UP-Drive

Automated Urban Parking and Driving

Research and Innovation Action - Horizon 2020
Grant Agreement Number 688652
Start date of the project: 01-01-2016, Duration: 48 months

Deliverable D 5.2

First development and integration cycle of lifelong mapping

Type: R

Status: Revision 1

Lead contractor for this deliverable: **ETHZ**

Due date of deliverable: 30.06.2017
Preparation date: 20.06.2017
Actual submission date: 05.07.2017

Coordinator: Wojciech Derendarz, VW

Project co-funded by the European Commission within HORIZON 2020		
Dissemination Level		
PU	Public	X
PP	Restricted to other programme participants (including the Commission Services)	
SEAM	Restricted to partners of the SEAM Cluster (including the Commission Services)	
RE	Restricted to a group specified by the consortium (including the Commission Services)	
CO	Confidential, only for members of the consortium (including the Commission Services)	

EXECUTIVE SUMMARY
This deliverable describes the lifelong mapping framework after the first development & integration cycle. All components, notably the metric and semantic map, the metric online localization, the semantic data aggregation and the map summarization are functional and integrated on the vehicles, fulfil their basic purposes and interact with each other in a limited fashion. All components deliver first evaluation results.

CONTRIBUTING PARTNERS		
	Company/Organisation	Name
Document Manager	**ETHZ**	
Partner 1	ETHZ	Cesar Cadena
Partner 1	ETHZ	Mathias Buerki
Partner 2	IBM	Ralf Kaestner

Table of Contents

1 Introduction

The goal of WP5 "Lifelong Localization & Mapping" is to provide a compact, customized, and metrically accurate map representation with which the vehicles can localize reliably in long-term and large-scale operations in urban environments, and where semantic data can be associated, stored and queried.

In this deliverable, the progress made in WP5 during the first specification and integration cycle is summarized, and the evaluation concept, together with appropriate metrics, is described. Where already possible, preliminary results are presented.

2 Summary of specifications and requirements

During the first development and integration cycle of lifelong mapping the responsible partners of the work-package (ETHZ and IBM) have focused mainly on framing and defining the requirements and specifications for both the localization and mapping frontends. This was done according to the expected use-cases and the scope of proposed target scenarios, and driven by thorough discussions between the partners. A complete derivation of these specifications and use-cases can be found in D5.1. To summarize, the following high-level tasks were completed:

- Designing the basic architecture for internal map representations.
- Designing interfaces between the cloud-based mapping backend and location queries (both metric and semantic).
- Defining which functionalities are to be available in the mapping backend.
- Defining both online and offline data exchange, as well as dataset archiving.

3 Summary of what is implemented and available

At the functional level, WP5 has implemented the first version of the metric localization (Task 5.4) and of the reference frame alignment (Task 5.2).

3.1 Metric Localization (ETHZ)

This framework is able to provide online localization capabilities, returning 6DoF transformations between the vehicle body frame and a map-based fixed reference frame. In addition to retrieving the vehicle's current pose, the ability to retrieve metric position estimates of historical vehicle poses and mapped landmarks is also possible. Furthermore, the framework is capable of locally storing maps after creation, and has the ability to augment these existing maps with newly acquired data.

3.1.1 Coordinate-Frames

Before presenting the evaluation results, all involved coordinate-frames are properly introduced in this subsection and illustrated in Fig. 3.1.1

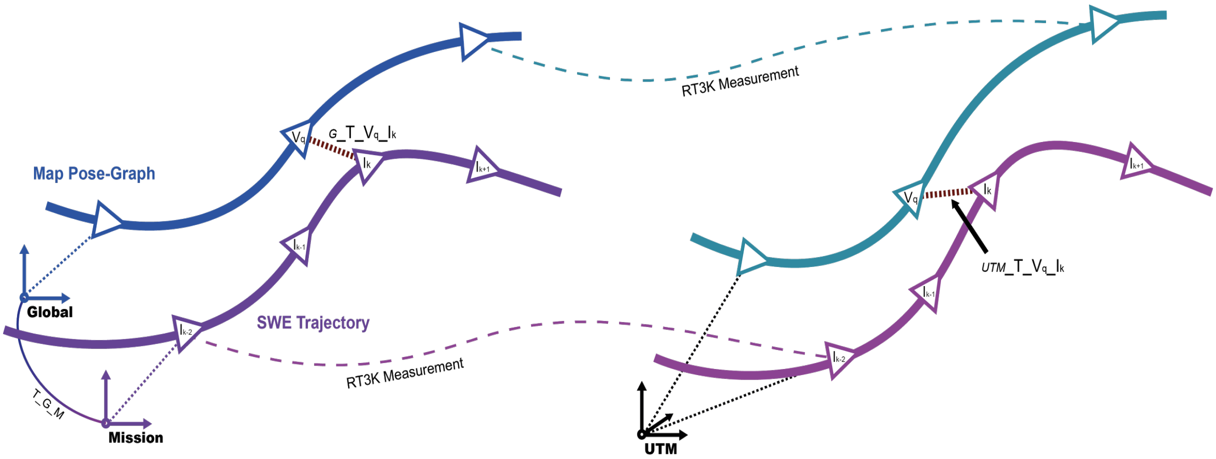


Figure 3.1.1: Schematic of the relationships between reference frames when evaluating the current localized (SWE) trajectory on the previous map (Map Pose-Graph) with the RT3K Measurements.

