FloMuSS – Fleet-Based Multi-Sensor System for the Continuous Acquisition of Spatially and Temporally High-Resolution Data of the Urban Streetscape

Siyu CHEN[1], Christoph EFFKEMANN[1], Dr.-Ing. Ralf BECKER[1], Ph.D. Lorenzo FERRONI[2], Dr.-Ing. Johannes LUDWIG[2], Ingmar SEITZ[3], Dr.-Ing. Conny LOUEN[3], Sajjad TABATABAEI[4] & Univ.-Prof. Dr.-Ing. Jörg BLANKENBACH[1]

[1] RWTH Aachen University, Geodätisches Institut und Lehrstuhl für Bauinformatik & Geoinformationssysteme, blankenbach@gia.rwth-aachen.de

[2] eagle eye technologies Deutschland GmbH

[3] RWTH Aachen University, Lehrstuhl und Institut für Stadtbauwesen und Stadtverkehr

[4] GELSENWASSER AG

DOI: 10.3217/978-3-99161-070-0-028, CC-BY4.0

1 Introduction

Rapid urbanization places increasing demands on cities as well as municipalities regarding transport planning, traffic and flood risk management, environmental monitoring, and infrastructure maintenance. All these tasks require comprehensive, up-to-date information of the urban area. Developing adaptive solutions to tackle these challenges necessitates accurate, comprehensive, georeferenced geometric-semantic road-space data as well as spatially resolved environmental observations (e.g., particulate matter (PM) for air quality, precipitation level, noise profiling) across the city.

However, current analog and spatial data sources are often sparse and outdated due to infrequent and fragmented update cycles. In the context of road-space mapping, available sources typically capture static features (e.g., manhole covers, trees, curbs) while omitting important areas such as parking lots, green strips, sidewalks, bike paths, and street furniture. Concurrently, environmental data are frequently collected by permanently installed measurement stations, which provide high-quality point measurements but do not adequately represent spatial variability across the entire urban area. Conventional data collection methods typically operate at discrete intervals (e.g., surveys every few years). However, they are not capable of maintaining the spatial and temporal resolution required for modern urban management and concepts such as Digital Twins. More cost-efficient approaches with higher temporal resolution would be desirable. Consequently, this motivates the development of an innovative and comparatively cost-effectively multi-sensor system for the systematic, regular and georeferenced collection of urban road-space and environmental data.

The idea behind the FloMuSS (Fleet-Based Multi-Sensor System) presented in this article is a low-cost sensor platform that should be able to be mounted on municipal service vehicles (e.g., waste collection trucks) that traverse the entire road network regularly, thereby ensuring

comprehensive coverage of the urban area and high-frequency updates (**Fig. 1**). The platform should allow for the installation of various sensors to enable georeferenced data collection of a wide range of parameters in urban areas.

In the underlying research project for the development of FloMuSS, we examine three exemplary use cases.

1. Streetscape monitoring: Capturing and continuous updating of transport infrastructure to support future-oriented road planning and targeted traffic management
2. Parking management: Improving the determination of parking space availability for dynamic traffic and parking space management
3. Pluvial flood risk management: Enhancing precipitation runoff models and predictions of flood hotspots for preventive heavy precipitation management (e.g., consideration of break lines in road space)

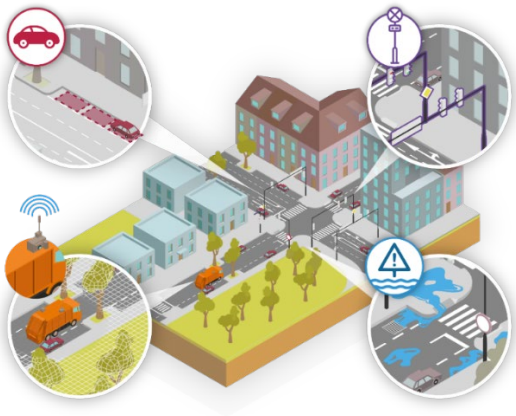

Fig. 1: Exemplary use cases in FloMuSS

Effective utilization of captured data requires precise georeferencing 6-Degrees-of-Freedom (DOF) pose of the sensors over time. For a moving vehicle acting as a sensor carrier, this necessitates the calculation of a time-continuous 6-DoF trajectory. This article focuses on calculating this trajectory using low-cost sensors in complex urban environments. The contribution of this paper is fourfold. First, we derive and formalize the requirements for a low-cost, vehicle-mounted multi-sensor system for urban streetscape acquisition in terms of positioning accuracy, spatial coverage, temporal resolution, and economic constraints. Second, we detail the system design and hardware integration, including sensor configuration and installation on municipal vehicles. Third, we present a robust real-time sensor fusion method that integrates Global Navigation Satellite System (GNSS), Inertial Measurement Unit (IMU), and images under degraded operating conditions such as prolonged GNSS signal loss and IMU data gaps. At last, we report results from test drives, analysing the quality of the estimated trajectories and dense point clouds under representative urban conditions and demonstrate the system's applicability of the collected data for selected use cases.

