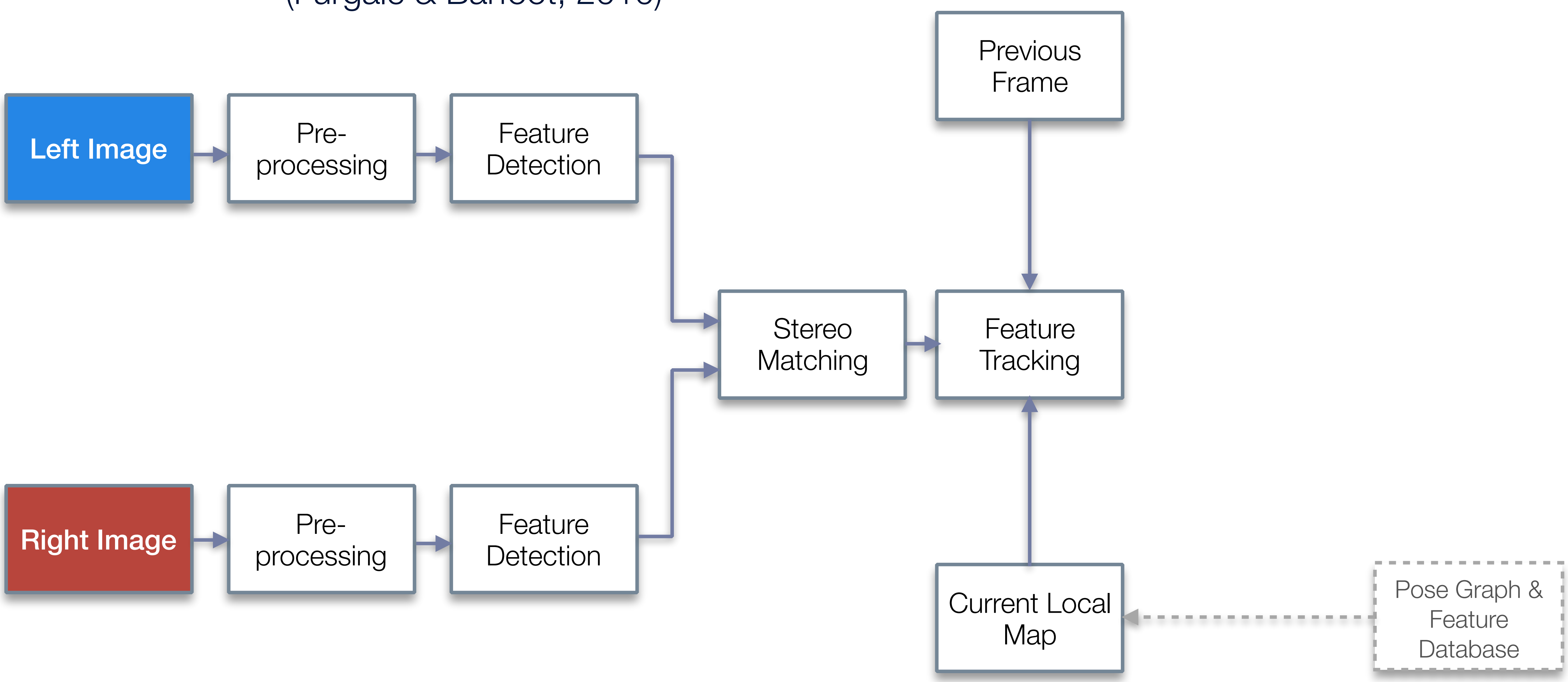
VT&R: Localization Pipeline

Stereo Pipeline
(Furgale & Barfoot, 2010)



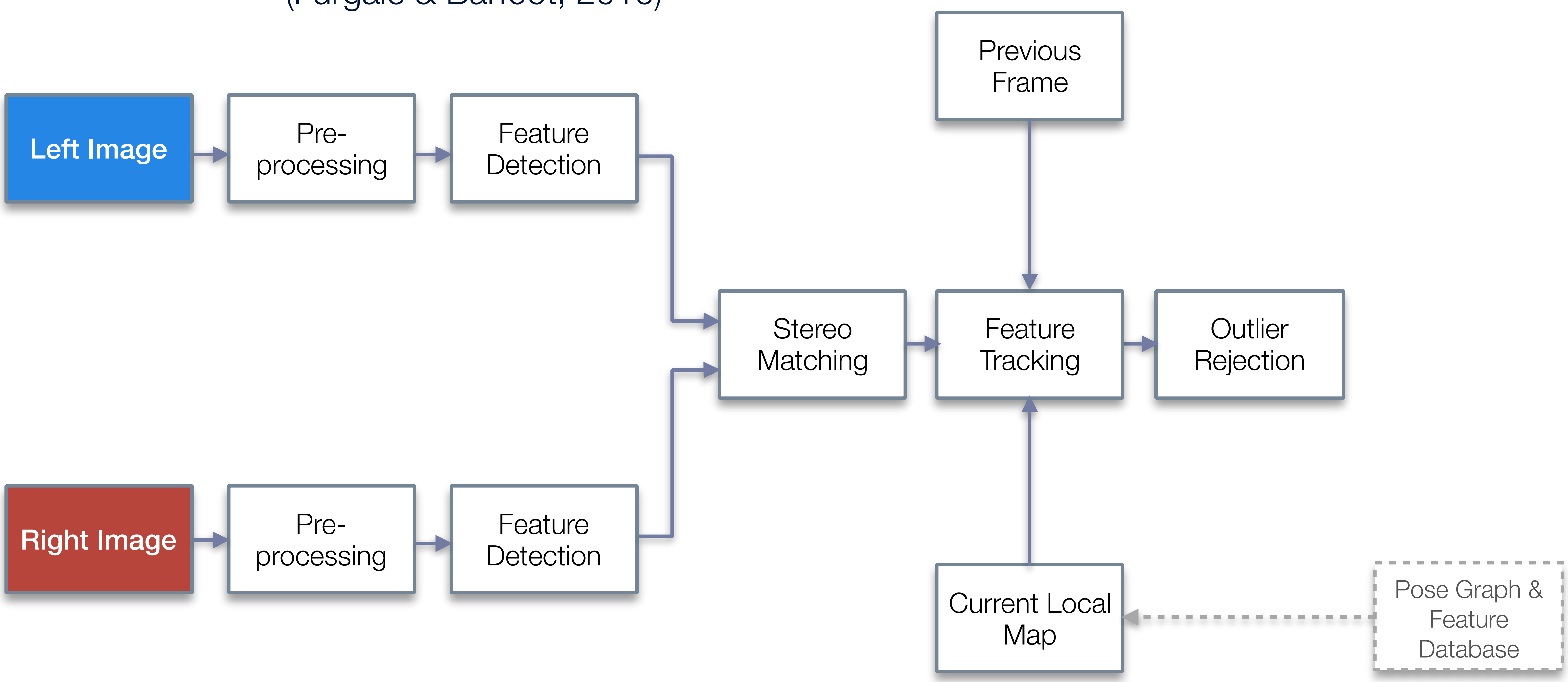
VT&R: Localization Pipeline

Stereo Pipeline
(Furgale & Barfoot, 2010)



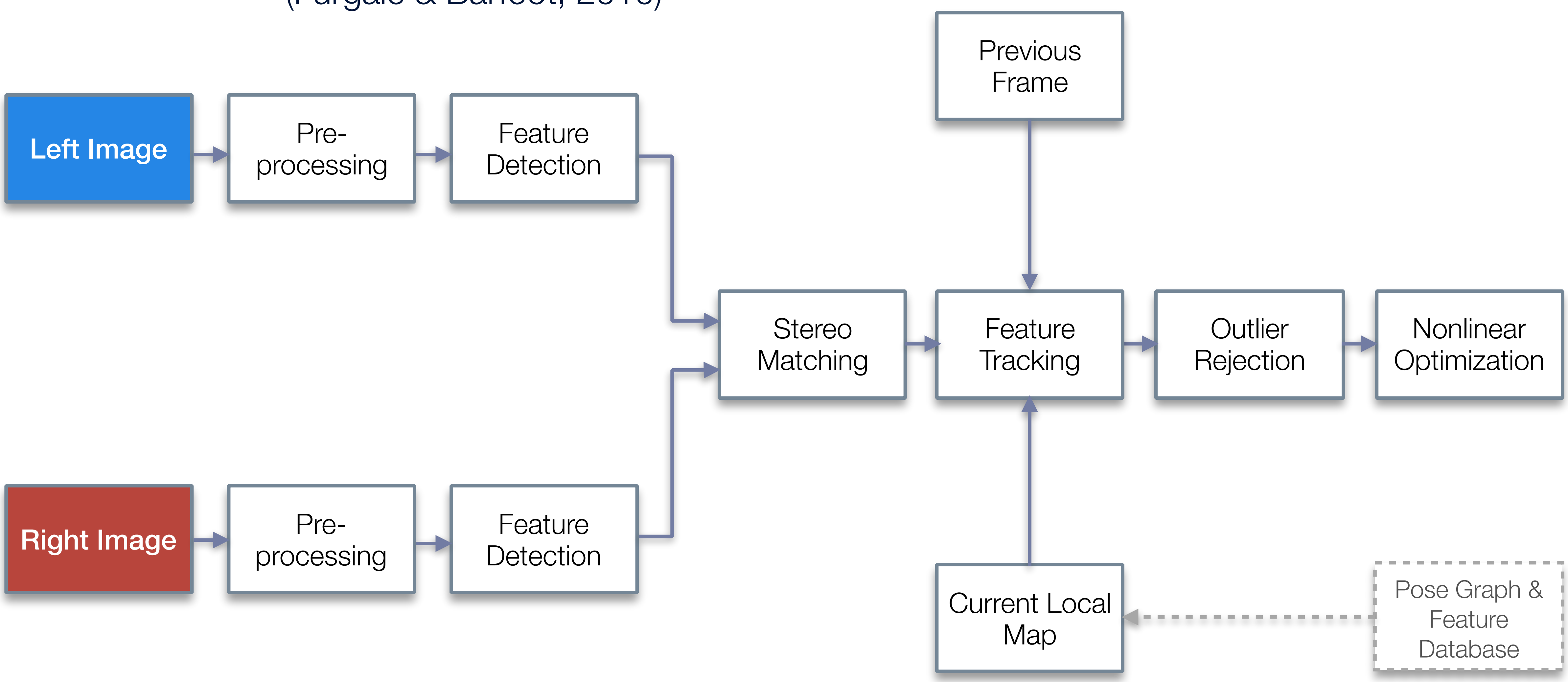
VT&R: Localization Pipeline

Stereo Pipeline
(Furgale & Barfoot, 2010)



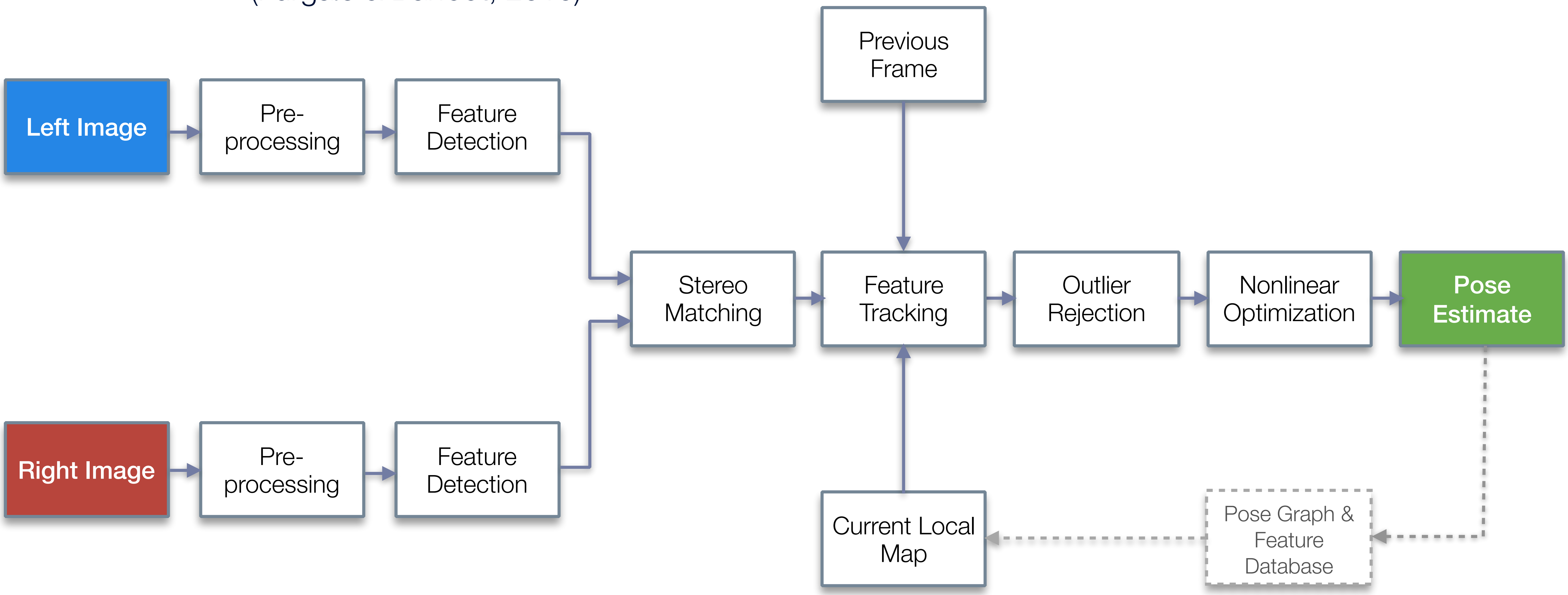
VT&R: Localization Pipeline

Stereo Pipeline
(Furgale & Barfoot, 2010)