All data in the map (vertices on the pose-graph, landmarks, etc.) can be expressed with respect to a map-fixed coordinate frame, denoted as “Global” (**G**). It is fix in the map, but can conceptually be chosen to lie anywhere in 3D space.

In addition to that, the reference coordinate frame for the localization system on the car is denoted by “**I**”, as it coincides with the body coordinate frame of the IMU sensor mounted in the car. While localizing a dataset against a mission, the current pose of the car is estimated and expressed wrt. a coordinate frame denoted by “Mission”(**M**). Hence, the corresponding transformation is denoted by **T_M_I** and expresses vectors expressed wrt. coordinate frame “**I**” in coordinate frame “**M**”.

This “Mission” coordinate frame serves as a locally smooth (jump-free) reference frame for the pose estimates of the Sliding-Window-Estimator (SWE). The transformation between the “Global” frame and the “Mission” frame is co-estimated together with the transformation **T_M_I** as soon as the car is localized against the map for the first time.

For every key-frame (usually corresponding to an image(s) acquisition), a pose-estimate (**T_M_I**k) is estimated, and in case the dataset is saved as a map session, a vertex (Vq) is created in the map pose-graph. For every such pose-estimate or vertex, a corresponding GPS measurement from the high-precision RTK/DGPS sensor is available. These measurements contain a 6DoF transformation **T_UTM_B** between the GPS sensor (“**B**”) and the UTM coordinate frame of reference for the corresponding UTM zone (in our case: 32N).

3.1.2 Subsystem Integration Status

The following subsystems have been implemented on the project's vehicle (Wolle) and a vehicle located in Zurich (Kermit) with a similar sensor setup.

- Sensor data integration: Integration of multi-camera (TopView) system, IMU and wheel odometry. Including intrinsic and extrinsic calibration (Fig. 3.1.2).

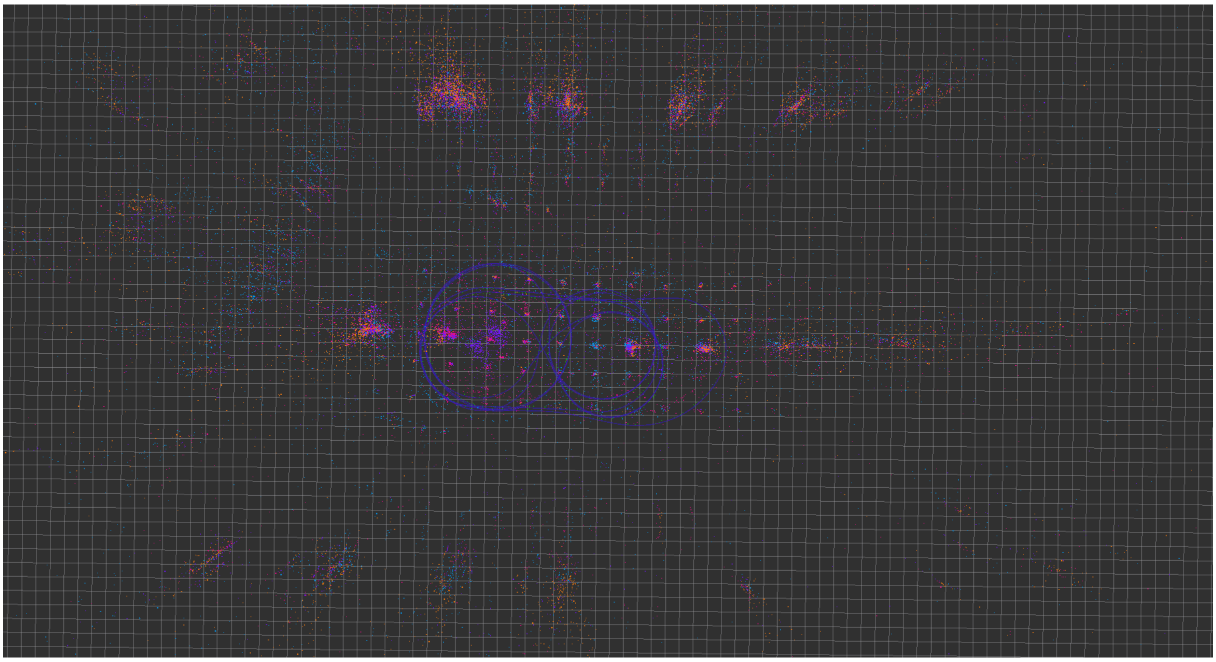
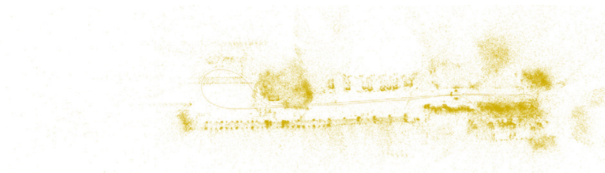
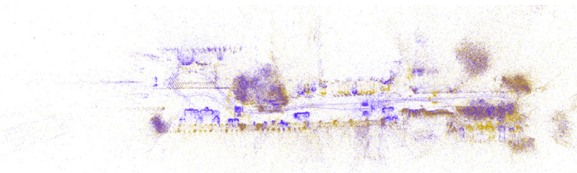


Figure 3.1.2: Bird's eye perspective of the 3D point-cloud map after extrinsics and intrinsics calibration. The different colors show the landmark positions from the four Top-View fish-eye cameras.

