Experimental Comparison of Visual-Aided Odometry Methods for Rail Vehicles

Florian Tschopp[1], Thomas Schneider[1], Andrew W. Palmer[2], Navid Nourani-Vatani[2], Cesar Cadena[1], Roland Siegwart[1], and Juan Nieto[1]

[1]Autonomous Systems Lab, ETH Zurich

[2]Siemens Mobility GmbH, Berlin

WHY ODOMETRY FOR RAIL VEHICLES

- **Rail vehicle localization** relies on **infrastructure-side Balises** (beacons)
- The rail segments are locked in **discretized large blocks**.
- Such a coarse locking leads to a **sub-optimal usage of the rail networks**.

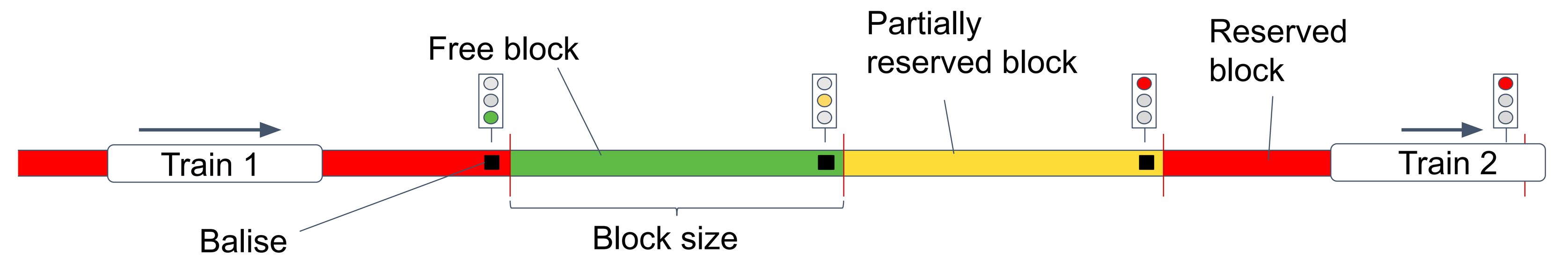

Fig. 1: Standard interlocking strategy on rail networks based on infrastructure-side positioning (and signalling). A block is released only if the trains position is reported leaving the block. If a block is reserved, the next train has to be able to stop in front of the block.

- Use of **moving blocks** without discretization to **increase the capacity** of the network.
- ➢ Requires **accurate and robust position and velocity estimation** of all vehicles.
- ➢ High level of safety by using **redundant and complementary sensors**.

VISUAL-AIDED ODOMETRY PIPELINES

Most promising available visual-aided odometry pipelines based on **filtering**, **sliding-window optimization** (SWO) and **batch optimization** are evaluated for rail applications.

Table I: Overview of visual-aided odometry approaches.

	Estimator type			Sensors			Comment
	EKF	SWO	Batch	Mono	Stereo	IMU	
ROVIO	x			x		x	Light-weight EKF using patch tracking.
VINS-Mono		x		x		x	Tightly coupled indirect monocular VI fusion.
Batch			x	x		x	Offline global batch VI bundle-adjustment.
ORB-SLAM2		x			x		Indirect stereo visual SLAM framework.
OKVIS		x			x	x	Keyframe-based tight coupling of stereo VI fusion.
Stereo-SWE		x			x	x	Tightly coupled VI fusion using depth as independent measurement.

EXPERIMENTAL EVALUATION

- Investigation of applicability, challenges, and limitations of current visual and visual-inertial motion estimation frameworks for rail applications.
- Evaluation against RTK-GPS ground truth on multiple datasets recorded in industrial, sub-urban, and forest environments.

Datasets

- `Trajectory 1` at low speeds ($\leq 25.5\,{}^{\mathrm{km}}/_{\mathrm{h}}$) in an industrial environment.
- `Trajectory 2` follows a public tram-line with speeds up to $52.4\,{}^{\mathrm{km}}/_{\mathrm{h}}$.