2 Related Work

Since this article focuses on the usability of mobile mapping systems (MMS) in urban areas, we review the topics of pose estimation for georeferencing in urban environments, 3D data acquisition (point clouds and from imaging), and the associated trade-offs between cost and update frequency.

Although commercial LiDAR-based MMS utilize high-end scanners and precise navigation units to deliver centimeter-level 3D point clouds *(Elhashash et al., 2022)*. However, their high acquisition and operational costs generally limit their deployment to one-off surveys or surveys conducted at longer temporal intervals, creating a data gap for applications like asset management and digital twin maintenance, which require high-frequency observations to detect changes over time. Consequently, there is a strong motivation for low-cost mapping concepts that trade ultimate geometric accuracy for higher temporal resolution and economic scalability.

Image-based systems have emerged as a cost-effective alternative to LiDAR-centric platforms *(Madeira et al., 2008; Frentzos et al., 2020)*. Beyond lower hardware costs, cameras provide rich RGB information crucial for semantic interpretation tasks, such as traffic-sign inventory and pavement classification, which can be automated via deep-learning pipelines. However, low-cost image-based MMS face specific challenges in urban environments. Purely vision-based approaches are sensitive to illumination and texture-less surfaces, while low-cost Real-Time Kinematic (RTK)-GNSS solutions suffer from signal blockage and multipath effects in "urban canyons".

To ensure accurate georeferencing under these conditions, multi-sensor fusion has become the standard solution. Integrating GNSS, IMU, and visual odometry allows for robust pose estimation even when individual sensors are degraded *(Elhashash et al., 2022; Fan et al., 2025)*. Even with these technical advances, most existing multi-sensor systems need to rely on costly components and are designed for specialized survey vehicles. There remains a lack of truly scalable solutions designed for integration into existing municipal fleets, such as waste collection trucks, which offer a promising strategy to achieve exhaustive spatial coverage at marginal operational cost *(Anjomshoaa et al., 2018)*.

3 System Requirements

The intended purpose of FloMuSS outlined in the introduction necessitates a system capable of operating across the entire urban road network. To address the gaps identified in the related work, the proposed system must adhere to specific requirements regarding scalability, real-time processing, and usability.

- Sensing configuration (GNSS-IMU-Camera): Relying on GNSS alone is often insufficient in dense urban environments due to signal blockage and multipath effects, resulting in intermittent availability and degraded positioning quality. While higher-grade IMUs can mitigate short GNSS outages, they increase system cost and still require exteroceptive constraints to bound drift over longer GNSS degradations. LiDAR-centric mobile mapping systems provide high-quality geometry but typically involve

substantially higher acquisition and operational costs, which limits update frequency and scalability for municipal fleet deployment. A camera-based approach offers a cost-effective source of complementary exteroceptive information. It supports drift-limited motion estimation when GNSS quality is poor, provides texture and appearance cues relevant for municipal inventory tasks (e.g., signage, lane markings, facade elements), and enables dense 3D reconstruction (point clouds) as a geometric data product when deployed as a stereo setup. We therefore adopt a GNSS–IMU–Camera configuration that balances robustness, cost, and information content.

- Modular sensor scalability: The system architecture must be sensor-agnostic to support diverse mapping tasks. The core configuration is based on GNSS-IMU-Camera sensing to capture RGB imagery and derive 3D point clouds. The design must also support the modular integration of environmental sensors, such as electrochemical gas arrays (measuring $PM_{2.5}$, NO_x) or spectral acoustic monitors for urban noise profiling, without altering the fundamental positioning framework.
- Real-time georeferencing: Integrating these platforms into active municipal workflows imposes specific processing requirements regarding real-time capability. Unlike post-processing techniques, the system requires an immediate georeferencing solution. A real-time estimation framework enables time-continuous pose estimation and immediate spatial registration of the captured road-space and environmental sensor streams, facilitating live streetscape monitoring (e.g., real-time pollution heatmaps) as well as downstream dense point-cloud generation. In this context, adopting a tightly couple GNSS-IMU-VSLAM (Visual Simultaneous Localization and Mapping) architecture is key to achieving real-time georeferencing under dynamic operating conditions. By leveraging temporal coherence and tight inertial coupling, VSLAM effectively rejects dynamic outliers, such as moving traffic and pedestrians, that typically degrade the global reconstruction techniques applied in standard Structure-from-Motion (SfM) solutions.
- Operational integration: Finally, to ensure economic viability, the system must be designed for "plug-and-play" deployment on non-specialized municipal vehicles. This dictates a low-maintenance form factor that minimizes post-processing efforts and does not interfere with the primary duties of the service vehicle (e.g., waste collection trucks).