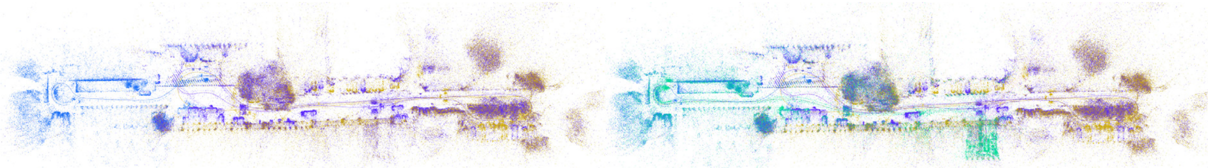
- Multi-Session Mapping: The capability to compute a unified consistent and accurate metric map from multiple dataset recordings under different appearance conditions is depicted in Fig. 3.1.3. A full geometrically consistent map is built after closing loops from a single dataset on a cloudy day (Fig. 3.1.3a). Subsequent datasets (i.e. sessions) were recorded on a sunny day, localized and registered against the pre-existing map (Fig. 3.1.3b - Fig. 3.1.3d). Fig. 3.1.3c and 3.1.3d further illustrate the capability to extend the spatial area of the map with new territory.



a. Map after bundle adjustment of the first session (07.06.2017) containing one loop closure after. Cloudy weather.



b. Second session (09.06.2017) localized on the previous map and bundle adjusted after a loop closure. Sunny weather.



c. *Third session (09.06.2017) localized and augmenting the map with a new explored area (blue).*

d. *Four sessions (09.06.2017) continues extending the map and closes the loop when revisiting back to the garage (cyan).*

Figure 3.1.3: Multi-session mapping over different days and weather conditions. Different colors denote different sessions.

- Localization: Online metric localization of a dataset against a map. In Fig. 3.1.4 we show the estimated trajectory of the Sliding-Window-Estimator (SWE T_G_I) expressed wrt. the map-fixed coordinate frame (G). In addition to that, the pose-graph (T_G_V*q*) of the map is shown in green.

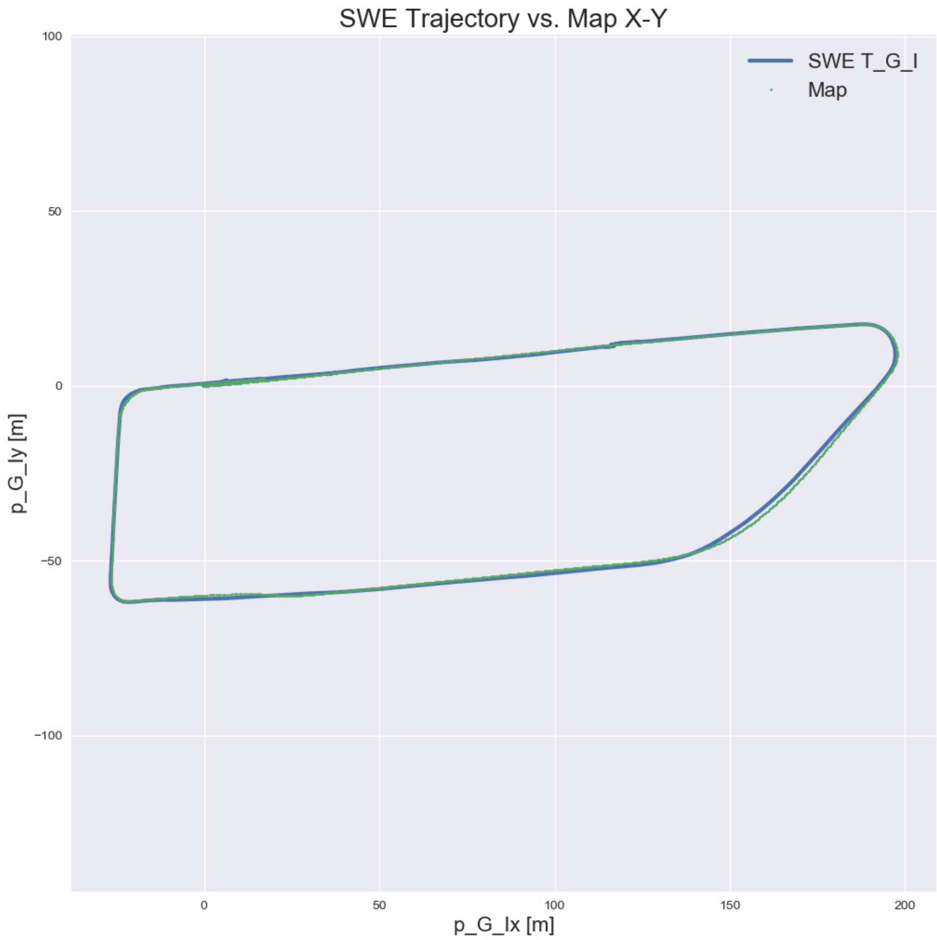


Figure 3.1.4: Estimated trajectory (blue) localized against a previous map. The pose-graph of the map is shown in green.