Collected using a custom-built stereo visual-inertial sensor (see Fig. 2) which is synchronizing all measurements in hardware:

- 2 x Basler acA1920-155uc ($2.3\,\mathrm{MP}$, global shutter, $20\,\mathrm{fps}$)
- ADIS16445 ($\pm 250\,{}^{\mathrm{deg}}/_{\mathrm{s}}, \pm 49\,{}^{\mathrm{m}}/_{\mathrm{s}^2}, 200\,\mathrm{Hz}, 0.011\,{}^{\mathrm{deg}}/_{\mathrm{s}\sqrt{\mathrm{Hz}}}, 0.001\,{}^{\mathrm{m}}/_{\mathrm{s}^2\sqrt{\mathrm{Hz}}}$)

Fig. 2: Sensor consisting of two cameras (stereo) and IMU.

Estimator Performance

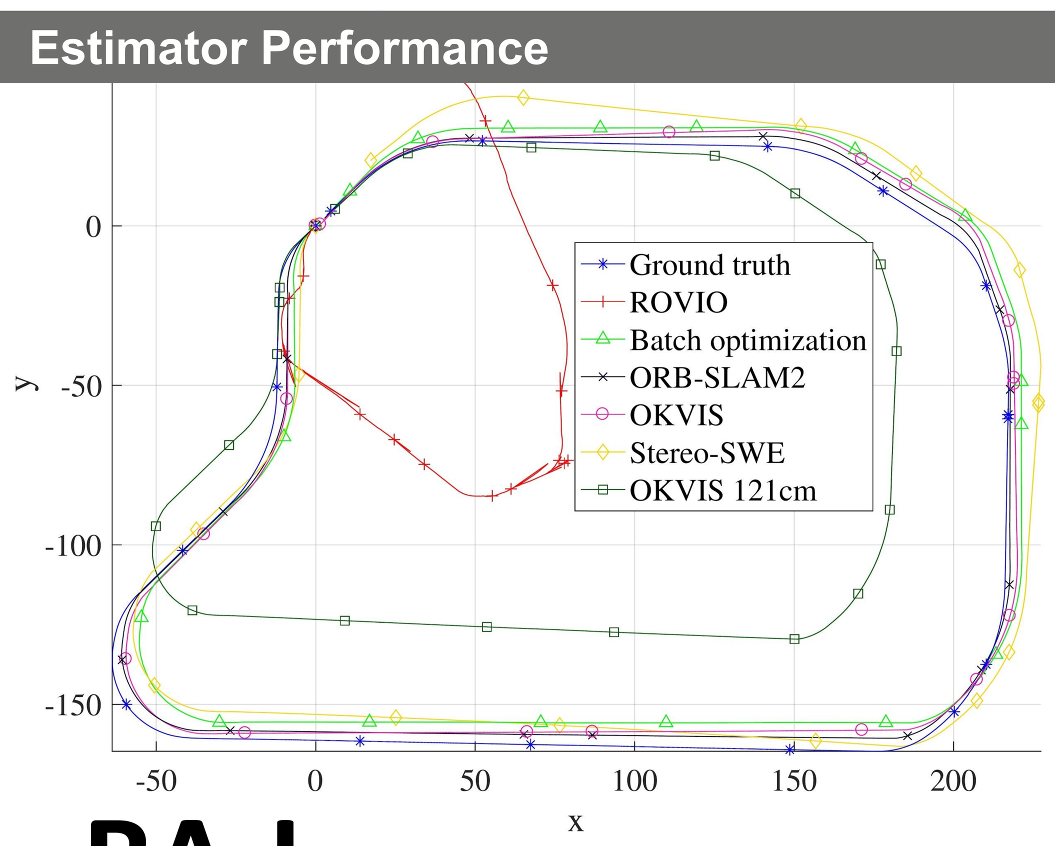

Fig. 3: Aligned paths of the different motion estimation pipelines on `trajectory 1`. Due to unrecoverable resets of the estimator, VINS-Mono is omitted here. If not stated otherwise, a 31 cm baseline is displayed.

