

# Bridging Observed and Modeled Cities: Multi-Band Consensus Footprints for MLS-to-CityGML Registration in Urban Environments

Marco Antonio Ortiz Rincón<sup>1</sup>, Yihui Yang<sup>1</sup>, Christoph Holst<sup>1</sup>

<sup>1</sup> Chair of Engineering Geodesy, TUM School of Engineering and Design, Technical University of Munich, Munich, Germany - {marco.ortiz, yihui.yang, christoph.holst} @tum.de

**Keywords:** Point Cloud Registration, CityGML, Multi-band Consensus, Coarse-to-Fine, Raster-based Alignment.

## Abstract

This paper presents a multimodal 3D data fusion workflow tailored for registering mobile laser scanning (MLS) point clouds to CityGML semantic building models (MLS-to-CityGML) in urban street environments. The proposed method first extracts stable footprint features from MLS fragments using a multi-band consensus filtering strategy, which suppresses noise and non-building elements in MLS observations. In parallel, 2D target features are derived from CityGML Level of Detail 2 (LoD2) building geometry by extracting and cleaning ground-contact footprint segments. The resulting representations are aligned in 2D using raster-based distance-transform matching, and the estimated pose is subsequently transferred to 3D through vertical alignment and refined using the plane-voxel generalized iterative closest point (PV-GICP) algorithm. Finally, MLS drift analysis is conducted based on adaptive fragmentation. Evaluation across five urban scenarios shows that the proposed multi-band consensus filter provides reliable coarse initialization. The mean horizontal residual of all evaluated scenes achieves 0.041 m after coarse registration, while subsequent PV-GICP refinement further reduces the overall mean residual from 0.065 m to 0.015 m. The experimental results demonstrate the workflow as a robust coarse-to-fine registration strategy for structured urban environments, thereby providing a reliable georeferencing basis for updating city-scale semantic models.

## 1. Introduction

Urban environments evolve continuously through redevelopment, refurbishment, and temporary interventions, creating a persistent demand for keeping 3D city models up to date (Arroyo Oñor et al., 2018, Eriksson and Harrie, 2021). In practice, however, update cycles often lag behind reality because they still rely heavily on manual inspection and editing, which are costly and difficult to scale across large areas (Wang et al., 2020). Meanwhile, many municipalities maintain semantic building or city models in standards such as CityGML (City Geography Markup Language), where LoD2 models are widely available and therefore serve as a natural reference for city-scale applications (Kolbe, 2009). Linking new observations to such existing reference models is a key prerequisite for automated or semi-automated model updates.

Mobile Laser Scanning (MLS) data is an attractive observation source in this context. MLS systems capture dense and accurate street-level geometry in the form of 3D point clouds and provide rich façade information that may be missing in airborne-based measurements (Puente et al., 2011). If MLS observations can be reliably aligned to a CityGML reference, they can support change detection, local quality assessment, and targeted updates of building geometry along streets and corridors (Stilla and Xu, 2023). This potential has motivated increasing research on LiDAR-to-model and multi-source registration, yet robust integration remains challenging in complex urban scenes (Wysocki et al., 2021, Xu et al., 2026).

A major difficulty lies in the mismatch between the “as-observed” MLS point clouds and the “as-modeled” LoD2 geometry. Urban MLS data contains vegetation, street furniture, moving objects, and occlusions, while point density and noise levels vary along the trajectory (Cheng et al., 2018). In contrast, CityGML LoD2 models represent generalized building structures and lack fine

façade details. Consequently, overlap between the two datasets can be limited and spatially inconsistent, leading to inaccurate or erroneous correspondence-based registration (Ma et al., 2025, Xu et al., 2026). In practice, coarse registration therefore becomes a critical bottleneck, as inadequate initialization can cause fine alignment methods to converge to incorrect local minima or fail entirely.

Even when MLS data is directly georeferenced by GNSS/IMU, urban signal degradation and trajectory drift can leave residual offsets to the CityGML reference. Fragment-wise registration is commonly used to mitigate drift; however, it introduces an inherent trade-off: smaller fragments reduce drift accumulation but may lack enough geometric constraints, whereas larger fragments preserve more structure yet remain affected by residual drift (Ortiz Rincón et al., 2026).

These challenges motivate an MLS-to-CityGML fusion strategy that improves coarse initialization through robust feature extraction and subsequently refines estimated transformations in a fragment-wise way. The main contributions of this work are:

- **Multi-band consensus filtering for MLS footprint features:** a learning-free filtering strategy that retains points from a master height band only when supported by observations from multiple additional bands within a local horizontal neighborhood, improving robustness to vegetation, sparse outliers, and local density variations.
- **A coarse-to-fine MLS-to-CityGML registration workflow:** the filtered MLS footprint features are aligned with wall-derived LoD2 footprint segments in 2D and refined in 3D using PV-GICP on individual MLS fragments.
- **Fragment-wise drift analysis:** a diagnostic tool for identifying residual inconsistencies across the refined transformations.

The remainder of this paper is organized as follows: Sec. 2 reviews related work on MLS registration and model-based alignment. Sec. 3 presents the proposed pipeline. Sec. 4 describes the experimental setup, evaluation metrics, and results. Sec. 5 discusses the main findings and limitations, and Sec. 6 concludes the paper.