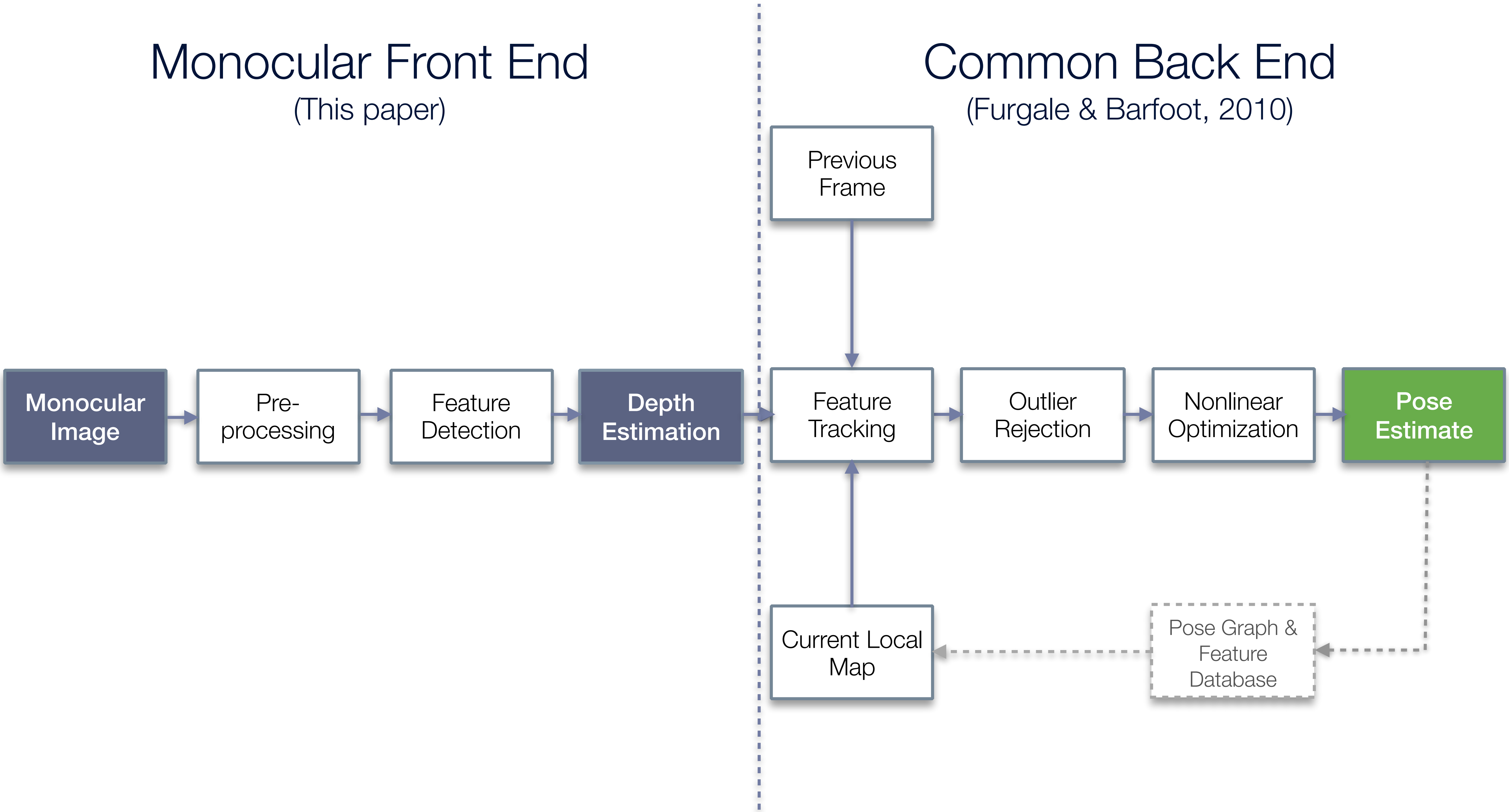


VT&R: Localization Pipeline

Stereo Pipeline
(Furgale & Barfoot, 2010)



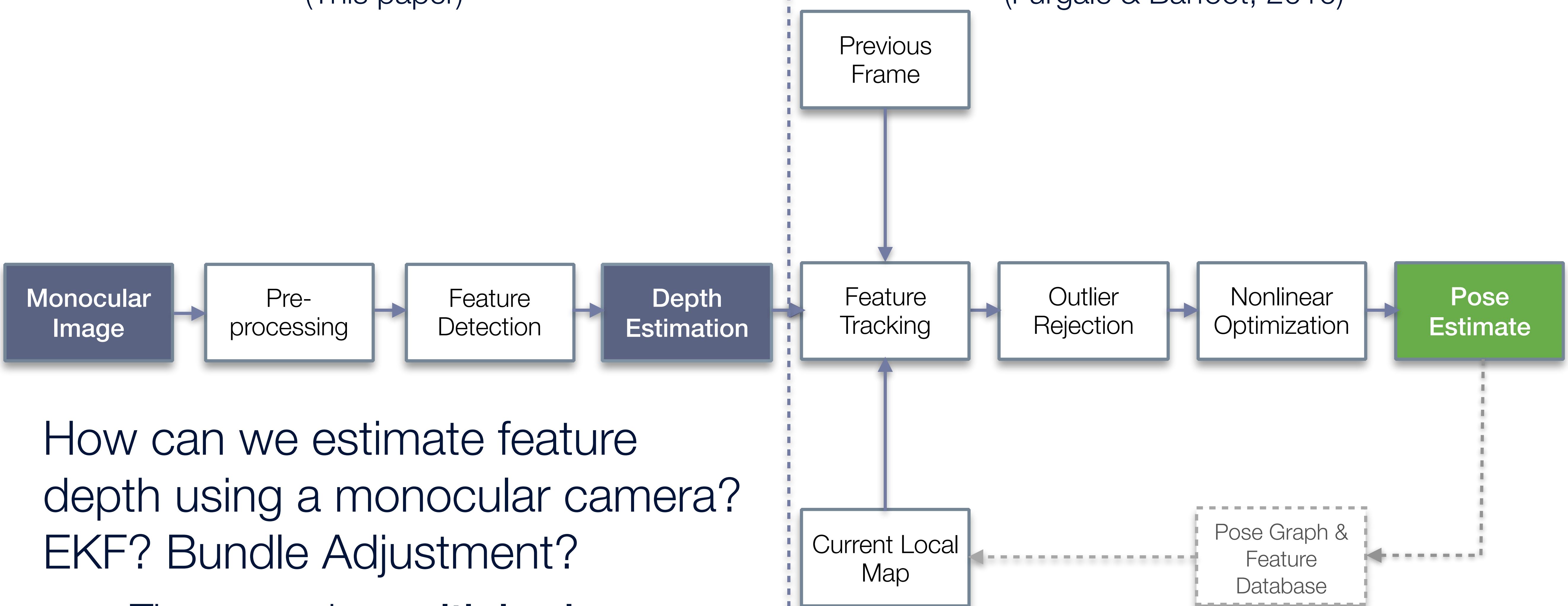
VT&R: Localization Pipeline



VT&R: Localization Pipeline

Monocular Front End
(This paper)

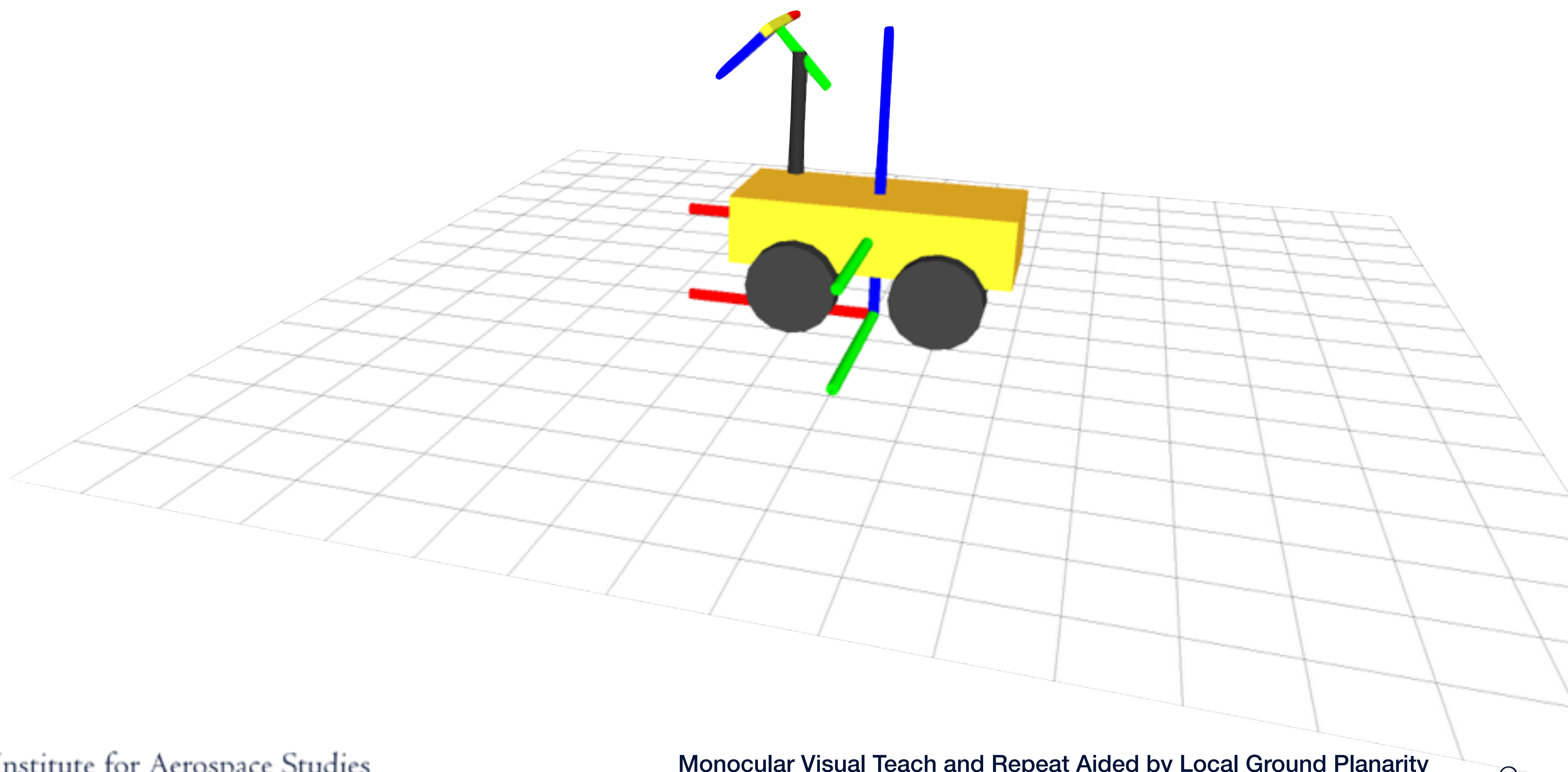
Common Back End
(Furgale & Barfoot, 2010)