4 System Configuration and Hardware Integration

To address the requirements, we developed FloMuSS as a modular, low-cost (~5,000 €) multi-sensor system. The setup prioritizes hardware-level synchronization, scalability, and robustness for deployment on municipal vehicles.

The core unit comprises two Stereolabs ZED X One global-shutter cameras forming a forward-looking stereo pair. Equipped with an Onsemi AR0234 sensor, each camera delivers a resolution of 1928 $(H) \times 1200$ (V) at a frame rate of 60 frames per second (fps). The optics feature a 2.2 mm focal length lens, providing a broad field of view of $110°(H) \times 79.6°(V)$, ideal for environmental perception. Each camera includes a factory-calibrated 6-axis IMU, eliminating the need for manual camera-IMU calibration. The IMU consists of an accelerometer with a measurement range of ± 12 g (resolution: 0.36 mg) and a gyroscope capable of

measuring ±1000 (degrees per second) dps (resolution: 0.03dps). Data is transmitted via automotive-grade GMSL2 cables to a quad capture card and processed by an NVIDIA Jetson AGX Orin. For positioning, a Drotek DP0601 RTK-GNSS receiver (based on the u-blox ZED-F9P module) is integrated. This multi-band receiver provides real-time PVT (Position, Velocity, Time) data at a maximum update rate of 8 Hz, along with a Pulse-Per-Second (PPS) signal. The PPS disciplines the Jetson's system clock (PTP master), which triggers the cameras.

The sensors are mounted on a rigid aluminum rail with adjustable sliders for baseline configuration (**Fig. 2**). The GNSS antenna is centered between the cameras to minimize lever-arm effects, and cameras are protected by 3D-printed housings. The integration of further sensors, e.g., for environmental data like air quality is possible. The system is installed on the roof of a municipal garbage vehicle using a detachable base plate (**Fig. 3**), ensuring an unobstructed field of view. Power is drawn from the truck's 12V socket via a DC/DC converter, facilitating rapid installation without modifying vehicle wiring. A right-handed vehicle-fixed coordinate system is defined at the left camera, which serves as the base for the following calibration, trajectory estimation and dense point cloud generation.

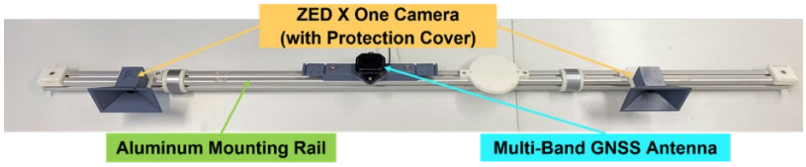

Fig. 2: System integration on the aluminum mounting rail.

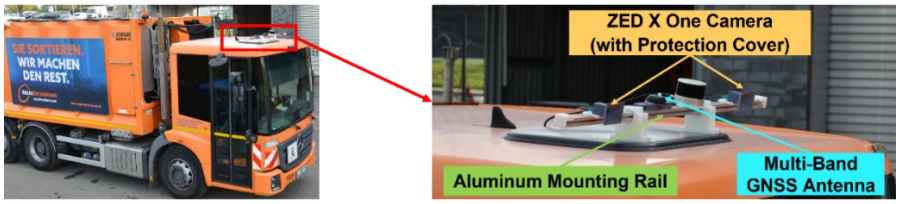

Fig. 3: System deployment on the municipal garbage vehicle.

5 Methodology

For FloMuSS, the determination of the current pose of the sensors mounted on moving carrier vehicles over time in a globally defined coordinate reference system (e.g., in the UTM coordinate reference system) is a fundamental prerequisite. This requirement necessitates the robust estimation of the 6-DoF camera trajectory, which serves as the spatio-temporal reference for aligning the sensor data. To ensure continuous and accurate georeferencing of this multi-modal data under complex real-world conditions, we developed the GNSS-IMU-VSLAM fusion pipeline for real-time camera pose estimation.