3.2 Reference Frame Alignment (IBM)

The task of aligning the different coordinate frames between the semantic map representation with the metric map has started. Two general types of coordinate frames are maintained within the project:

- Map data that is referenced with respect to GPS, usually expressed in WGS84. The highly accurate dynamic lane models (DLMs) provided by VW are specified in this frame. Certain classes of semantic information, such as road / lane markings, traffic signs and parking spaces are also partially available in this frame.
- Map data that is referenced with respect to metric visual odometry. Certain classes of semantic information, such as road / lane markings, traffic signs, dynamic objects as reported by on-board sensors are translated into that frame.

For the purpose of data fusion and online decision making (i.e., driving decisions), a transform estimate between these two frames is required. This estimate is maintained at a local level, which ensures maximal local overlap without resorting to the costly and unnecessary deformation of the involved frames further at a global level. This design choice has far reaching consequences and in particular serves as a facilitator of lifelong mapping and continuous atomic updatability of the overall map framework.

We have furthermore developed the following components to facilitate development in the coming project phases:

- Writing a parser for VW's DLM format and translation into a standard XML-serializable format using the Boost graph library.
- Tools for visualizing this standard format in ROS' RViz visualizer.

3.2.1 Approach

In a nutshell, our reference frame alignment approach involves the following sequential steps:

- Instantiate the semantic map representation based on the XML Schema exchanged with VW (see Fig. 3.2.1).
- Construct a road graph based on the semantic map’s connectivity of driving lanes (see Fig. 3.2.1and Fig. 3.2.2).
- Acquire (or simulate) the following observations:
  - Highly noisy GPS measurements (in WGS84 frame)
  - An odometry with incremental positioning and orientation errors (in local frame, see Fig. 3.2.3)
- Solve a non-linear optimization problem subject to the following cost constraints:

- GPS-to-odometry position error (point-to-point distance)
- Road graph-to-odometry position error (point-to-line distances for the K nearest road graph segments)
- Odometry-to-odometry incremental pose update error (expressed in SE(3))

The results of the alignment are represented in UTM frame.

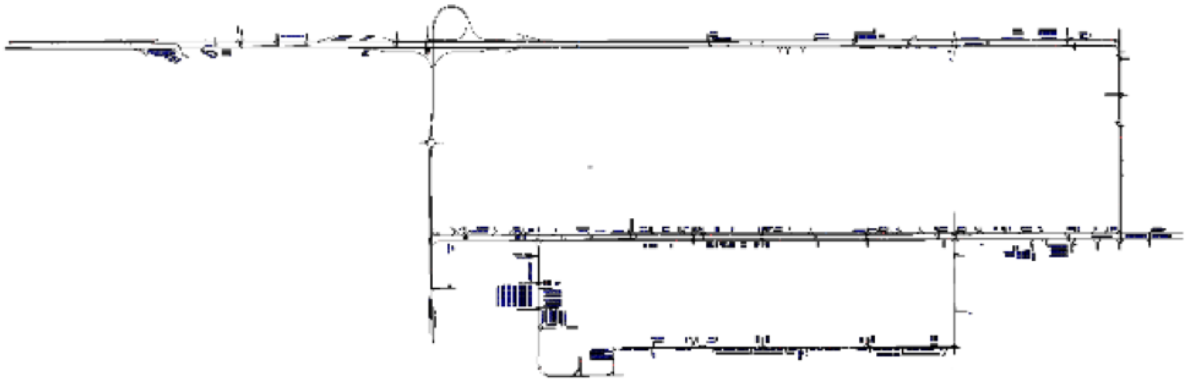


Figure 3.2.1: Semantic road graph visualization in RViz. The overall depicted graph represents a map area of approximately 3 km x 1 km.