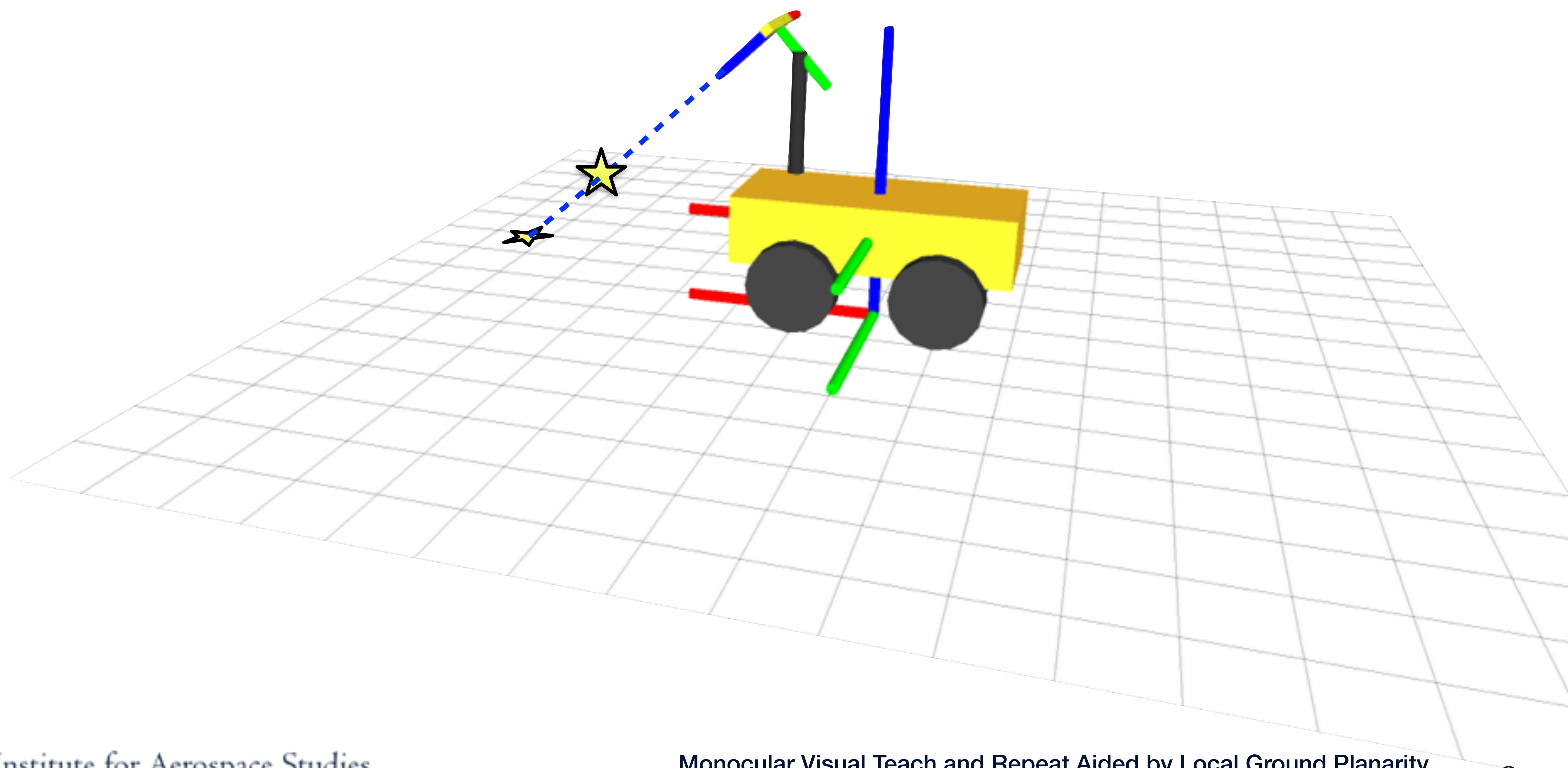
How can we estimate feature depth using a monocular camera? EKF? Bundle Adjustment?

- These require **multiple views**
- Need **scale** for control

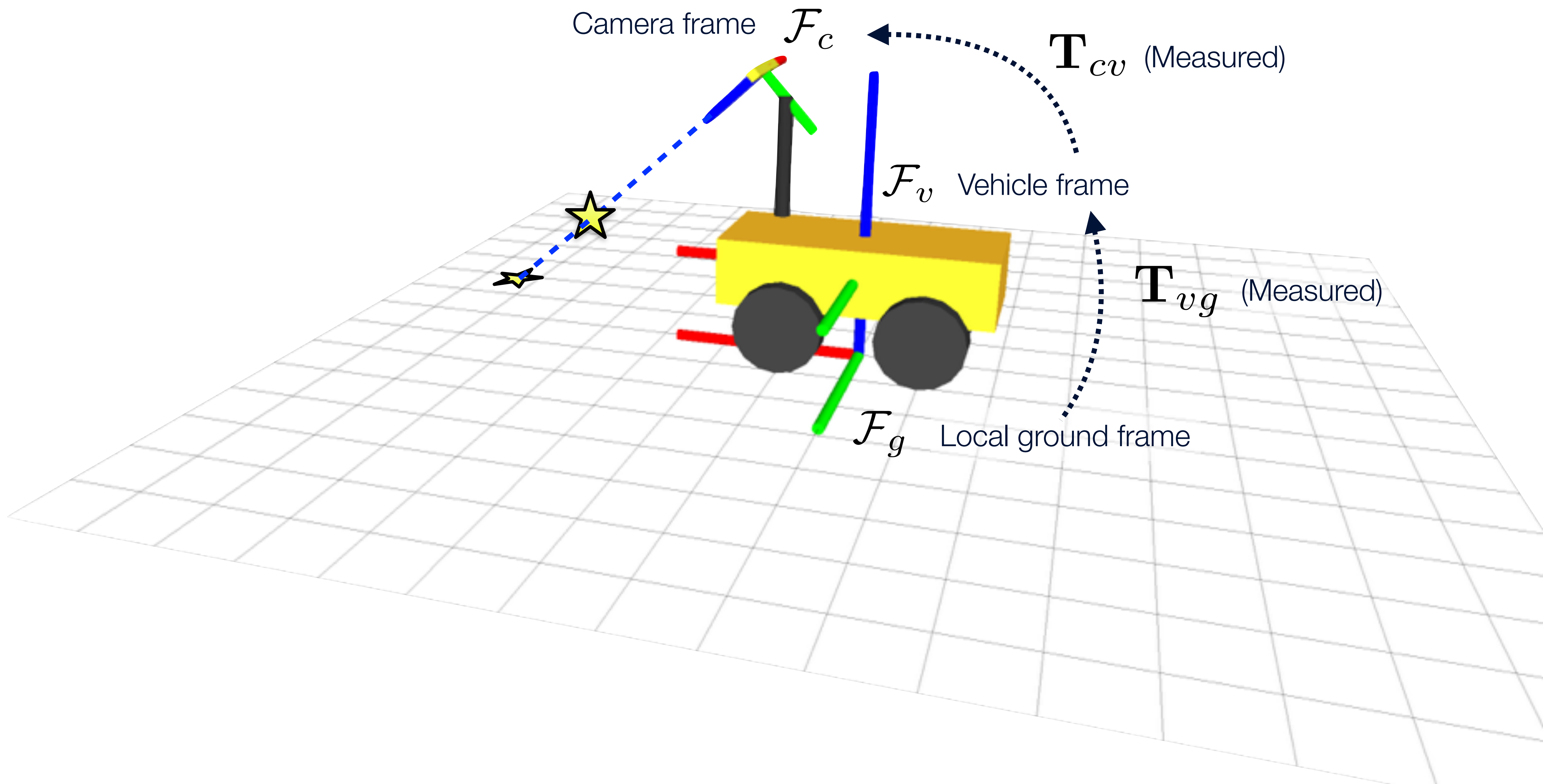
Depth Estimation: Local Ground Planarity



Depth Estimation: Local Ground Planarity



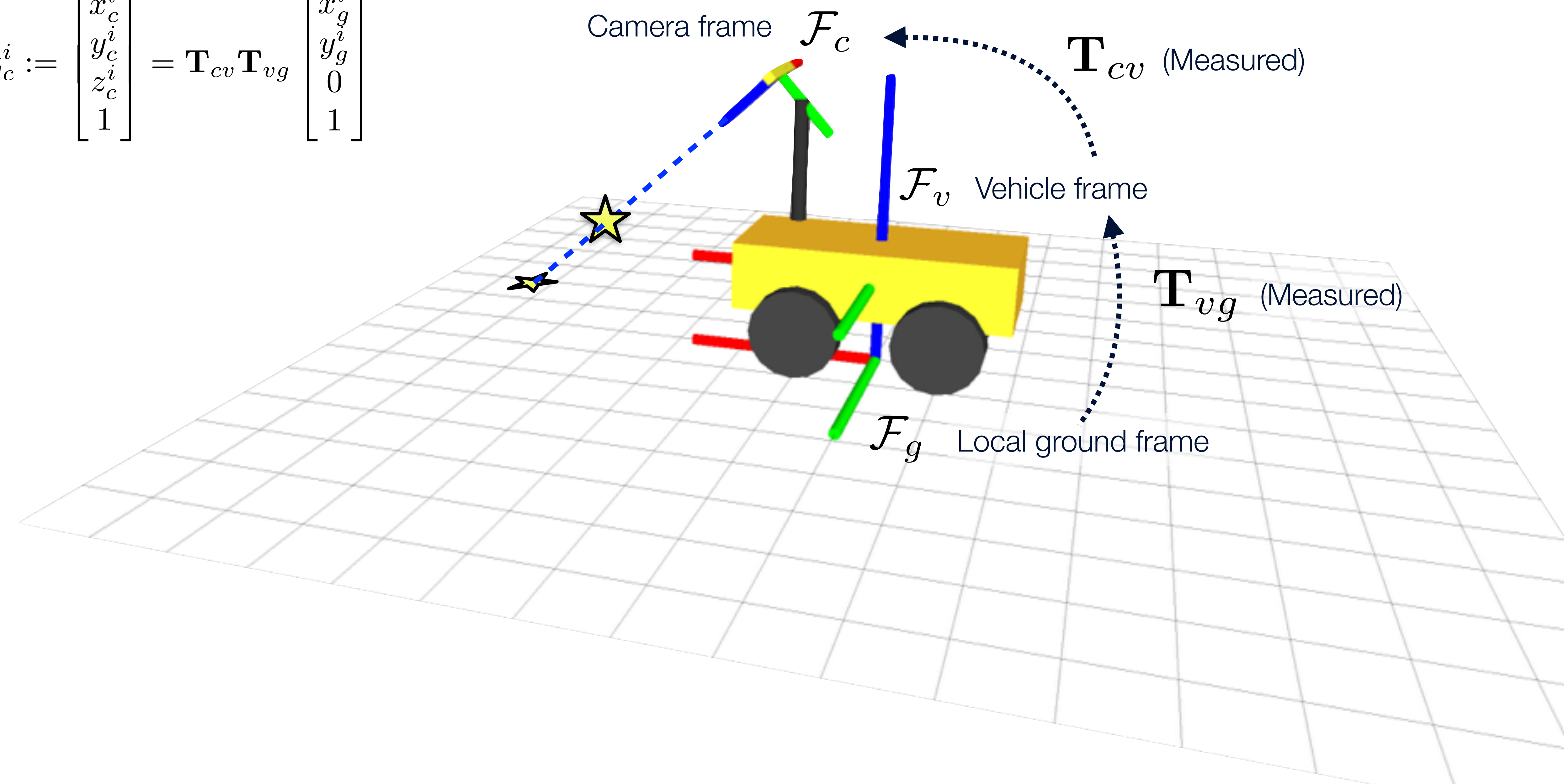
Depth Estimation: Local Ground Planarity



Depth Estimation: Local Ground Planarity

Ground feature coordinates in camera frame:

$$\mathbf{z}_c^i := \begin{bmatrix} x_c^i \\ y_c^i \\ z_c^i \\ 1 \end{bmatrix} = \mathbf{T}_{cv}\mathbf{T}_{vg}\begin{bmatrix} x_g^i \\ y_g^i \\ 0 \\ 1 \end{bmatrix}$$