5.1 System States for Sensor Fusion Framework

The objective is to estimate a time-continuous 6-DoF pose for the vehicle body frame $\mathcal{F}^{b}$, defined here to coincide with the left camera ($\mathcal{F}^{c_L}$). We utilize a local East-North-Up (ENU)

navigation frame $\mathcal{F}^n$, with the origin set at the initial GNSS/RTK fix. The minimal state vector at time t_k is defined as $x_k^{\text{pose}} = [p_k^n, \theta_k]^T$, where $p_k^n \in R^3$ is the position and θ_k is the unit quaternion representation of the rotation matrix $R_{b,k}^n \in SO(3)$. Given the estimated body pose, the poses of the right camera ($\mathcal{F}^{c_R}$) and the GNSS antenna phase center ($\mathcal{F}^g$) are derived via fixed, known extrinsic calibrations.

5.2 Real-Time Pose Estimation via GNSS-IMU-VSLAM Fusion

Our implementation builds upon the GNSS-stereo-inertial solution *(Cremona et al., 2023)*, an extension of ORB-SLAM3 *(Campos et al., 2021)*, adapting it specifically for the urban operational domain. A critical distinction of our system is the optionality of the inertial stream. Our system is designed to operate as a stereo-only GNSS-VSLAM when inertial data are unavailable, with potentially reduced robustness and accuracy compared to the full GNSS-IMU-VSLAM configuration. It automatically upgrades to a full GNSS-IMU-VSLAM configuration when valid IMU readings are detected. GNSS measurements are introduced as unary factors in this optimization, ensuring that global position observations continuously correct the local map drift and the pose estimate. To incorporate global positioning, GNSS measurements, first transformed into the ENU frame $\mathcal{F}^n$, must be associated with the visual keyframes. We employ a temporal proximity association strategy similar to *(Cremona et al., 2023)*. Let t_i denote the timestamp of the i-th keyframe and $\widehat{p_g}(t)$ represent the continuous-time GNSS antenna position, for each assigned keyframe, we query the closest GNSS measurement in time. The measurement is associated with keyframe-i if the absolute time difference is within a rigorous tolerance threshold Δt_{gnss}: $|t_i - t_{\text{gnss}}| < \Delta t_{\text{gnss}}$, and the keyframe-$i$ is added to the subset $\mathcal{I}_{\text{gnss}}$, otherwise, the keyframe remains unconstrained by global positioning. This selective association prevents stale or asynchronous GNSS data from corrupting the tightly coupled optimization.

Following the coordinate definitions, the full state vector x_i for the keyframe-i in the backend optimization includes the navigation states and the inertial biases:

$$x_i = \left[q_{b,i}^n,\ p_i^n,\ v_i^n,\ a_{\text{bias},i}^b,\ \omega_{\text{bias},i}^b\right]^T \tag{1}$$

where p_i^n and v_i^n are the position and linear velocity of the vehicle body frame expressed in $\mathcal{F}^n$, and $q_{b,i}^n$ is the unit quaternion representing the rotation $R_{b,i}^n$. The terms $a_{\text{bias},i}^b$ and $\omega_{\text{bias},i}^b$ denote the slowly time-varying biases for the accelerometer and gyroscope, respectively.

We construct a factor graph optimization problem over a local window of keyframes $\mathcal{I}$ and the set of visible 3D landmarks $\mathcal{J}$. The optimization targets the keyframe states $\{x_i\}_{i\in\mathcal{I}}$ and landmark positions $\{m_j\}_{j\in\mathcal{J}}$, where $m_j \in R^3$. The total cost function is a sum of visual, inertial, and global positioning residuals. The local bundle adjustment is formulated as:

$$\{x_i^*, m_j^*\} = \arg\min_{\{x_i\},\{m_j\}} \Big(\sum_{(i,j)\in\mathcal{O}} \rho_{\text{vis}} \left(r_{ij}^{\text{vis}\ T}\ \Sigma_{ij}^{\text{vis}\ -1}\ r_{ij}^{\text{vis}} \right) + \sum_{(i,i-1)\in\mathcal{I}_{\text{imu}}} \rho_{\text{imu}} \left(r_{i,i-1}^{\text{imu}\ T}\ \Sigma_{i,i-1}^{\text{imu}\ -1}\ r_{i,i-1}^{\text{imu}} \right) + \sum_{i\in\mathcal{I}_{\text{gnss}}} \rho_{\text{gnss}} \left(r_i^{\text{gnss}\ T}\ \Sigma_i^{\text{gnss}\ -1}\ r_i^{\text{gnss}} \right) \Big) \tag{2}$$