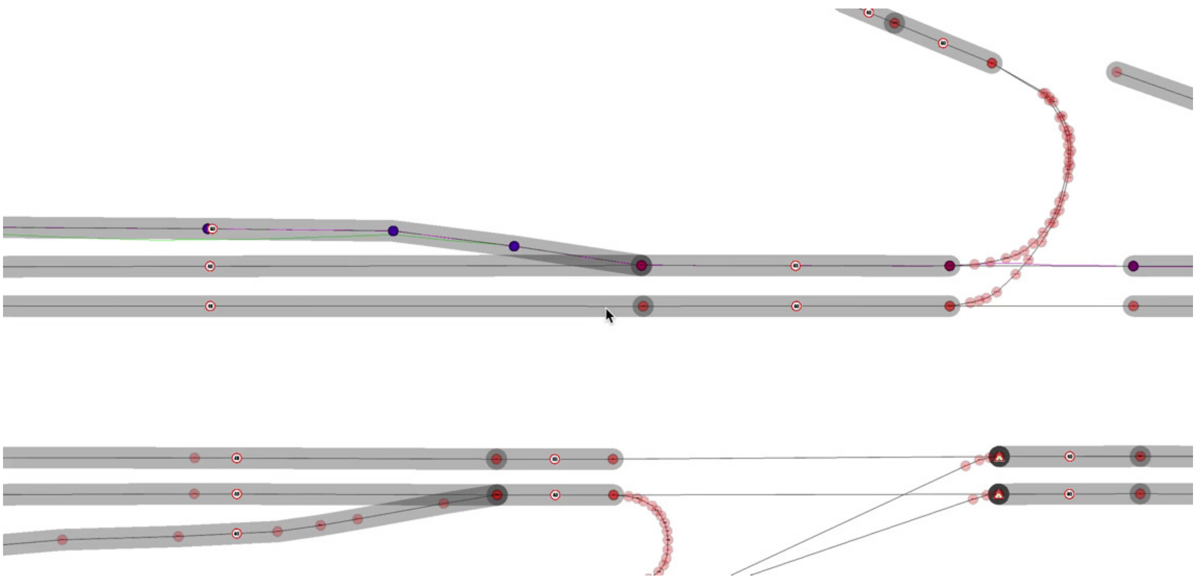


Figure 3.2.2: Detailed view of the semantic road graph visualization in RViz. Lines represent lanes, and red circles correspond to geographic positions defining a lane. Line sections shaded in gray denote semantic waypoints on the lanes, such as speed limit areas, traffic lights, and pedestrian crossings.

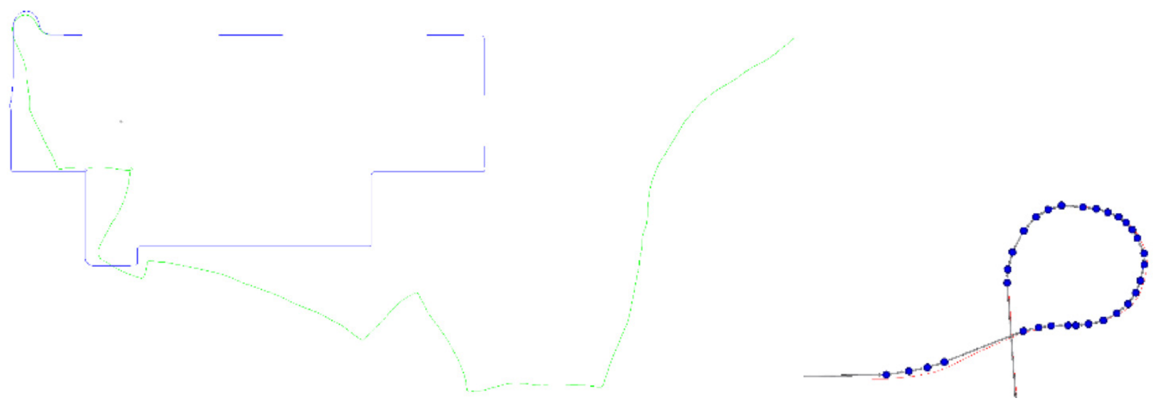


Figure 3.2.3: Ground truth odometry sampled from the road graph (blue line) and simulated odometry (green line), modeling both translational and rotational errors. From afar, the aligned odometry (red line) perfectly overlays the ground truth. A detailed view of the aligned odometry is depicted on the right, revealing sub-meter divergence from the ground truth in some places.

3.2.2 Expected Integration

The envisioned point of integration of our reference frame alignment approach is the generally unknown transform link between the semantic map and the metric map used for localizing the vehicle: by timestamp, the alignment solution can thus be queried for a transform estimate between the maps which enables the following specific applications:

- Anchor observations and their inference results (e.g., the availability of parking spots) in the semantic map (requires the transform from the metric map in local map frame to the semantic map in WGS84 frame).
- Represent priors from the semantic map (e.g., the expected occurrence of a pedestrian crossing) in the local navigation frame (requires the transform from the semantic map in WGS84 frame to the metric map in local map frame).

4 Evaluation

4.1 Evaluation for Metric Localization (ETHZ)

In this section we describe the criteria used for evaluation of the metric localization system performance. These criterias are expressed in terms of key-performance-indicators **KPI**s.

The following table describes our **KPI**s.

KPI Name	Description
#Inliers	Number of geometrically consistent matches between the keypoints in the live-images and the landmarks in the map. Evaluated on a per-keyframe basis. For every live-image, keypoints are extracted and BRISK descriptors are computed. These features are then matched against landmarks in the map, before inlier matches are distinguished from outliers using RANSAC.
LRLE	Local relative localization error. For each localized key-frame *k*, the transformation between the estimated pose of the vehicle and the nearest vertex *q* in the pose-graph of the map is computed and denoted by **SWE_T_nnV***q*_**I***k*. This transformation can be compared with the transformation between the current RTK DGPS measurement (**T_UTM_I***k*) and the RTK DGPS measurement of vertex *q* (**T_UTM_nnV***q*), denoted by **RT3K_T_nnV***q*_**I***k*. An illustration of these transformations is depicted in Figure 3.1.1.