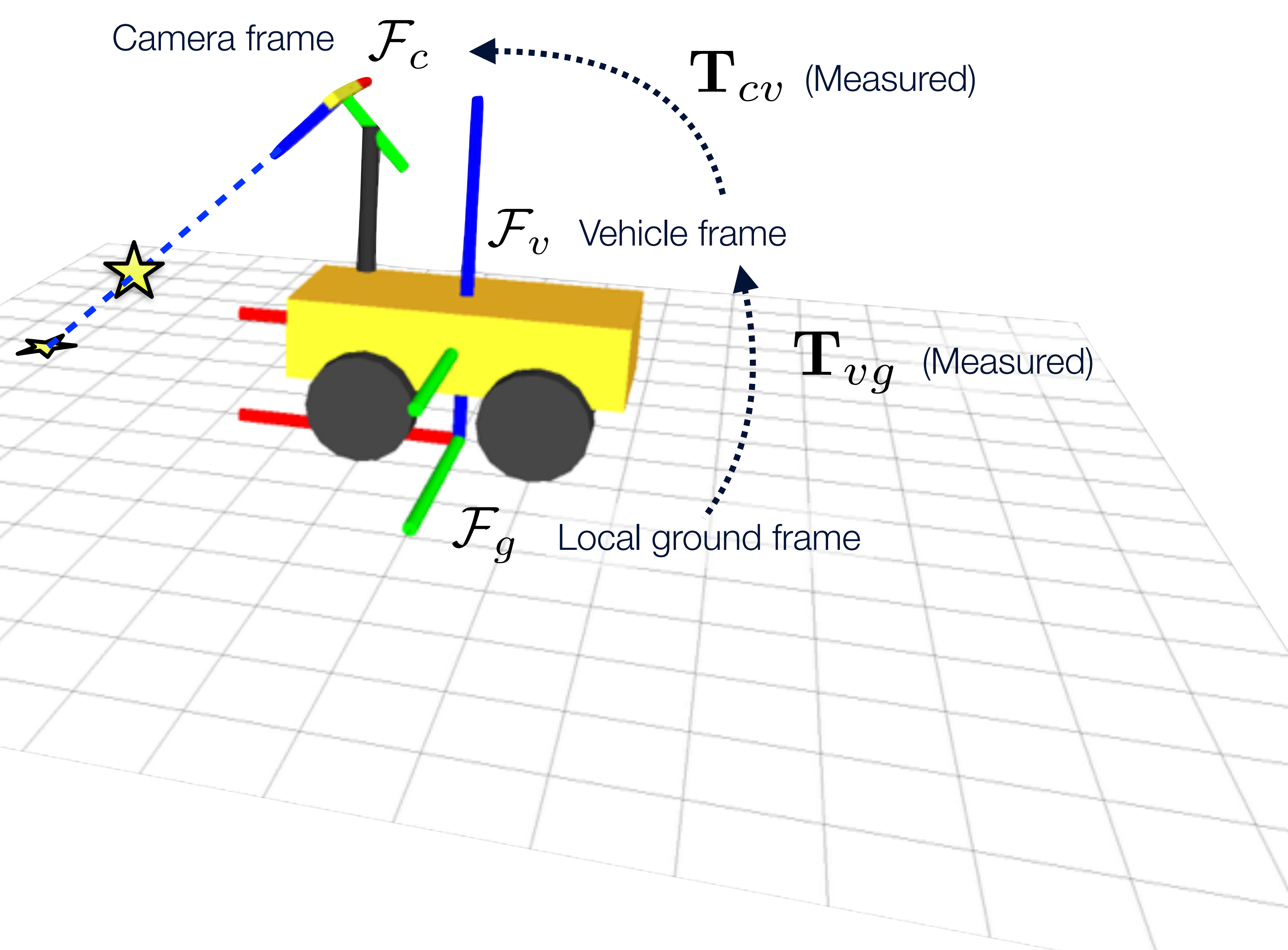
Depth Estimation: Local Ground Planarity

Ground feature coordinates in camera frame:

$$\mathbf{z}_c^i := \begin{bmatrix} x_c^i \\ y_c^i \\ z_c^i \\ 1 \end{bmatrix} = \mathbf{T}_{cv}\mathbf{T}_{vg} \begin{bmatrix} x_g^i \\ y_g^i \\ 0 \\ 1 \end{bmatrix}$$

Projective coordinates of feature point:

$$\mathbf{p}^i := \begin{bmatrix} p_x^i \\ p_y^i \\ 1 \end{bmatrix} = \frac{1}{z_c^i}\mathbf{z}_\mathbf{c}^\mathbf{i}$$



Depth Estimation: Local Ground Planarity

Ground feature coordinates in camera frame:

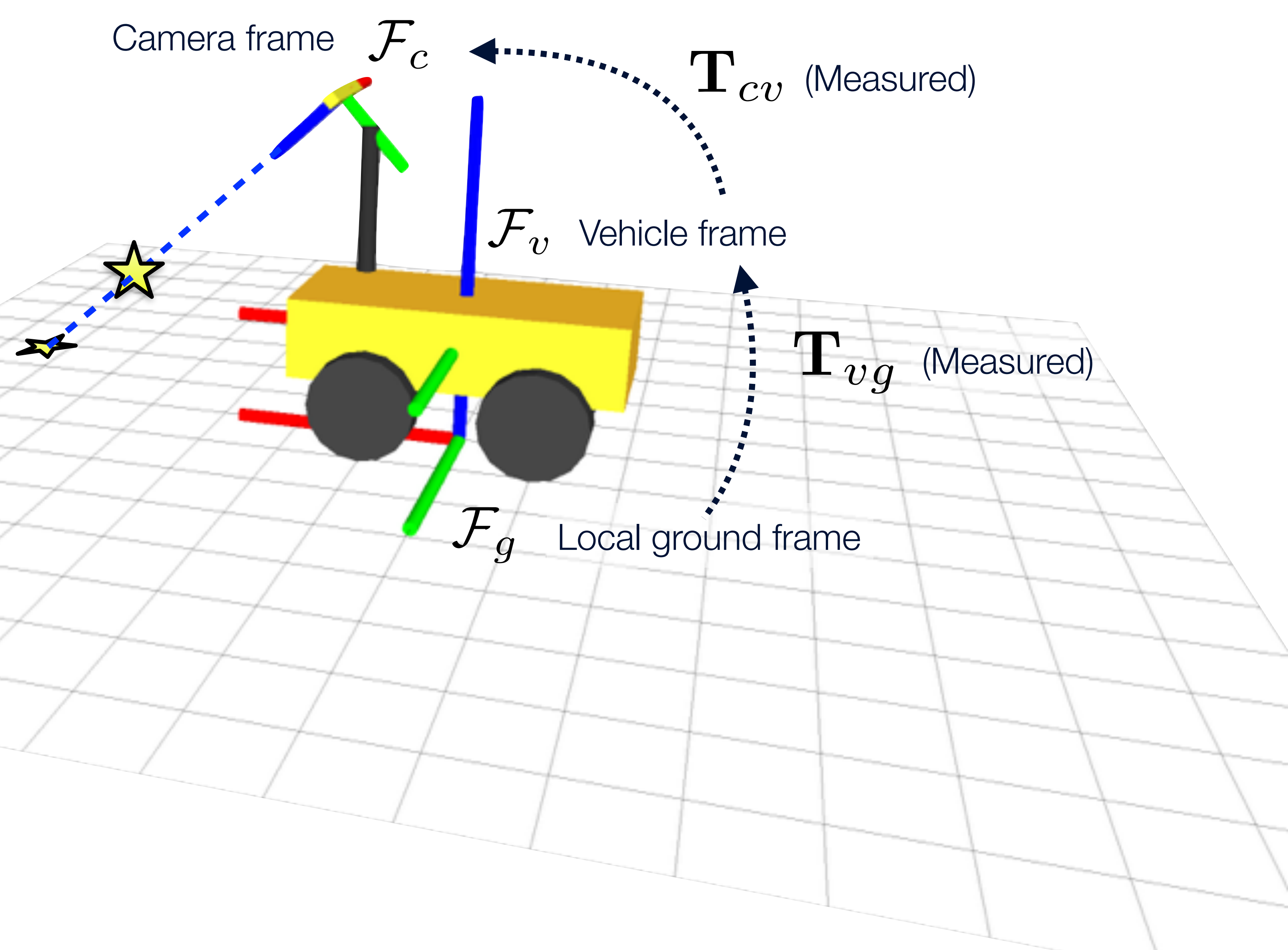
$$\mathbf{z}_c^i := \begin{bmatrix} x_c^i \\ y_c^i \\ z_c^i \\ 1 \end{bmatrix} = \mathbf{T}_{cv}\mathbf{T}_{vg} \begin{bmatrix} x_g^i \\ y_g^i \\ 0 \\ 1 \end{bmatrix}$$

Projective coordinates of feature point:

$$\mathbf{p}^i := \begin{bmatrix} p_x^i \\ p_y^i \\ 1 \end{bmatrix} = \frac{1}{z_c^i}\mathbf{z}_\mathbf{c}^\mathbf{i}$$

Image coordinates of feature point:

$$\mathbf{y}^i := \begin{bmatrix} u^i \\ v^i \\ 1 \end{bmatrix} = \mathbf{K}\mathbf{p}^i$$



Depth Estimation: Local Ground Planarity

Ground feature coordinates in camera frame:

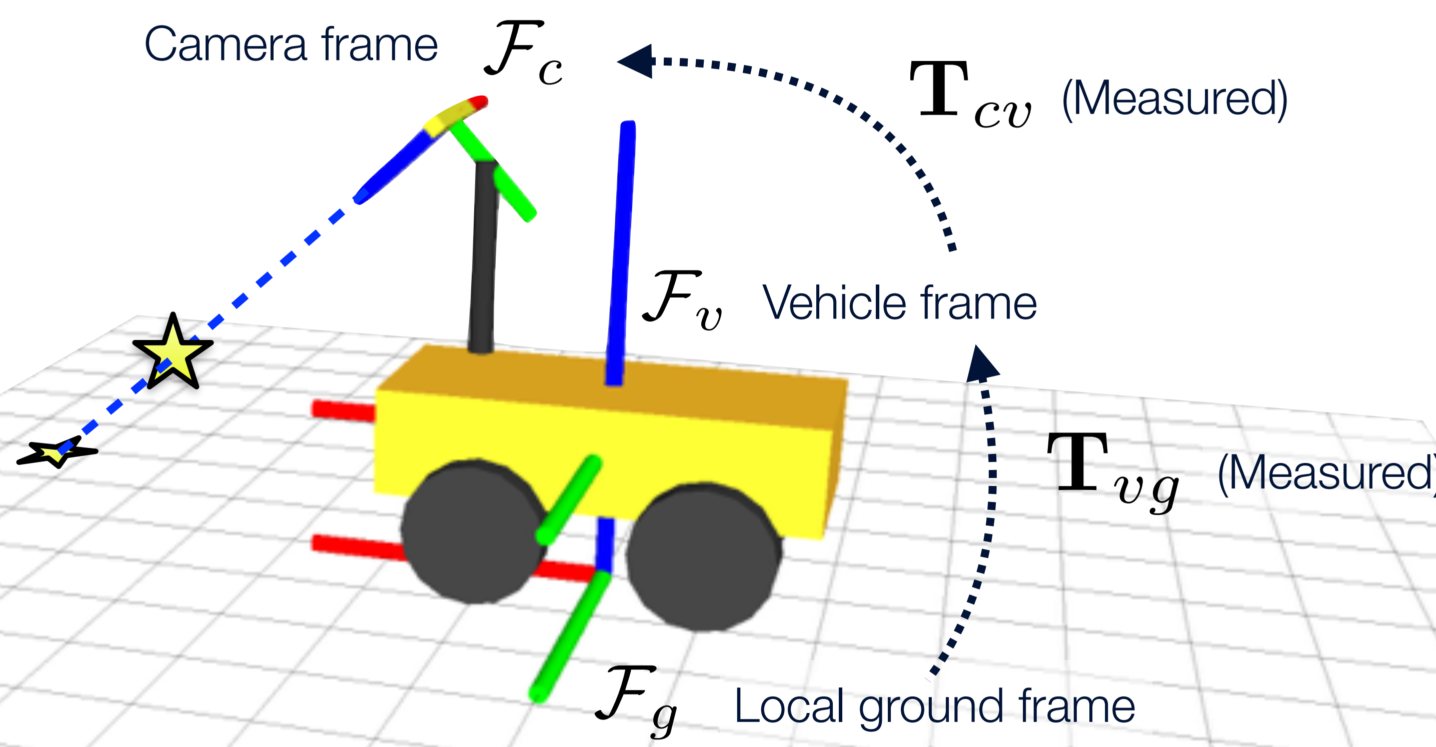
$$\mathbf{z}_c^i := \begin{bmatrix} x_c^i \\ y_c^i \\ z_c^i \\ 1 \end{bmatrix} = \mathbf{T}_{cv}\mathbf{T}_{vg} \begin{bmatrix} x_g^i \\ y_g^i \\ 0 \\ 1 \end{bmatrix}$$

Projective coordinates of feature point:

$$\mathbf{p}^i := \begin{bmatrix} p_x^i \\ p_y^i \\ 1 \end{bmatrix} = \frac{1}{z_c^i}\mathbf{z}_\mathbf{c}^\mathbf{i}$$

Image coordinates of feature point:

$$\mathbf{y}^i := \begin{bmatrix} u^i \\ v^i \\ 1 \end{bmatrix} = \mathbf{K}\mathbf{p}^i$$



This works well for 3DOF **visual odometry**, but not so well for **mapping**!