where $\rho(\cdot)$ denotes robust loss functions employed to downweight outliers. The visual residual r_{ij}^{vis} encodes the reprojection error of a landmark-j observed in keyframe-i. It is defined as the difference between the observed stereo coordinate u_{ij} and the projection of the landmark:

$$r_{ij}^{\text{vis}} = u_{ij} - \pi\left(T_{c_L,i}^{n\ -1}\ m_j\right) \tag{3}$$

where $\pi(\cdot)$ is the stereo pinhole projection function and $T_{c_L,i}^{n}$ is the left camera pose derived from the body frame states. The information matrix Σ_{ij}^{vis} is scaled by the feature extraction scale level, adhering to the standard ORB-SLAM3 formulation.

When valid IMU data is available between consecutive keyframes $i-1$ and i, we employ preintegration theory to synthesize a relative motion constraint. The residual $r_{i,i-1}^{\text{imu}}$ penalizes deviations between the pre-integrated relative measurements (rotation, velocity, and position increments) and the estimates predicted by the states x_{i-1} and x_i. The covariance $\Sigma_{i,i-1}^{\text{imu}}$ is derived by propagating the continuous-time accelerometer and gyroscope noise densities through the integration period. If IMU data is missing, the set $\mathcal{I}_{\text{imu}}$ is empty, effectively reducing to a visual-GNSS bundle adjustment.

The GNSS residual enforces consistency between the estimated body pose and the raw position measurement provided by the receiver. This factor accounts for the lever arm offset r_g^b:

$$r_i^{\text{gnss}} = \widehat{p_{g,\iota}^{n}} - \left(p_i^n + R_{b,i}^n r_g^b\right) \tag{4}$$

Specially, the covariance Σ_i^{gnss} is dynamic and it is populated using the reported horizontal and vertical accuracy metrics from the RTK-GNSS receiver at each epoch. This allows the optimization to naturally trust the visual-inertial odometry more when GNSS signal quality degrades and rely on GNSS when satellite visibility is high.

6 Results and Discussion

The primary goal of the proposed low-cost multi-sensor system and processing pipeline is to enable robust, city-wide georeferencing of data collected from regularly operating municipal fleet vehicles, including under challenging urban conditions such as GNSS degradation and partial sensor outages. Building on the resulting globally consistent pose estimates, we additionally derive georeferenced dense stereoscopic point clouds for urban inventory applications. Consequently, we evaluate the system not only just on trajectory metrics, but also

by assessing the consistency of the resulting point clouds against target structures (e.g., road edges, facades) overlaid on high-resolution orthophotos from Geoportal NRW[1].

6.1 MMS Reference System

To conduct the evaluation, the FloMuSS system was rigidly mounted onto an eagle eye survey vehicle, which is already equipped with a commercial, high-resolution MMS (**Fig. 4**). In 3D point cloud data capturing projects for highways, eagle eye has demonstrated an absolute accuracy of 1 cm in position and height at driving speeds of up to 100 km/h. Given the superior accuracy of the MMS's high-grade solution, which claims centimeter-level global accuracy, its trajectory serves as the ground truth reference. However, because the sensors are mounted at different positions on the vehicle roof, the raw trajectories are not directly comparable. To resolve this, the MMS ground truth trajectory was spatially transformed into the coordinate system of the FloMuSS left camera. This transformation uses pre-calibrated extrinsic parameters to rigorously account for the 3D lever-arm effect and boresight alignment, ensuring that the evaluation measures algorithmic performance rather than spatial offsets.

Fig. 4: Eagle eye MMS reference system plus FloMuSS low-cost system

6.2 Performance of GNSS-IMU-VSLAM Fusion

The GNSS-IMU-VSLAM pipeline was evaluated using data recorded in May 2025 in Herzogenrath, Germany. **Fig. 5** visualizes the trajectories of the left (magenta) and right (orange) cameras alongside the GNSS antenna (green), and depicts a segment with dense roadside vegetation causing significant GNSS signal disturbance. While the GNSS receiver typically achieves centimeter-level accuracy in open areas, the canopy cover in this section causes the solution quality to deteriorate drastically, resulting in positioning errors of up to 3 meters or total signal loss. Despite this extreme volatility, the fused camera trajectories remain smooth and accurately aligned with the road axis, maintaining the correct lever-arm offset from the antenna. This demonstrates the system's ability to bridge short periods of GNSS degradation using visual-inertial constraints. Specifically, for the quantitative analysis, we evaluated the above-mentioned trajectory segment with a total length of 150 m. Within this section, the system achieved a positioning accuracy of 0.15 m Root Mean Square Error (RMSE) relative to the eagle eye MMS ground truth.