The first KPI refers to the number of matched inliers. In Fig. 4.1.1 we show the resulting number of inliers when localizing a dataset and the associated localization success counting. Localization success is defined depending on a number of parameters, such as minimum number of matches, a RANSAC reprojection threshold, and minimum number of inliers.

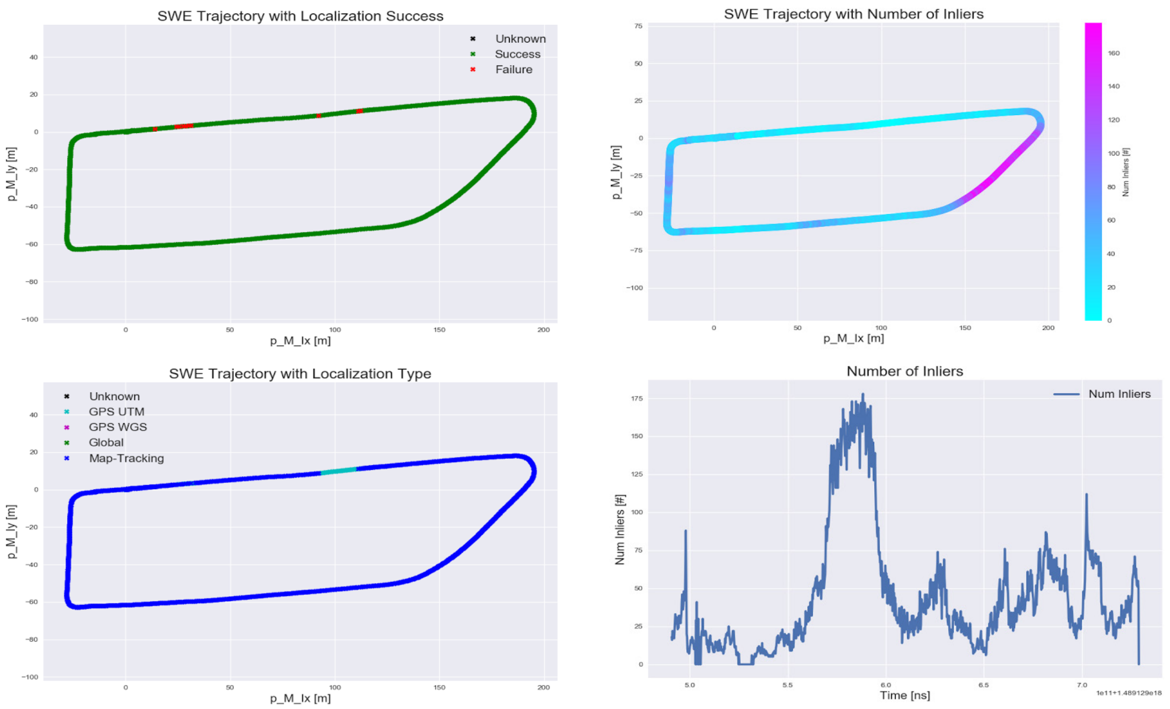


Figure 4.1.1: on the left, current trajectory depicting successful/failed localization attempts (top) and target map reference (bottom). On the right, the number of matched inliers per frame is shown on the trajectory (top) and over the time of the mission (bottom).

For the second KPI (localization error) we show the bird's view perspective of the following trajectories, all expressed in the UTM coordinate frame:

*RT3K*_**T_UTM_B***k*	RT3K measured
*SWE*_**T_UTM_B***k*	*RT3K*_**T_UTM_nnV** * **T_nnV_G** * *SWE*_**T_G_M** * *SWE*_**T_M_I***k* * **T_I_B**

T_I_B denotes the pre-calibrated transformation between the IMU and RT3K sensor-body coordinate frames.

T_nnV_G denotes the transformation between the map reference coordinate-frame and the nearest-neighbor (wrt. *SWE*_**T_G_I***k* = *SWE*_**T_G_M** * *SWE*_**T_M_I***k*) pose-graph vertex.

The blue trajectory labeled "RT3K Map" denotes the RT3K measurements (*RT3K*_**T_UTM_V***q*) of all the vertices of the pose-graph.