Depth Estimation: Uncertainty?

The ground isn't perfectly flat, but locally it's close!

Idea: Model the ground plane as a Gaussian distribution on SE(3)

Depth Estimation: Uncertainty?

The ground isn't perfectly flat, but locally it's close!

Idea: Model the ground plane as a Gaussian distribution on SE(3)

Covariance of 2D feature pixel coordinates

Covariance of ground plane distribution on SE(3)

$$\mathbf{R}_i = \begin{bmatrix} \mathbf{R}_{\mathbf{y}^i} & \mathbf{0} \\ \mathbf{0} & \mathbf{R}_{\mathbf{T}_{vg}} \end{bmatrix}$$

Depth Estimation: Uncertainty?

The ground isn't perfectly flat, but locally it's close!

Idea: Model the ground plane as a Gaussian distribution on SE(3)

Covariance of 2D feature pixel coordinates

Covariance of ground plane distribution on SE(3)

$$\mathbf{R}_i = \begin{bmatrix} \mathbf{R}_{\mathbf{y}^i} & \mathbf{0} \\ \mathbf{0} & \mathbf{R}_{\mathbf{T}_{vg}} \end{bmatrix}$$

$$\mathbf{G}_i = \begin{bmatrix} \frac{\partial \mathbf{z}_c^i}{\partial \mathbf{y}^i} & \frac{\partial \mathbf{z}_c^i}{\partial \mathbf{T}_{vg}} \end{bmatrix}$$

Observation model Jacobian

Depth Estimation: Uncertainty?

The ground isn't perfectly flat, but locally it's close!

Idea: Model the ground plane as a Gaussian distribution on SE(3)

Covariance of 2D feature pixel coordinates

Covariance of ground plane distribution on SE(3)

$$\mathbf{R}_i = \begin{bmatrix} \mathbf{R}_{\mathbf{y}^i} & \mathbf{0} \\ \mathbf{0} & \mathbf{R}_{\mathbf{T}_{vg}} \end{bmatrix}$$

$$\mathbf{G}_i = \begin{bmatrix} \frac{\partial \mathbf{z}_c^i}{\partial \mathbf{y}^i} & \frac{\partial \mathbf{z}_c^i}{\partial \mathbf{T}_{vg}} \end{bmatrix}$$

Observation model Jacobian

$$\mathbf{z}_c^i \sim \mathcal{N}(\bar{\mathbf{z}}_c^i, \mathbf{G}_i \mathbf{R}_i \mathbf{G}_i^T)$$

Gaussian approximation of 3D feature coordinates

Depth Estimation: Uncertainty?

The ground isn't perfectly flat, but locally it's close!

Idea: Model the ground plane as a Gaussian distribution on SE(3)

Covariance of 2D feature pixel coordinates

Covariance of ground plane distribution on SE(3)

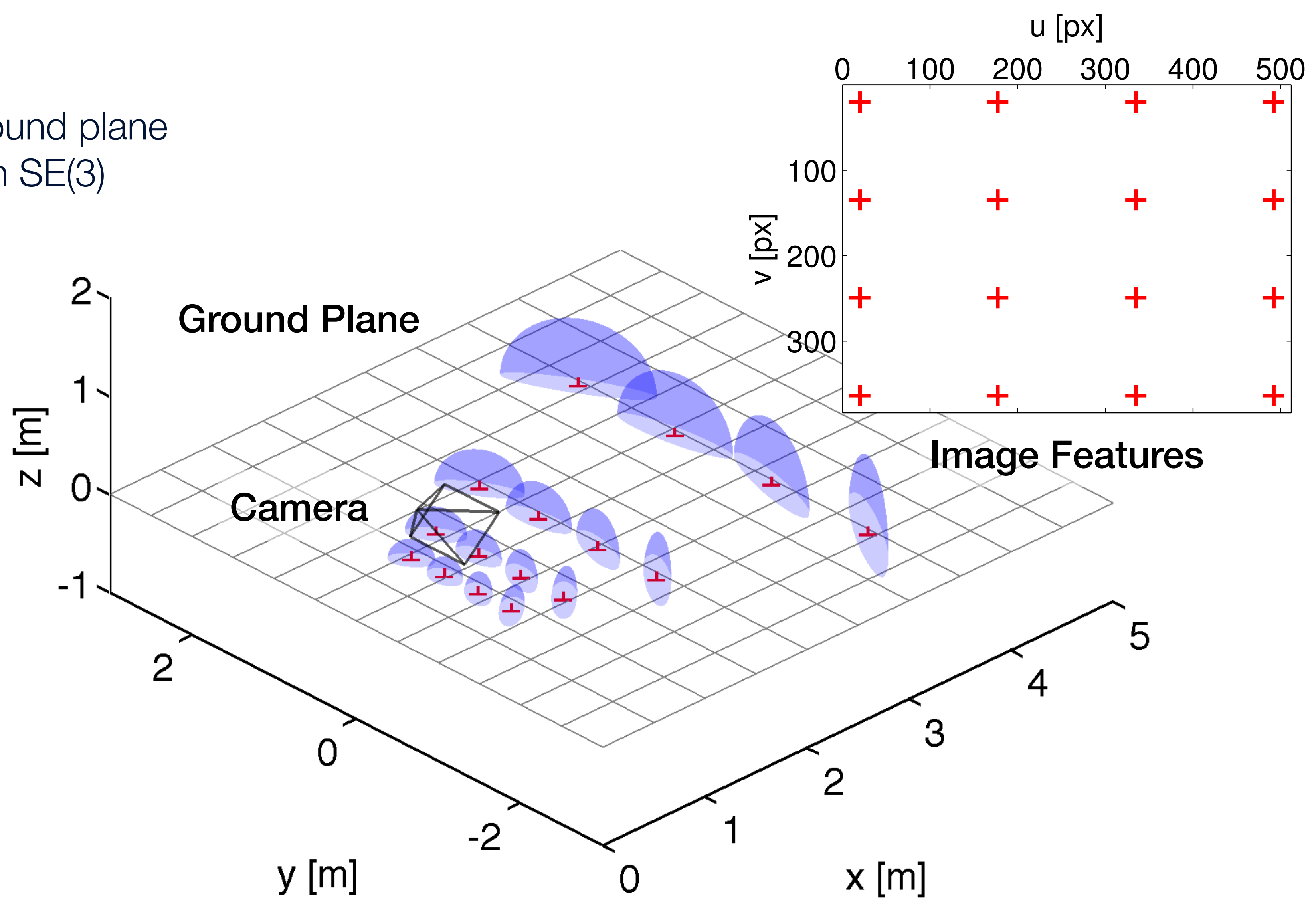
$$\mathbf{R}_i = \begin{bmatrix} \mathbf{R}_{\mathbf{y}^i} & \mathbf{0} \\ \mathbf{0} & \mathbf{R}_{\mathbf{T}_{vg}} \end{bmatrix}$$

$$\mathbf{G}_i = \begin{bmatrix} \frac{\partial \mathbf{z}_c^i}{\partial \mathbf{y}^i} & \frac{\partial \mathbf{z}_c^i}{\partial \mathbf{T}_{vg}} \end{bmatrix}$$

Observation model Jacobian

$$\mathbf{z}_c^i \sim \mathcal{N}(\bar{\mathbf{z}}_c^i, \mathbf{G}_i \mathbf{R}_i \mathbf{G}_i^T)$$

Gaussian approximation of 3D feature coordinates



Monocular Visual Teach and Repeat Aided by Local Ground Planarity
Lee Clement, Jonathan Kelly, and Timothy D. Barfoot

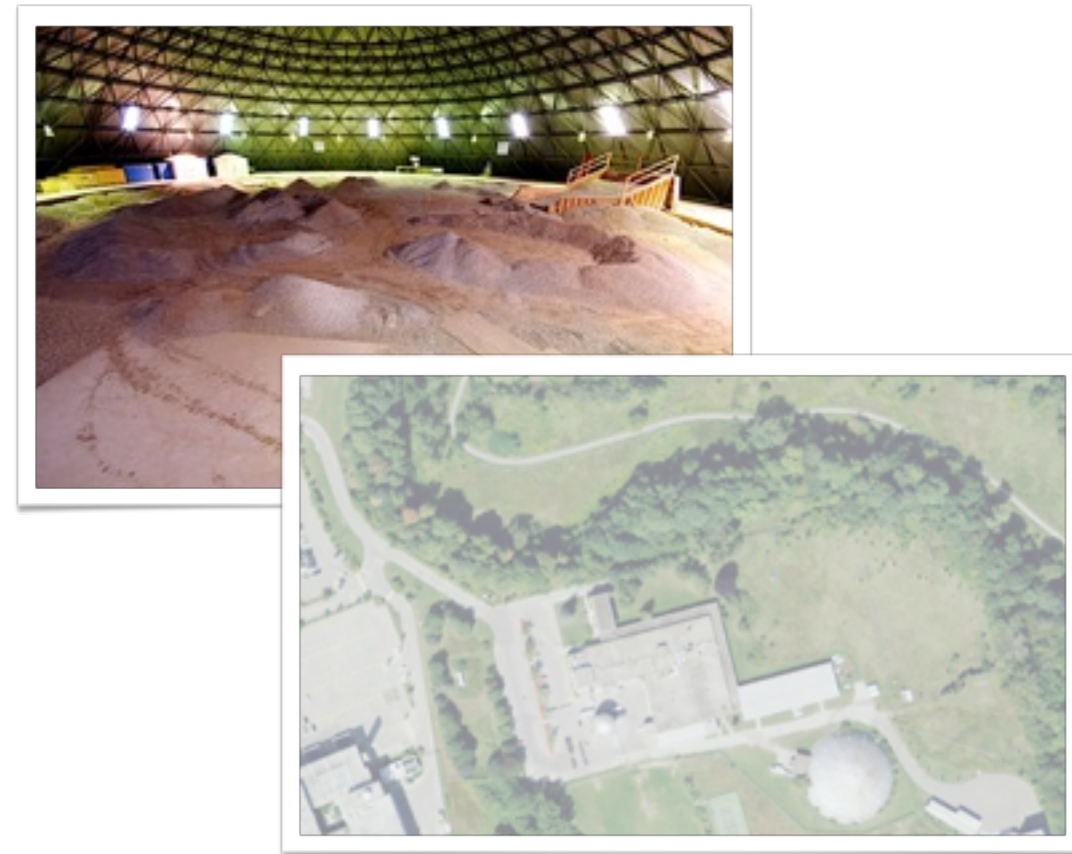
Field Testing: Goals

Field Testing: Goals

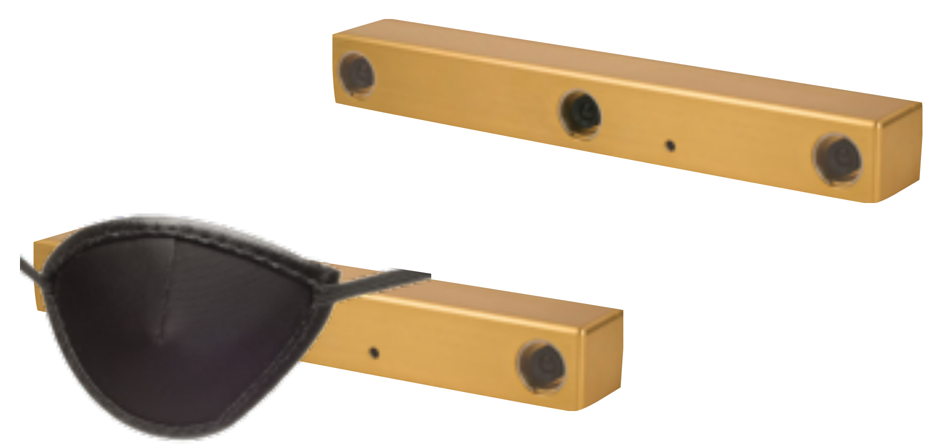
1

Characterize the performance of Monocular VT&R over long routes in different conditions.