[1] https://www.geoportal.nrw/

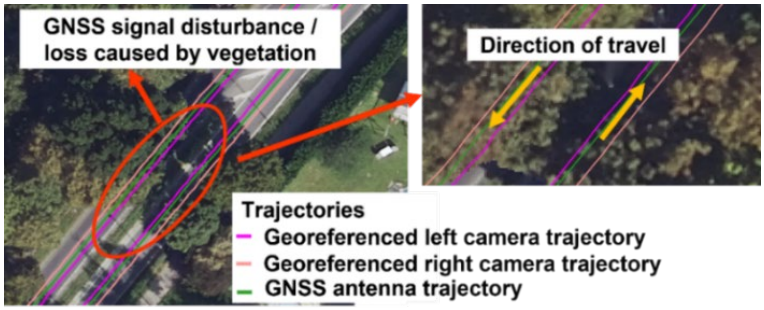

Fig. 5: GNSS-IMU-VSLAM under GNSS signal degradation.

6.3 3D Dense Point Cloud Generation

To evaluate the quality of camera data for geometric measurements or change detection, Multi-view Stereo (MVS) technique was used to reconstruct 3D dense point cloud from recorded stereo images. The photogrammetric alignment was initialized with precise camera poses derived from the GNSS-IMU-VSLAM fusion. This external trajectory integration is essential to bridge GNSS-denied zones, where relying on standard photogrammetric alignment in these areas would result in significant absolute position shifts or incomplete reconstruction. Furthermore, utilizing these predefined camera constraints accelerates the pipeline by removing the need for the computationally intensive initial alignment step.

Fig. 6 illustrates the cloud-to-cloud differences to a reference 3D point cloud from the eagle eye MMS, using a color-coded representation. Deviations in particularly relevant areas, such as the road surface, are predominantly below 0.05 m. Points located farther from the stereo cameras (e.g., facades and roofs) exhibit larger deviations of approximately 0.1-0.3 m. A quantitative comparison along a representative cross section is shown in **Fig. 7**, highlighting the vertical agreement between the reconstructed and reference point clouds.

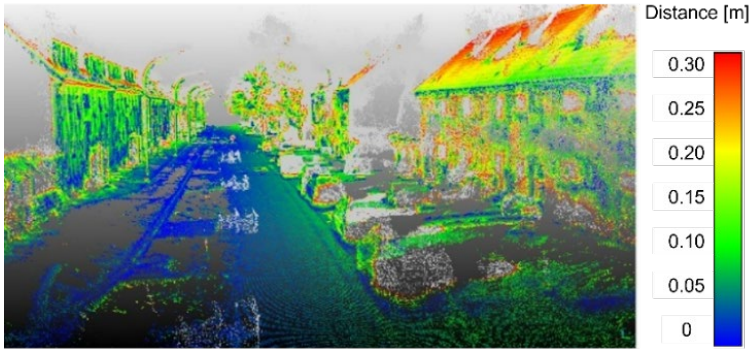

Fig. 6: Cloud-to-cloud distance between the 3D point cloud reconstructed from georeferenced stereo images and the reference point cloud from the eagle eye MMS in Herzogenrath.

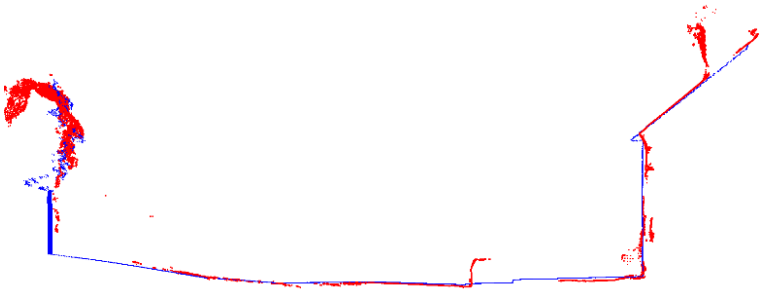

Fig. 7: Cross-sectional comparison of the reconstructed 3D point cloud (red) and the reference eagle eye MMS point cloud (blue) along a representative road segment.