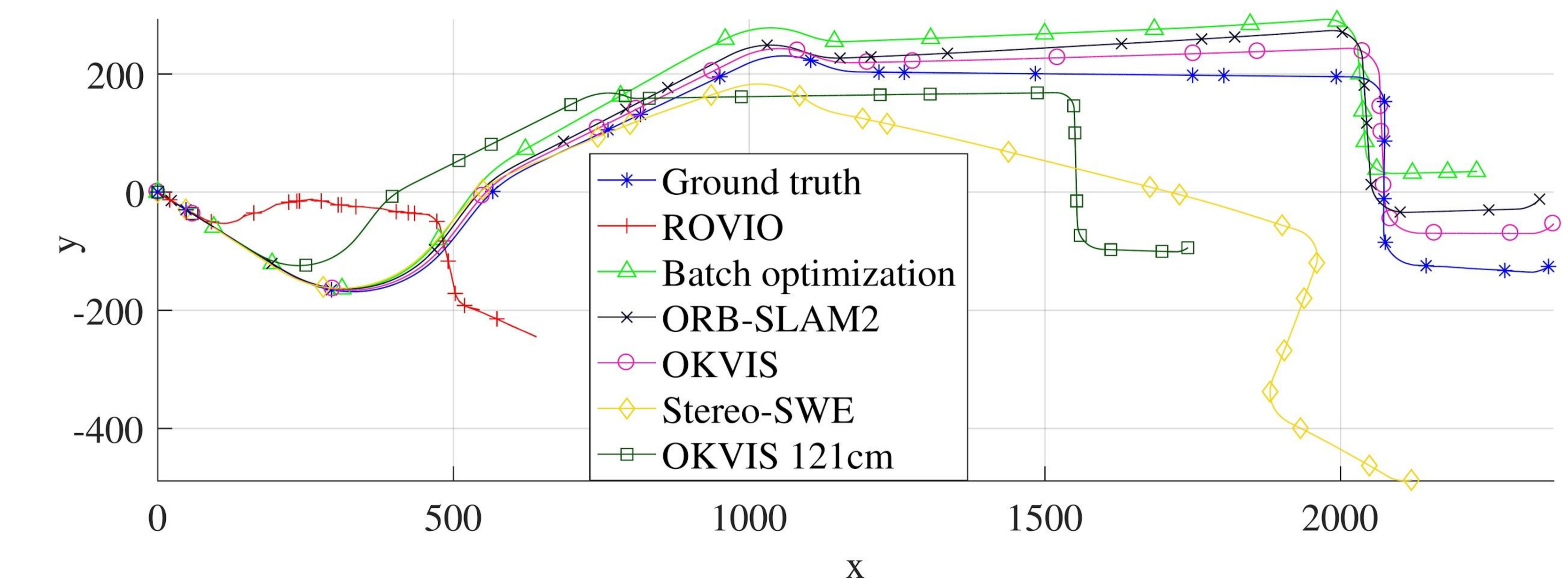

Fig. 4: Aligned paths of the different motion estimation pipelines on `trajectory 2`. Due to unrecoverable resets of the estimator, VINS-Mono is omitted here. If not stated otherwise, a 31 cm baseline is displayed.

- ROVIO and VINS-Mono fail to work properly due to **locally un-observable IMU biases** caused by constraint motion and constant velocity scenarios.
- Bias observability can be partially recovered using global bundle adjustment.
- Using stereo constraints, **metric scale is observable** enabling accurate motion estimation.
- Incorporating inertial information is beneficial to **increase accuracy and robustness** in challenging scenarios as described below.

Table II: Median estimation errors (distance in %, heading in deg/m). [1]Contains resets of the estimator.

	Trajectory	`Trajectory1`		`Trajectory2`	
	Segment length	10 m	50 m	10 m	50 m
Visual-inertial	ROVIO [2][1]	66.570/0.0490	67.723/0.0578	75.292/0.0269	75.149/0.0210
	VINS-Mono [3][1]	5.060 / 0.1033	10.589 / 0.4093	43.552 / 0.0408	45.339 / 0.0525
	Batch optimization [4]	7.092 / 0.0322	2.899 / 0.0066	12.361 / 0.0153	4.239 / 0.0084
Stereo visual	ORB-SLAM2 [5]	**2.138** / **0.0204**	3.054 / 0.0093	1.786 / 0.0078	1.829 / **0.0033**
Stereo visual-inertial	OKVIS [6]	2.152 / 0.0219	**2.850** / **0.0070**	**1.428** / **0.0074**	**1.110** / 0.0038
	Stereo-SWE [7]	2.845 / 0.0249	4.029 / 0.0128	3.710 / 0.0099	3.840 / 0.0087

Challenges

During our investigation, several challenging scenarios were observed:

- **A**: High speeds ($> 40\,{}^{\mathrm{km}}/_{\mathrm{h}}$) can cause problems for non-optimized feature tracking algorithms.
- **B**: Visual aliasing is prominent especially if no inertial information is available.
- **C**: Fast changing illumination and reflections are challenging for all visual pipelines.