Field Testing: Goals



1. **Characterize** the performance of Monocular VT&R over long routes in different conditions.



2. **Compare** the performance of Monocular VT&R to Stereo VT&R on the same routes, using the same hardware.

Field Testing: Hardware

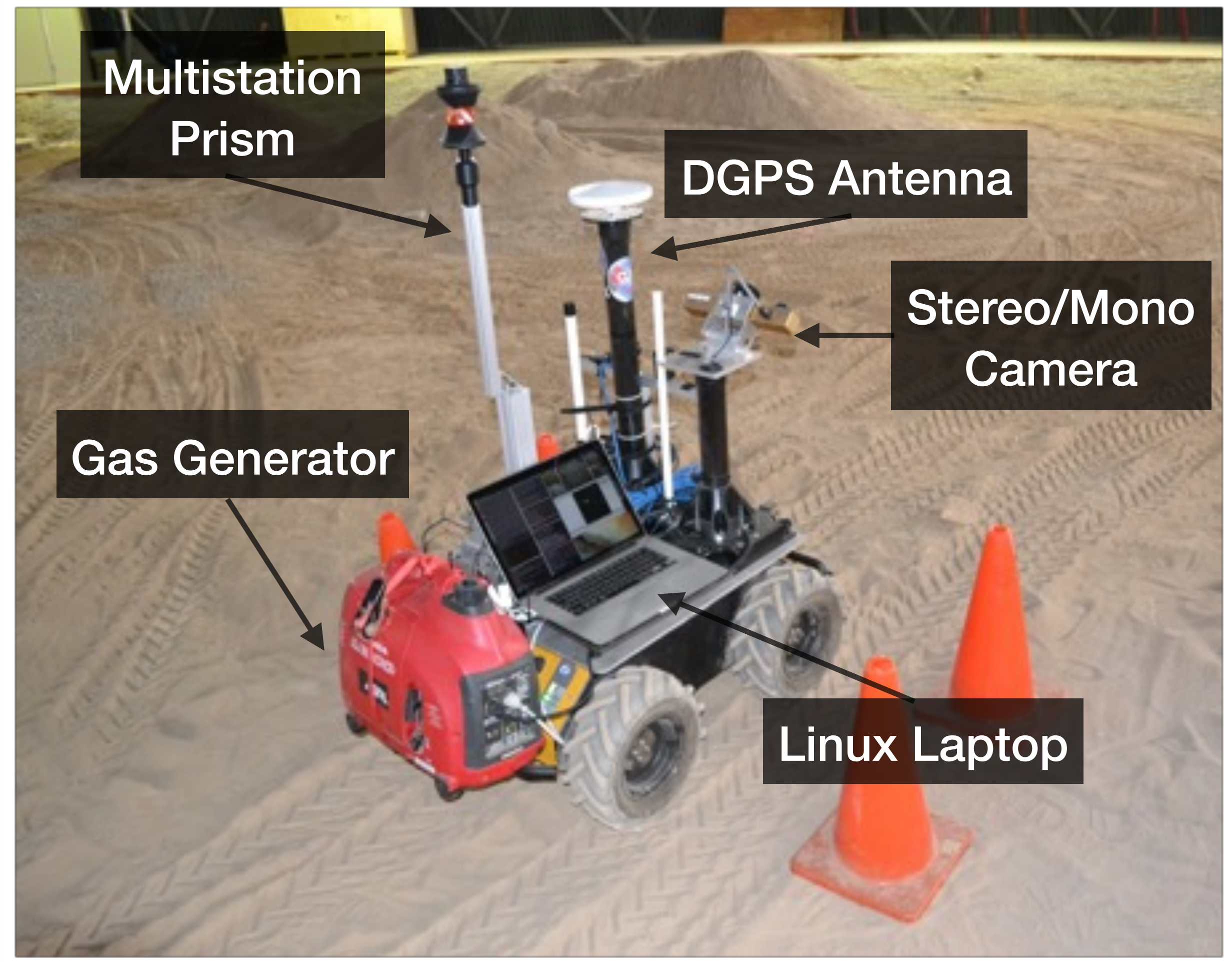
Vehicle:
- Clearpath Husky A200 Rover

Sensor:
- PointGrey Bumblebee XB3 Stereo Camera (1 m from ground, 47° to horizontal)

Computation:
- MacBook Pro (Ubuntu 12.04, ROS Hydro)

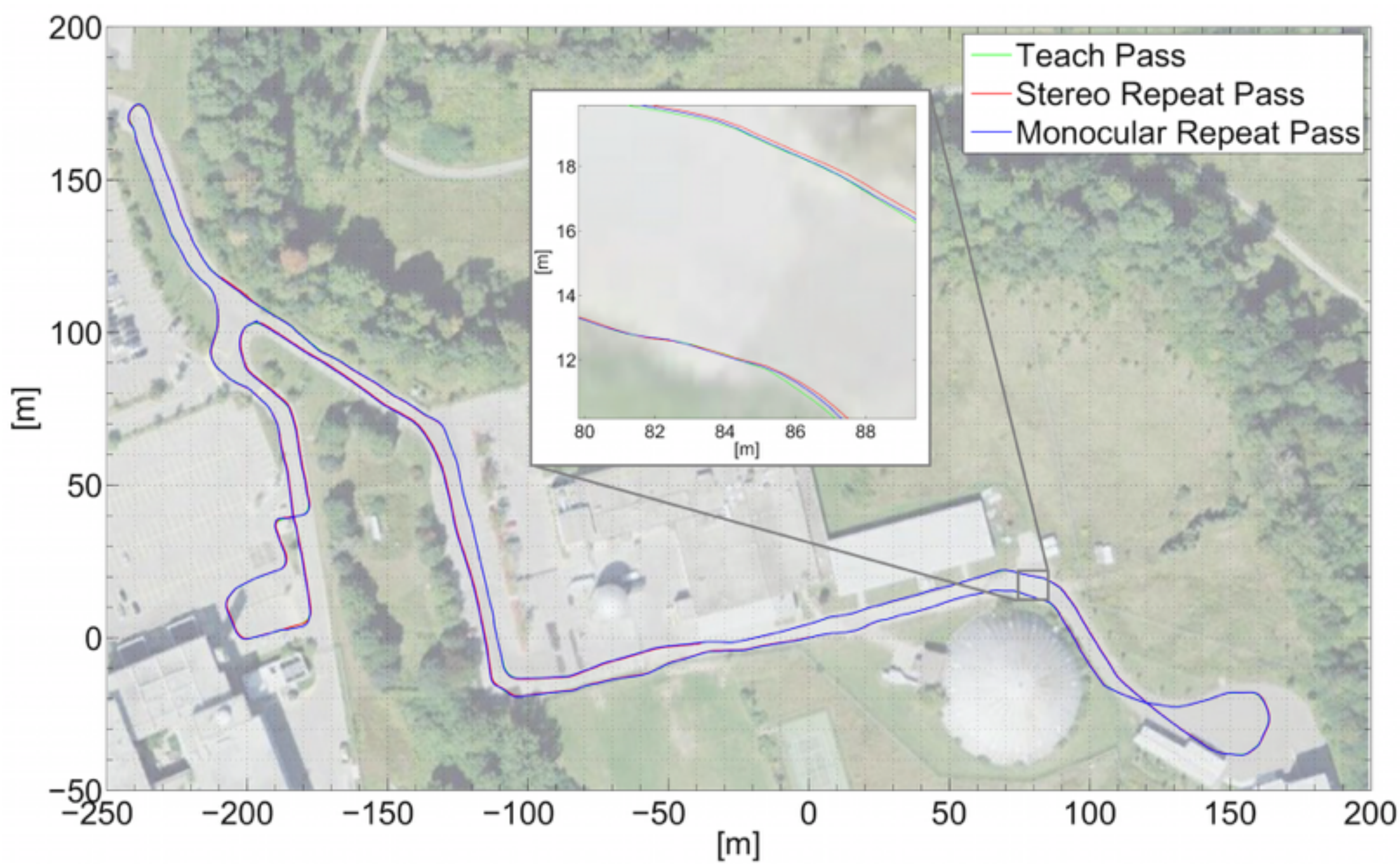
Ground Truthing:
- Ashtech DG14 DGPS (Outdoor)
- Leica Nova MS50 Multistation (Indoor)