6.4 Evaluation of Dense Point Cloud for the Exemplary Use Cases

The usability of the derived 3D point cloud data for the exemplary use cases depends on the system's ability to recognize and map specific objects in the street environment:

- Road Geometry: Detection of curbs, lane markings, and surface types.
- Objects & Furniture: Classification of traffic lights, bollards, and street furniture.
- Parking Semantics: Identification of parked vehicles and interpretation of regulatory signage.
- 3D Topology / digital surface and terrain models (DSM/DTM): High-precision height measurement of vertical break edges for hydraulic analysis.

We evaluated the data quality of the resulting 3D point clouds based on four key indicators:

- Completeness: The completeness of all objects of a category along the captured road section.
- Differentiability: The ability to distinguish distinct features within the data.
- Geometric Accuracy: The magnitude of deviations in location or size of an object.
- Homogeneity: The consistency of data quality for a certain object category along the captured road section.

To ensure practical relevance, these indicators were tested on critical road infrastructure features, including curbs, curb ramps, roadway boundaries, road markings, traffic signs (including text legibility), and stationary vehicles. The assessment was performed via manual visual inspection of the 3D point clouds generated from identical road segments, selected for their diverse infrastructure and varying cross-sections. For each indicator, specific objects were compared side-by-side between the FloMuSS sensor and the eagle eye MMS. The evaluation accounts for the different sensing modalities: RGB color fidelity is used for both the MMS and the vision-based FloMuSS system, while LiDAR reflection intensity is only available for the MMS. For the pluvial flood risk management scenario, the evaluation assessed whether the point clouds' geometric accuracy and spatial resolution are sufficient for generating DTM. Validation involved visual comparisons at hydraulically challenging locations to test the differentiability of key features. The analysis specifically focused on determining if flow-impeding structures (e.g., walls) and terrain elevation differences could be reliably distinguished from temporary elements, such as vehicles.

Using the eagle eye high-end MMS as the reference system for benchmarking, the FloMuSS system was found to be suitable for the targeted conceptual road planning, traffic management and parking management use cases. In terms of *completeness*, the eagle eye MMS performed considerably better than the low-cost system as expected. The latter exhibited gaps in the 3D point cloud that resulted in some objects being missed entirely. Due to variations in the presence and density of point clouds, the *homogeneity* of object detection in the low-cost system has room for improvement. Regarding *differentiability*, the data quality of the low-cost system was much closer to that of the reference system. Moderate weaknesses were observed in the detection of traffic signs and in identifying the boundary between the road surface and the shoulder. In these cases, the features were sporadically not distinguishable from the background. The *geometric accuracy* of the FloMuSS system, which is predominantly below 0.05 m in the relevant areas (see Section 6.3), proved to be sufficient for conceptual road planning, traffic and parking management use cases, for which a very high accuracy is generally not required. Minor disadvantages compared to the reference system arise from the lower point density, which complicates the precise determination of object dimensions, such as those of curbs. Overall, the processing and interpretation of raw point cloud data remain challenging for road planning and traffic management practitioners. This highlights the need for more standardized and user-friendly data representations, as well as for automated object detection and classification methods, which will be addressed in future work.

For the pluvial flood risk management use case, the evaluation focused on assessing whether the spatial resolution and geometric accuracy of the 3D point clouds are sufficient for generating DTMs. Sensor data from all three systems were visually compared at hydraulically challenging locations. The analysis examined the ability to distinguish terrain elevation differences, flow-impeding structures such as walls and curbs, and temporary objects like parked vehicles. The results indicate that the generated 3D point clouds provide an adequate basis for DTM generation in urban street environments, supporting flood hotspot identification and hydraulic analysis. For further flood-related applications that require higher resolution, data from a high-end MMS is still needed.

7 Conclusion and Outlook

This study demonstrates that the proposed FloMuSS system and sensor-fusion based processing pipeline enable reliable georeferencing of vehicle-borne sensor data even under challenging urban conditions, including GNSS signal loss and IMU temporal data gaps. By integrating GNSS, IMU, and visual information within a real-time GNSS-IMU-VSLAM processing pipeline, robust camera trajectories and globally referenced 3D point clouds can be obtained from data collected by regularly operating municipal fleet vehicles.

This paper presented a pipeline that utilizes a real-time GNSS-IMU-VSLAM framework. The selection over pure SfM was motivated by two factors:

- Dynamic robustness: VSLAM leverages temporal coherence and tight inertial coupling to effectively reject dynamic outliers (e.g., moving traffic, pedestrians) that typically degrade global reconstruction technique applied in SfM.