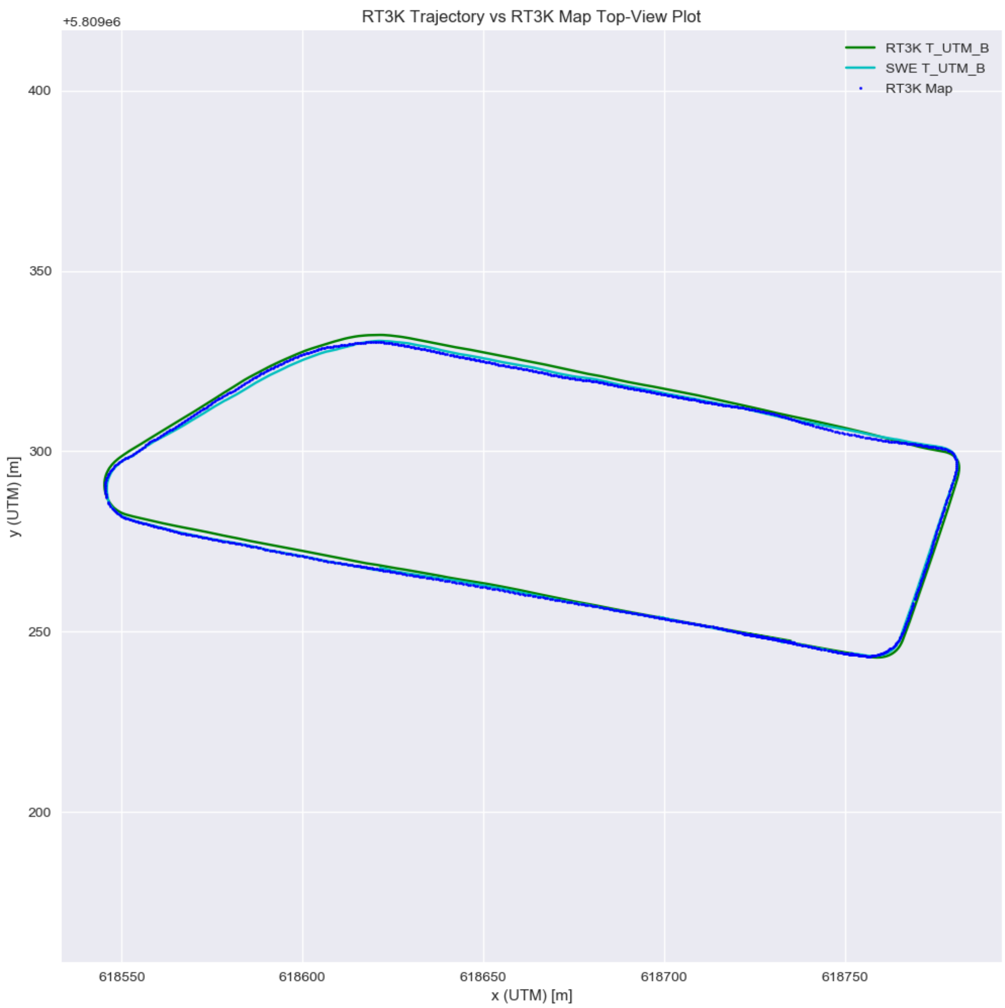


Figure 4.1.2: Current localized trajectory (cyan), the corresponding current RT3K measurements (green) and the RT3K measurements of the map (blue).

Fig. 4.1.3 finally shows the local relative localization error based on the following transformation: **T_lrle** = *SWE*_**T_l***k*_**nnV** * *RT3K*_**T_nnV_l***k.*

This transformation denotes the local difference in translation and orientation between the estimate from the Sliding-Window-Estimator wrt. the nearest neighbor vertex in the pose-graph, as compared to the estimate from the RT3K DGPS sensor, also wrt. the same nearest neighbor vertex in the pose-graph.

This transformation (**T_lrle**) can be computed for every localized key-frame. From this set of transformations, root-mean-squared errors are computed for the translation (x, y, as well as x,y,z) and the orientation (L2 norm of the axis-angle representation).