Other:
- 1 kW Gas Generator



Field Testing: 1.4 km UTIAS Outdoor Route

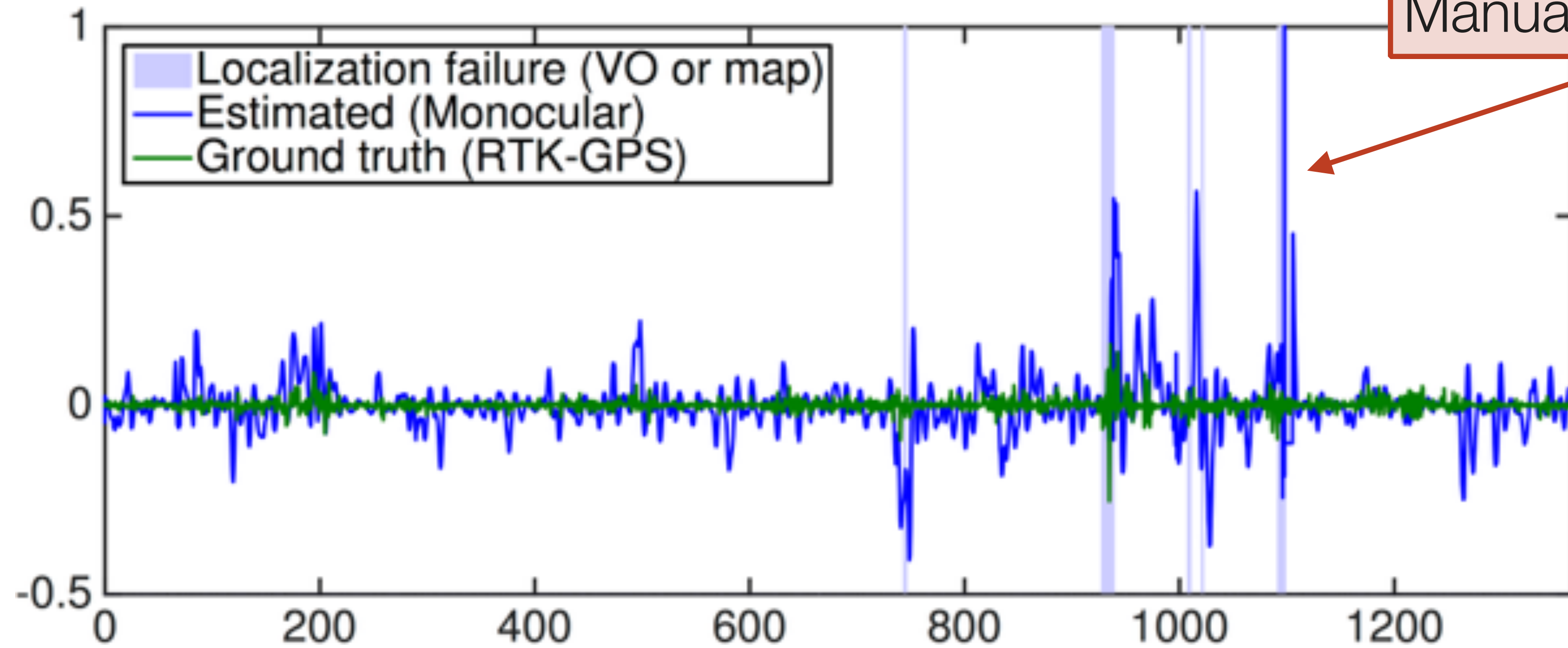
Ground Truth: RTK-GPS



Field Testing: 1.4 km UTIAS Outdoor Route

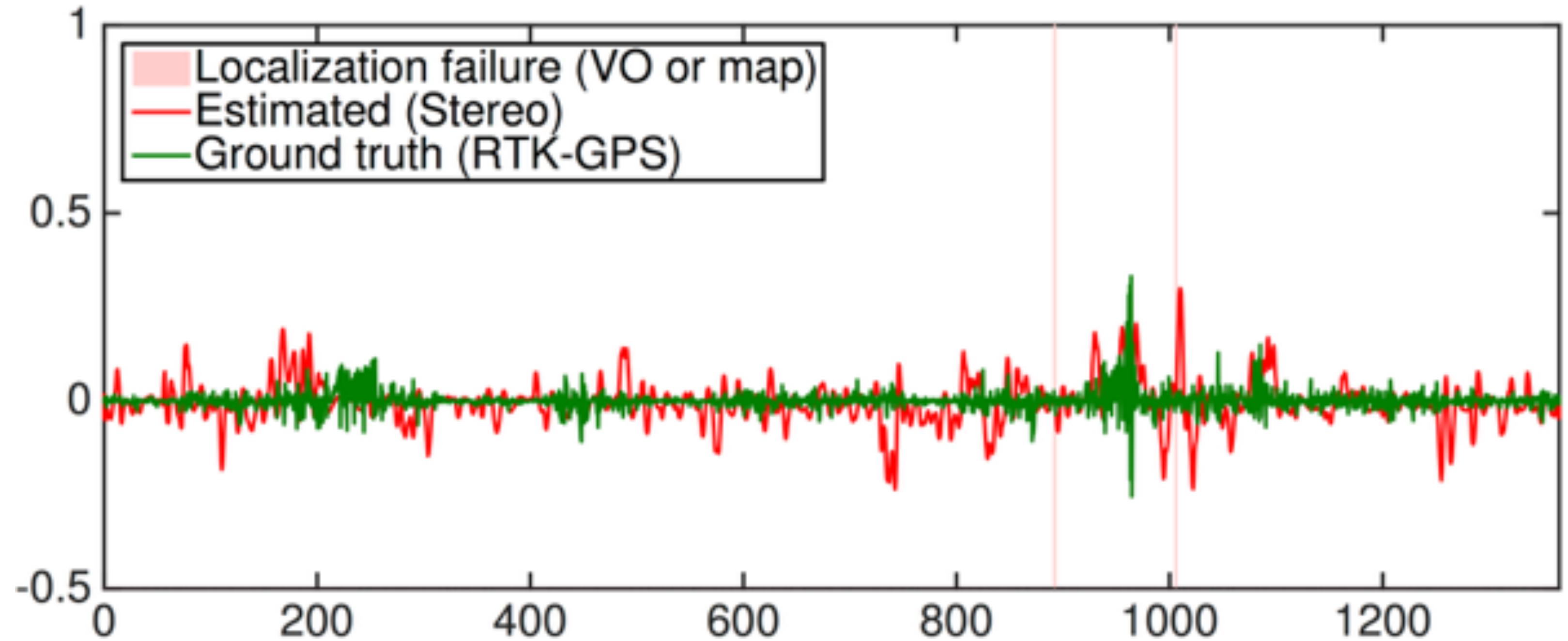
Manual intervention

Lateral Path-Tracking Error [m]



Monocular
99.9% autonomy

Lateral Path-Tracking Error [m]

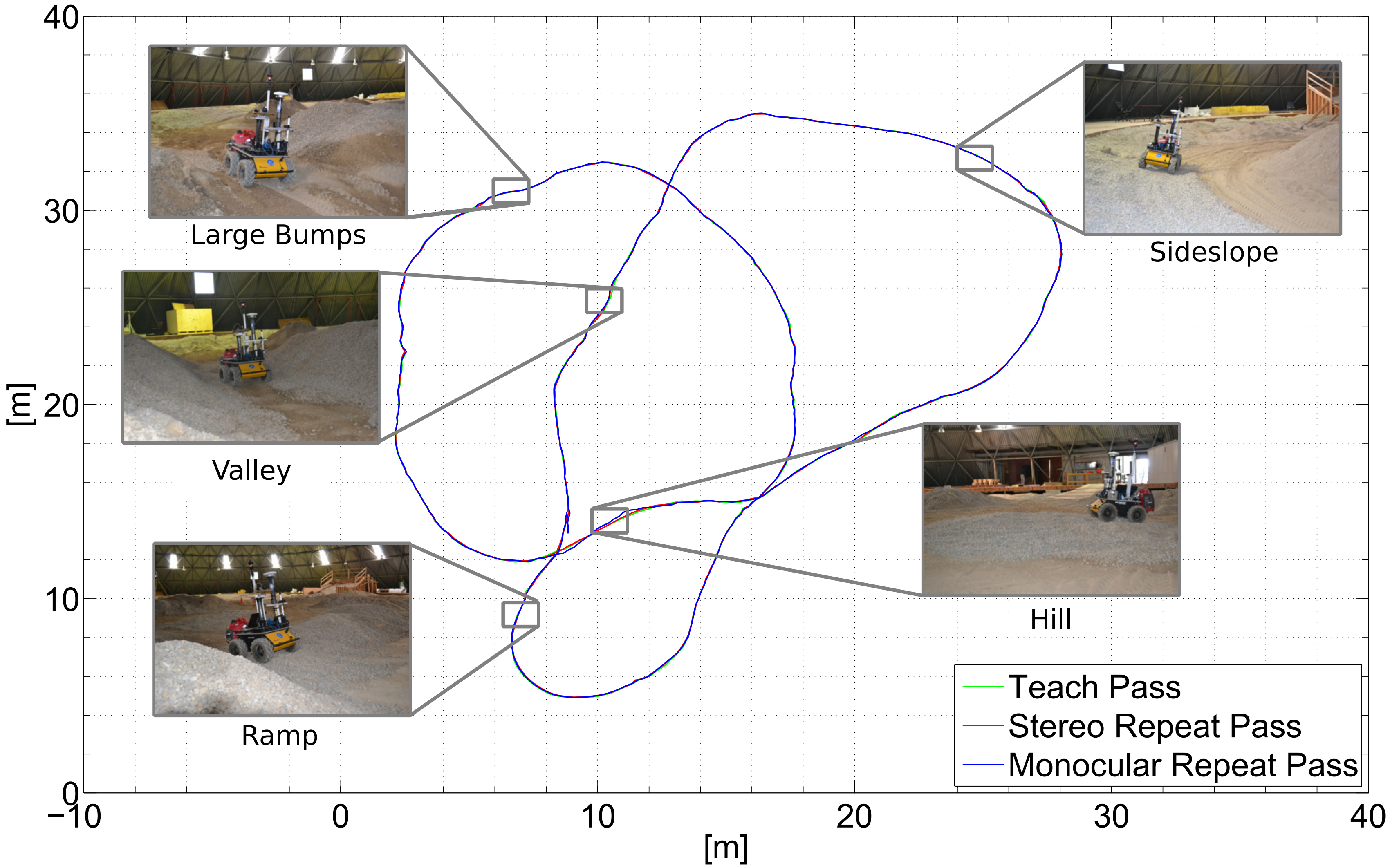


Stereo
100% autonomy

Distance Traveled [m]

Institute for Aerospace Studies
UNIVERSITY OF TORONTO

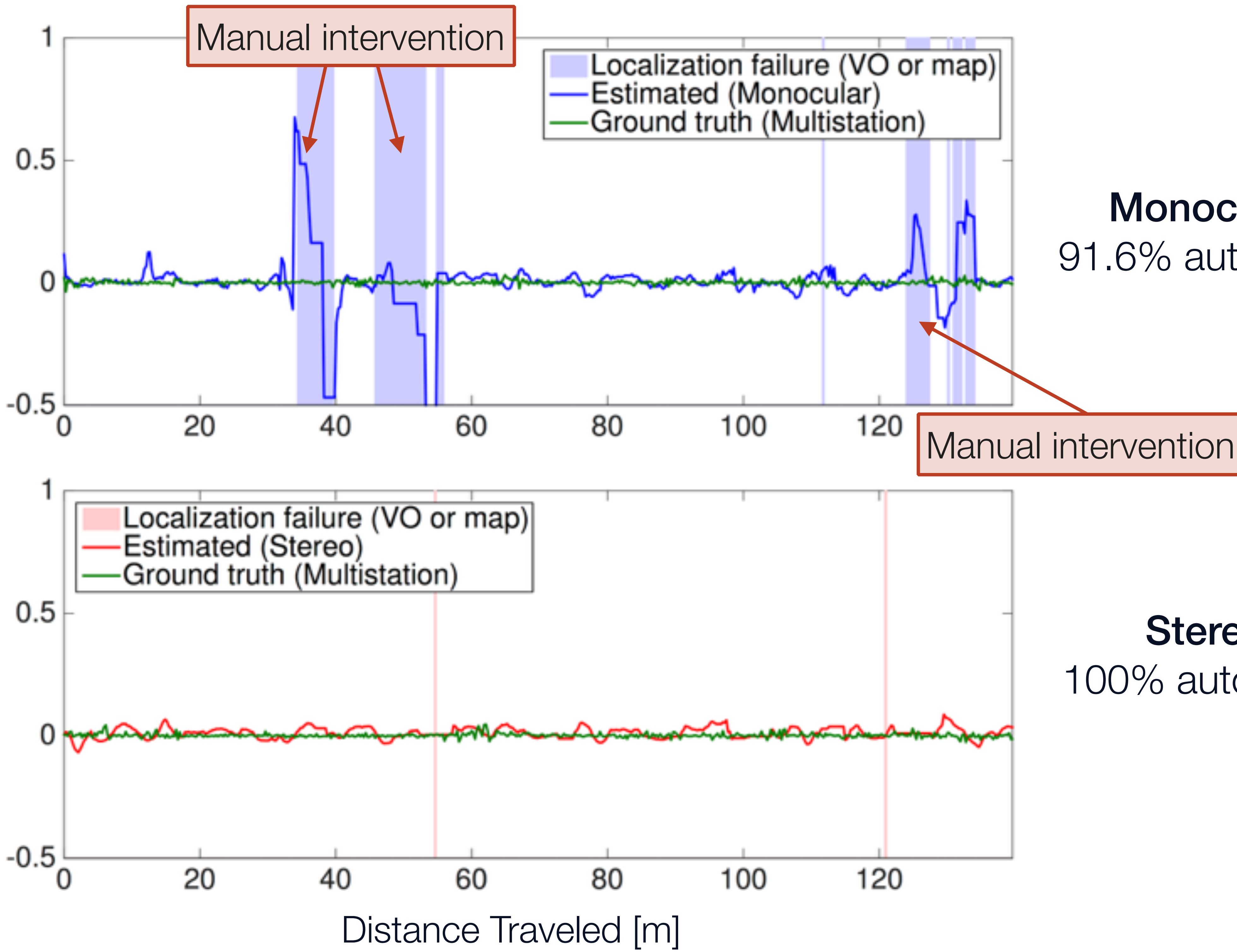
Field Testing: 140 m MarsDome Indoor Route



Ground Truth:
Leica Nova MS50 Multistation

Field Testing: 140 m MarsDome Indoor Route

Lateral Path-Tracking Error [m]



Monocular
91.6% autonomy

Lateral Path-Tracking Error [m]

Stereo
100% autonomy

Field Testing: Summary

Trial	Route	Path length	Repeat speed	Local start time (UTC-4)			Autonomy rate	
				Teach	Mono	Stereo	Mono	Stereo
1	Outdoor	1370 m	0.6 m/s	09:56:46	10:35:10	12:08:30	99.71%†	100.00%
2	Outdoor	1360 m	0.6 m/s	11:45:40	12:22:26	13:43:49	99.88%	100.00%
3	Outdoor	1361 m	0.6 m/s	13:26:41	14:00:12	15:20:12	99.74%	100.00%
4	Indoor	126 m	0.3 m/s	13:32:23	13:40:53	14:02:46	96.28%	100.00%
5	Indoor	140 m	0.3 m/s	12:18:57	12:32:20	12:59:11	91.60%	100.00%

	Mono	**Stereo**
Total distance driven	4298 m†	4357 m
Total distance autonomously traversed	99.41%	100.00%