- Capability for live monitoring: The real-time estimation enables immediate georeferencing of the environmental sensor streams, facilitating live streetscape monitoring (e.g., real-time pollution heatmaps) rather than post-process analysis.

The results for the exemplary use cases show that FloMuSS system can provide a viable and up-to-date data basis for urban street-space management. While the resulting point cloud quality does not fully match that of high-end MMS, the achieved accuracy and spatial resolution are sufficient for conceptual road planning and traffic management tasks, where very high geometric accuracy of the order of 1-2 cm is often not required. Furthermore, the approach also shows strong potential for pluvial flood risk management. The continuous acquisition of dense point clouds provides a suitable basis for generating DTMs and identifying flow-impeding structures in urban environments. However, when higher resolutions and accuracies are required for pluvial flood risk management tasks, the use of high-end MMS remains essential and sensible.

However, while the system effectively handles GNSS outages and partially missing IMU data, the reliance on sequential estimation reveals a susceptibility to irregular image acquisition rates and temporal data gaps. These operational irregularities can destabilize the real-time fusion pipeline. In addition, when inertial measurements are unavailable for extended periods, the estimation can remain feasible in a stereo-only configuration, but typically with reduced accuracy and robustness (e.g., increased drift and less stable heading). For such segments, post-processed, globally consistent reconstruction and trajectory refinement can be advantageous. To address these, future development will focus on integrating supplementary global optimization strategies. These post-processing mechanisms can serve as a robustness fallback, designed to bridge temporal gaps and recover trajectories in data segments where standard sequential tracking assumptions are violated.

Nevertheless, the availability of dense 3D point clouds alone is insufficient to generate direct operational value. At present, the system primarily produces raw 3D point cloud data, which places a significant processing burden on municipal practitioners and limits seamless integration into existing planning workflows. To fully exploit the potential of continuous, fleet-based sensing, the data must be transformed into standardized and semantically enriched object layers, such as curbs, lane boundaries, parking spaces, or regulatory signage, allowing users to work with meaningful, planning-relevant information rather than unstructured geometry. Achieving this transformation requires robust and automated workflows for object detection, segmentation, and classification. Automation is essential to ensure consistent data quality, reproducibility, and cost efficiency, and to enable frequent updates of street-space inventories with minimal manual effort. Future work should therefore focus on scalable analysis pipelines and interoperable data standards that translate continuously collected multi-sensor data into actionable information for municipal planning and pluvial flood risk mitigation, thereby improving the suitability of the resulted geometric–semantic urban data for applications such as urban digital twin generation.

Acknowledgements

The authors would like to thank all partners for their excellent cooperation. We gratefully acknowledge the project funding from the German Federal Ministry of Digital and Transport under the FloMuSS funding measure (grant number 01FV1013-A and B).

References

ANJOMSHOAA, A., DUARTE, F., RENNINGS, D., MATARAZZO, T. J., DESOUZA, P. & RATTI, C. (2018): City scanner: Building and scheduling a mobile sensing platform for smart city services. IEEE Internet of Things Journal, 5 (6), 4567–4579.

CAMPOS, C., ELVIRA, R., GÓMEZ RODRÍGUEZ, J. J., MONTIEL, J. M. M. & TARDÓS, J. D. (2021): ORB-SLAM3: An accurate open-source library for visual, visual–inertial, and multimap SLAM. IEEE Transactions on Robotics, 37 (6), 1874–1890.

CREMONA, J., CIVERA, J., KOFMAN, E. & PIRE, T. (2024): GNSS-stereo-inertial SLAM for arable farming. Journal of Field Robotics, 41 (7), 2215–2225.

ELHASHASH, M., ALBANWAN, H. & QIN, R. (2022): A review of mobile mapping systems: From sensors to applications. Sensors, 22 (11), 4262.

FAN, Z., ZHANG, L., WANG, X., SHEN, Y. & DENG, F. (2025): LiDAR, IMU, and camera fusion for simultaneous localization and mapping: A systematic review. Artificial Intelligence Review, 58 (6), 1–59.

FRENTZOS, E., TOURNAS, E. & SKARLATOS, D. (2020): Developing an image-based low-cost mobile mapping system for GIS data acquisition. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 43, 235–242.

MADEIRA, S., GONÇALVES, J. & BASTOS, L. (2008): Low cost mobile mapping system for urban surveys. In 13th FIG Symposium on Deformation Measurement and Analysis & 4th IAG Symposium on Geodesy for Geotechnical and Structural Engineering, LNEC, Lisbon, 12–15.