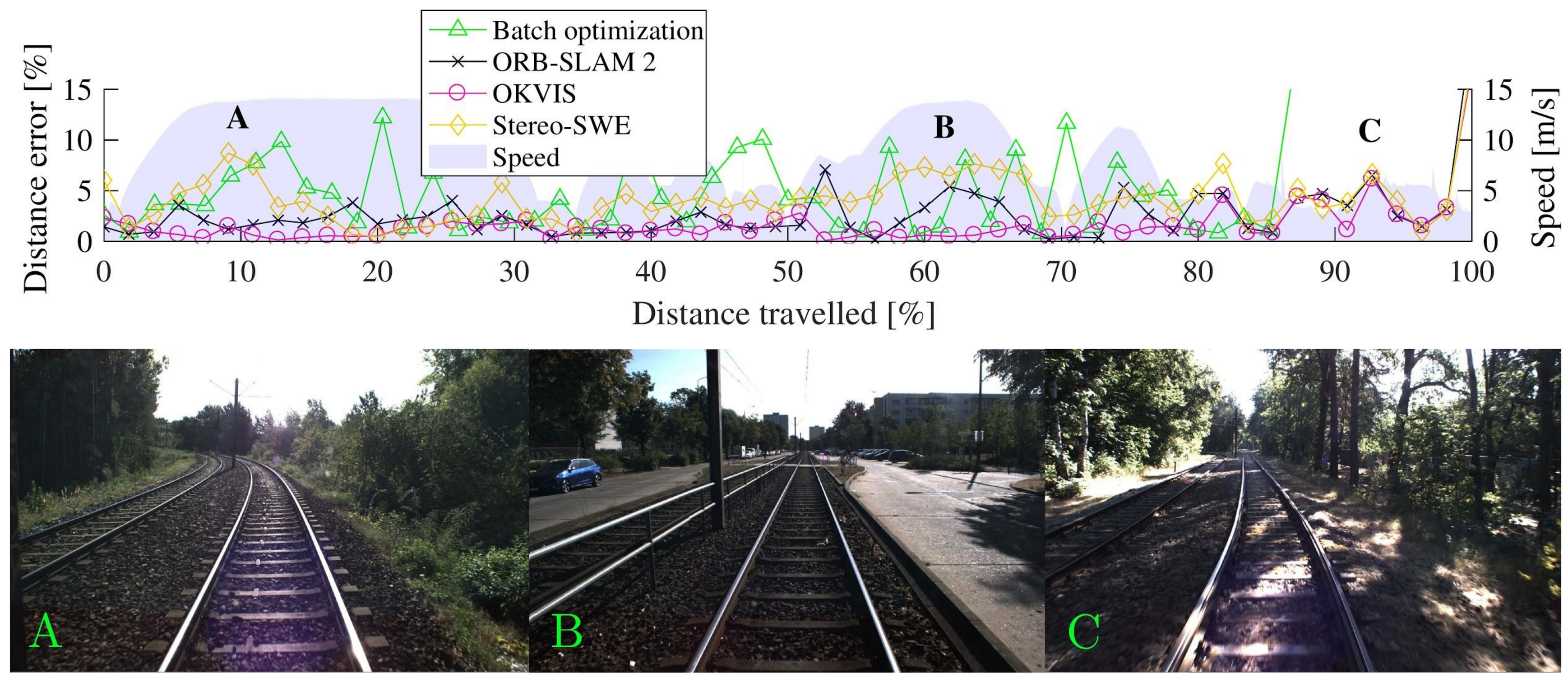

Fig. 5: Top: Errors of the best performing pipelines during `trajectory2`. The letters indicate selected challenging scenarios. Bottom: Camera images of the challenging scenarios.

CONCLUSIONS

- **Precise and robust position and velocity estimation** is essential to **increase railway capacity** due to coarse interlocking strategies.
- Even without enforcing specific motions, visual-aided odometry methods can achieve **high accuracy odometry** on rail vehicles.
- **Monocular VIO fails** due to locally unobservable IMU biases, stereo vision enables **consistent motion estimation**. Inertial information increases **robustness and accuracy**.
- Visual-aided odometry in isolation is **not reliable enough** for safety critical application but can complement for failure cases of other sensors (such as GPS, wheel odometry, ...).

ACKNOWLEDGMENTS

This work was supported by Siemens Mobility, Germany. The authors would like to thank Andreas Pfrunder for his help in initial data collections and evaluations.