Field Testing: Summary

Trial	Route	Path length	Repeat speed	Local start time (UTC-4) Teach	Mono	Stereo	Autonomy rate Mono	Stereo
1	Outdoor	1370 m	0.6 m/s	09:56:46	10:35:10	12:08:30	99.71%†	100.00%
2	Outdoor	1360 m	0.6 m/s	11:45:40	12:22:26	13:43:49	99.88%	100.00%
3	Outdoor	1361 m	0.6 m/s	13:26:41	14:00:12	15:20:12	99.74%	100.00%
4	Indoor	126 m	0.3 m/s	13:32:23	13:40:53	14:02:46	96.28%	100.00%
5	Indoor	140 m	0.3 m/s	12:18:57	12:32:20	12:59:11	91.60%	100.00%

	Mono	**Stereo**
Total distance driven	4298 m†	4357 m
Total distance autonomously traversed	99.41%	100.00%

Field Testing: Summary

				Local start time (UTC-4)			Autonomy rate	
Trial	Route	Path length	Repeat speed	Teach	Mono	Stereo	Mono	Stereo
1	Outdoor	1370 m	0.6 m/s	09:56:46	10:35:10	12:08:30	99.71%†	100.00%
2	Outdoor	1360 m	0.6 m/s	11:45:40	12:22:26	13:43:49	99.88%	100.00%
3	Outdoor	1361 m	0.6 m/s	13:26:41	14:00:12	15:20:12	99.74%	100.00%
4	Indoor	126 m	0.3 m/s	13:32:23	13:40:53	14:02:46	96.28%	100.00%
5	Indoor	140 m	0.3 m/s	12:18:57	12:32:20	12:59:11	91.60%	100.00%

	Mono	**Stereo**
Total distance driven	4298 m†	4357 m
Total distance autonomously traversed	99.41%	100.00%

Field Testing: Summary

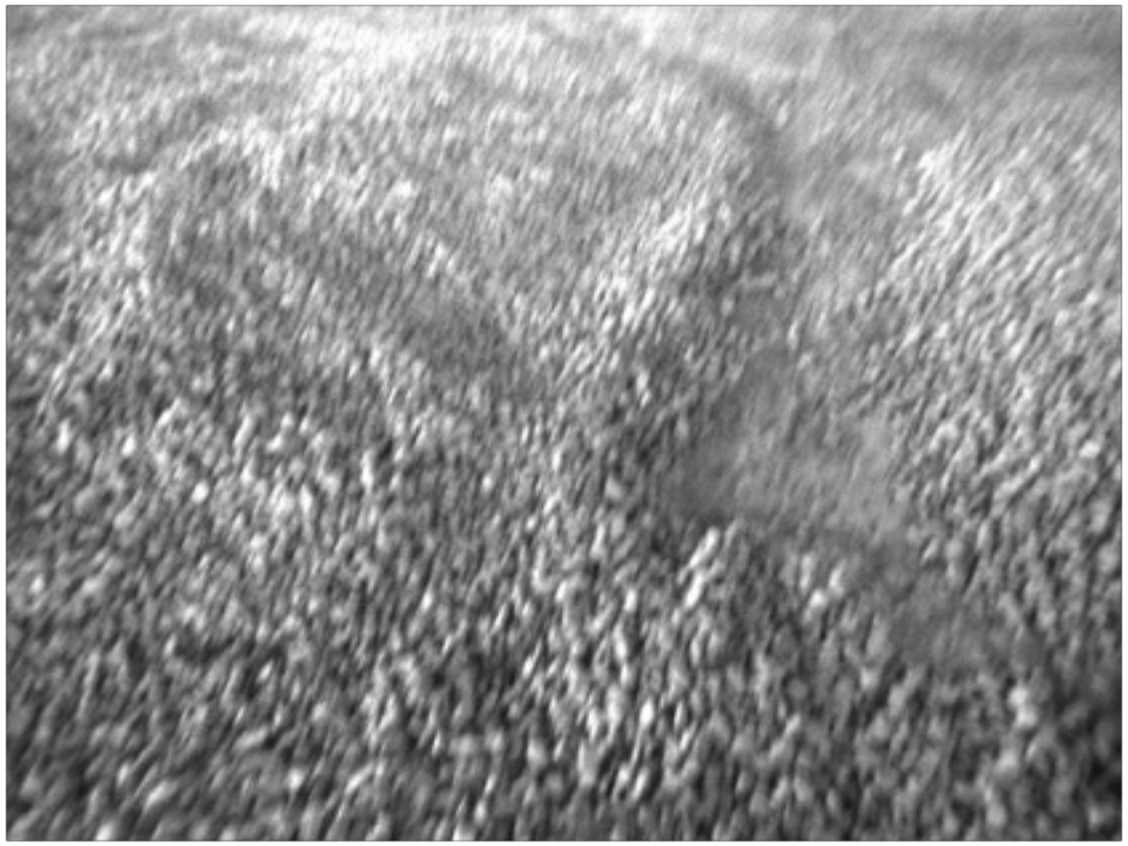
Trial	Route	Path length	Repeat speed	Local start time (UTC-4)			Autonomy rate	
				Teach	Mono	Stereo	Mono	Stereo
1	Outdoor	1370 m	0.6 m/s	09:56:46	10:35:10	12:08:30	99.71%†	100.00%
2	Outdoor	1360 m	0.6 m/s	11:45:40	12:22:26	13:43:49	99.88%	100.00%
3	Outdoor	1361 m	0.6 m/s	13:26:41	14:00:12	15:20:12	99.74%	100.00%
4	Indoor	126 m	0.3 m/s	13:32:23	13:40:53	14:02:46	96.28%	100.00%
5	Indoor	140 m	0.3 m/s	12:18:57	12:32:20	12:59:11	91.60%	100.00%

	Mono	**Stereo**
Total distance driven	4298 m†	4357 m
Total distance autonomously traversed	99.41%	100.00%

Why were monocular autonomy rates lower?

Not because of rough terrain!

Field Testing: Failure Cases



Motion Blur (Low Light)

Self-similar Terrain

Field Testing: Failure Cases

Motion Blur (Low Light)

Self-similar Terrain

+

High spatial uncertainty

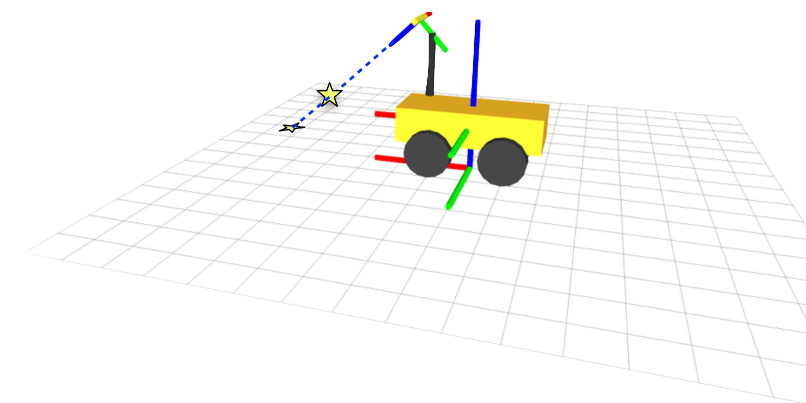
=

Bad feature matching!

Summary: Monocular Visual Teach & Repeat

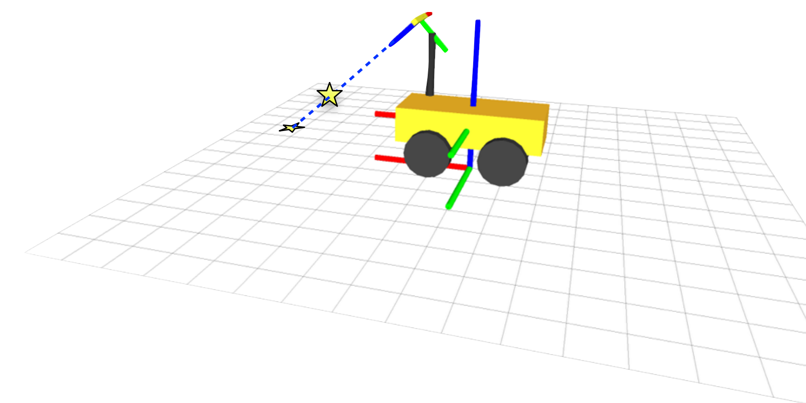
Summary: Monocular Visual Teach & Repeat

1 The flat-ground approximation works for high-accuracy monocular route following, given an appropriate uncertainty model.

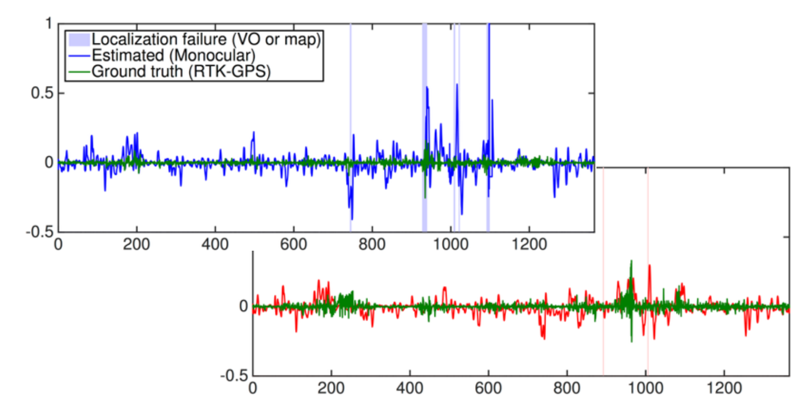


Summary: Monocular Visual Teach & Repeat

1 The flat-ground approximation works for high-accuracy monocular route following, given an appropriate uncertainty model.



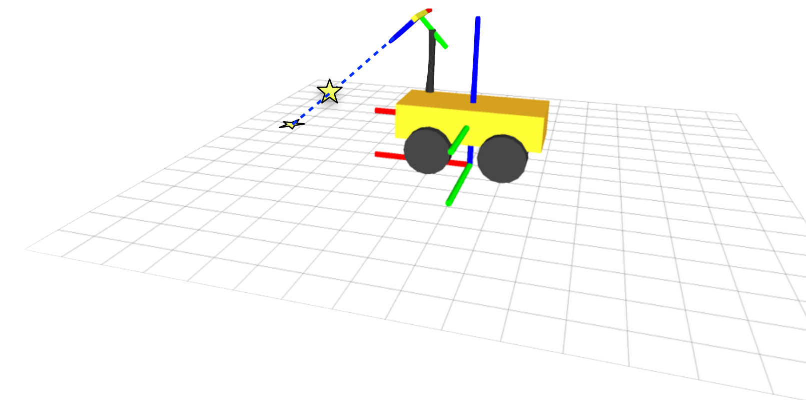
2 Monocular VT&R is just as accurate as Stereo VT&R, but less robust in places where feature matching is hard.



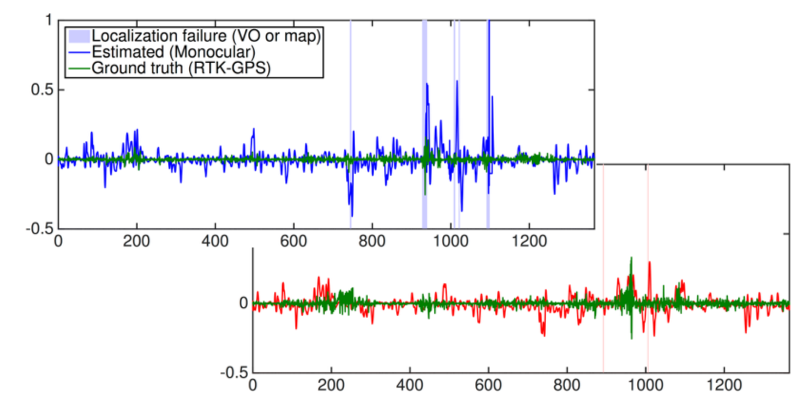
Institute for Aerospace Studies
UNIVERSITY OF TORONTO

Summary: Monocular Visual Teach & Repeat

1 The flat-ground approximation works for high-accuracy monocular route following, given an appropriate uncertainty model.



2 Monocular VT&R is just as accurate as Stereo VT&R, but less robust in places where feature matching is hard.



3 Existing monocular robots can now do repetitive navigation tasks autonomously, without additional sensors.

Thanks!
Questions?

Email: lee.clement@mail.utoronto.ca

Web: http://utias.utoronto.ca