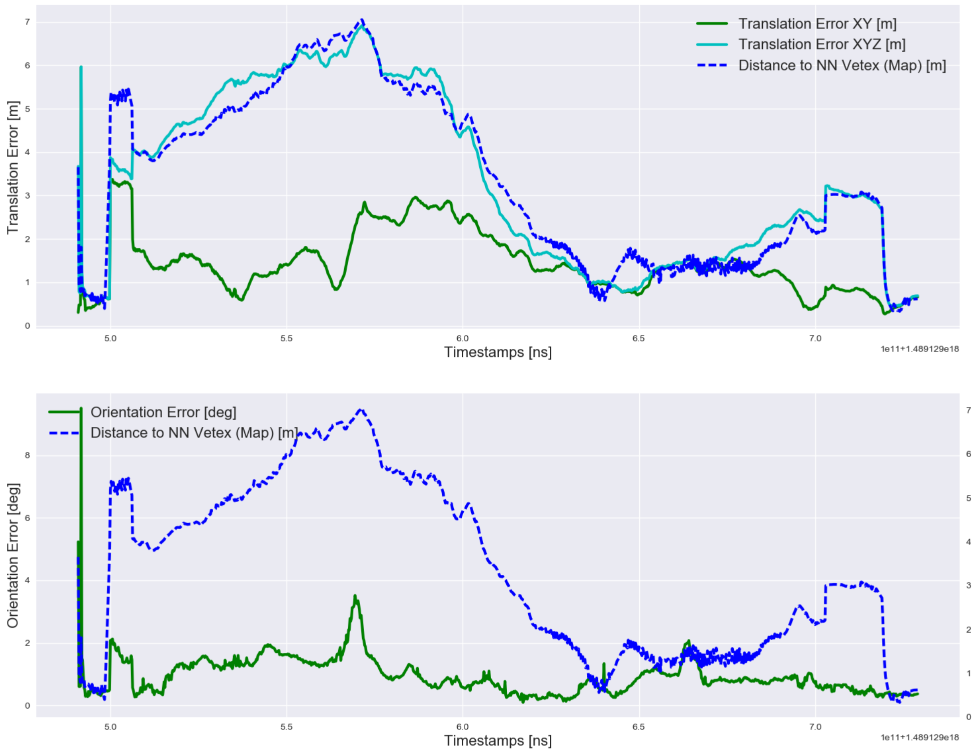


Figure 4.1.3: Localization errors over time of the mission in the XY-plane, XYZ, and orientation.

The example dataset evaluated in Fig. 4.1.3 shows orientations errors in the expected range of < 2 degrees. The errors in translation, however, exceed the expected range by far, also in the 2D plane (x,y). We believe this error to mainly be due to the inconsistent biases of the UTM pose measured by the RT3K sensor. This is also supported by Fig. 4.1.2, which shows the Sliding-Window-Estimator to localize the car to the right side of the pose-graph, while the RT3K measurements show the car to be to the left side of the RT3K measurements recorded for the vertices of the pose-graph. We plan to evaluate these metrics on a far wider range of example datasets and environments in order to support our claim about Fig. 4.1.3 and to be able to retrieve actually meaningful and trustfully accuracy values also for the translational errors. However, at the time of writing this document the datasets required for this task have not been available yet, and further evaluations have thus to be postponed until sufficient data will be available.

4.2 Evaluation for Reference Frame Alignment (IBM)

Evaluation of the alignment results posed a difficult challenge as absolute ground truth was generally unavailable: The objective of the alignment was to minimize the divergence between the semantic road graph and the path taken by the vehicle. Hence, we considered the road graph as relative ground truth and evaluated the alignment results by means of the residual alignment error. This could only be performed in simulation due to the ambiguity of the errors inflicted by real world measurements: For a similar evaluation involving reference GPS observations, the vehicle would have to drive exactly on the lanes prescribed by the road graph, which was impractical to achieve.

In our simulated experiments, we generated the following observations from the road graph representation:

- A connected path of sequential road graph positions in WGS84 frame which we assumed the vehicle to follow.
- A sequence of (relative) ground truth GPS measurements consisting in equi-distant vehicle positions in WGS84 frame, subsampled from the road graph path.
- A (relative) ground truth odometry consisting in equi-distant vehicle positions in the local odometry frame, subsampled from the road graph path.
- A sequence of simulated GPS measurements obtained by adding random noise to all components of the GPS ground truth.
- A simulated odometry obtained by adding random noise to the translational and rotational components of the ground truth odometry updates (motion model).

In order to simulate the alignment performance for a true experimental setup, we chose rather pessimistic error parameters, adhering to the degraded performance of consumer-grade GPS and odometry sensors:

- GPS latitude/longitude standard deviation of 3*10-5 deg, altitude standard deviation of 1.5 m
- Odometry distance standard deviation of 0.1 m/m, rotation standard deviation of 1.0 deg/m

The typical assumed path length of the vehicle amounted to approximately 6 km.

To quantify the alignment results, we calculated the alignment error as the mean and standard deviation of the point-to-point distance between the simulated and the (relative) ground truth odometry, where we found

- A mean offset of about 0.4 m (mostly longitudinal, which can be explained by the point-to-line distance constraints of the optimized cost function).
- A standard deviation of about 0.3 m.
- Locally bound errors due to “lane switches” in cases where road graph lanes are close to each other (e.g., a lane diverges from or converges to another lane). Avoiding such errors could be achieved by cross-checking the optimized odometry for topological consistency with the road graph, under loss of generality and added complexity (no analytical solution) of the implemented approach.

5 Summary and outlook

In this document, we have presented the lifelong mapping results of the 1st development & integration cycle. The implementation progress has been described and the evaluation concept along with the appropriate metrics summarized. The following results have been achieved during the 1st development phase:

- Requirements and specification for the localization and mapping software layers, data structure and interfaces have been derived.
- All sensors used for localization and mapping on the vehicle platform have been calibrated and can collect datasets for localization and mapping on regular basis.
- Initial version of all metric mapping components and localization have been fully integrated on the vehicle and deliver first evaluation results.
- Initial internal map representations and interfaces between metric and semantic map components have been specified.

Following the spiral life-cycle model, the final lifelong mapping framework after the 2nd development & integration cycle will be described in D5.3. All components will reach full maturities and the results of each module's final performance evaluation will be presented in detail. WP5 goals for the upcoming period are as follows:

- Integration of localization and mapping software stack on IBM cloud infrastructure.
- Further improvement of the localization software, e.g. multi-session mapping and offline summarization.
- Further improvement and full integration of global localization, robustness and scalability.
- Integration of efficient queries to the semantic map.