## 2. Related Work

This section reviews three lines of work most relevant to our pipeline: registration methods for urban MLS point clouds (Sec. 2.1), footprint or plan-view representations with raster-based alignment (Sec. 2.2), and drift-aware fragment-wise strategies for MLS data (Sec. 2.3).

### 2.1 MLS point cloud registration

Urban MLS registration is commonly formulated as the estimation of a rigid transformation between a source and a target dataset, for scan-to-scan alignment, scan-to-map localization, or integration into a reference representation (Pomerleau et al., 2015). Most approaches rely on ICP-style refinement, which alternates between correspondence estimation and transformation updates (Besl and McKay, 1992, Chen and Medioni, 1991).

Generalized-ICP models local neighborhoods as anisotropic distributions and provides a plane-to-plane alignment that is well suited to man-made surfaces (Segal et al., 2009). Piecewise-ICP aims to enhance the robustness and computational efficiency of registration in complex and dynamic environments (Yang and Holst, 2025). However, these fine registration methods remain highly dependent on a suitable initialization, especially under cluttered conditions and partial overlap. Coarse alignment is therefore often established using descriptor-based correspondences, such as FPFH (Rusu et al., 2009), or dedicated global registration methods, such as FGR (Zhou et al., 2016).

In MLS-to-model registration tasks, however, reliable initialization remains challenging due to a large amount of repetitive local geometric structures in LoD2 models. This limitation motivates the use of alternative global features derived from LoD2 representations that are less sensitive to correspondence ambiguity.

### 2.2 Footprint features and raster-based alignment

A common strategy to reduce ambiguity in urban MLS-to-model registration is to focus on plan-view footprints that are shared across different data sources. In street-level scanning, building-related points are often selected near façades or walls, projected onto a 2D plane, and converted into outlines or line segments that serve as compact registration primitives (Hammoudi et al., 2009, Yang et al., 2013). Recent studies have shown that the reliability of such footprints depends strongly on effective noise suppression and on choosing an appropriate height range for extracting building boundaries (Rottmann et al., 2022).

Once both datasets are represented in 2D, raster-based alignment offers a correspondence-free alternative to direct 3D matching. A typical approach rasterizes edges or occupancy and estimates a 2D rigid transformation by evaluating a distance transform of the template, enabling efficient chamfer-style matching (Borgefors, 1986, Borgefors, 1988). Similar ideas have been used for scan-to-map alignment and localization, for example by matching edge maps to OpenStreetMap or occupancy grids

(Floros et al., 2013). Related concepts have also been applied to refine footprint geometry against rasterized contour footprint features (Zhuo et al., 2018).

While these methods scale well and avoid explicit correspondences, they remain sensitive to discretization effects and, in particular, to the quality of the extracted 2D representations. This motivates filtering strategies that stabilize footprint-like MLS representations before alignment.

### 2.3 MLS drift and fragment-wise registration strategies

Large-scale MLS acquisitions are affected by time-dependent georeferencing errors due to GNSS degradation, IMU drift, residual calibration errors, etc. These effects can accumulate along the trajectory, so a single rigid transformation may be insufficient for long corridors. Trajectory-based strategies address this issue by estimating piecewise corrections, typically through trajectory segmentation, local alignment, and subsequent global adjustment that distributes residuals across segments (Gressin et al., 2012, Takai et al., 2013).

In practice, fragment-wise registration is also widely used to mitigate MLS drift and enable local refinement. Consequently, quality-controlled and adaptive fragmentation strategies have been proposed to identify unreliable sections and reduce fragmentation failures (Kalenjuk and Lienhart, 2022, Ortiz Rincón et al., 2026). Structural priors derived from 3D city models can further support trajectory improvement by providing stable geometric constraints (Lucks et al., 2021). Overall, the existing studies motivate workflows that strengthen building-related evidence at the fragment level and explicitly analyze fragment-wise transformations to diagnose residual drift effects.

## 3. Methodology

As discussed in Sec. 2, MLS-to-CityGML registration in urban scenes is challenged by noise, varying point density, partial overlap, and drift between fragments. We therefore propose an automatic pipeline that can robustly align MLS point clouds with CityGML models in a coarse-to-fine strategy, as illustrated in Figure 1. Specifically, the MLS point clouds are preprocessed using multi-band consensus filtering (Sec. 3.1), and the CityGML model footprints are also extracted (Sec. 3.2). Then, a raster-based coarse alignment is applied to the extracted 2D footprints from two data sources, followed by a 3D refinement using PV-GICP (Sec. 3.3). In addition to registration accuracy, residual drift effects are also evaluated through a fragment-wise consistency analysis based on local start-to-end deviations. The details of each step are presented in the following subsections.

### 3.1 MLS data preprocessing

Each MLS point cloud is transformed into a compact 2D footprint feature set that emphasizes vertically persistent structures, typically building façades. This is achieved by estimating a ground reference, slicing the point cloud into height bands above the ground, and retaining only those plan-view locations that are consistently supported across multiple bands.

**3.1.1 Ground plane estimation** A local ground plane is estimated to define a stable height reference. Ground seed candidates are generated by subdividing the horizontal bounding box into an  $n_x \times n_y$  grid and selecting the minimum-elevation point in each occupied cell. This strategy yields a spatially distributed

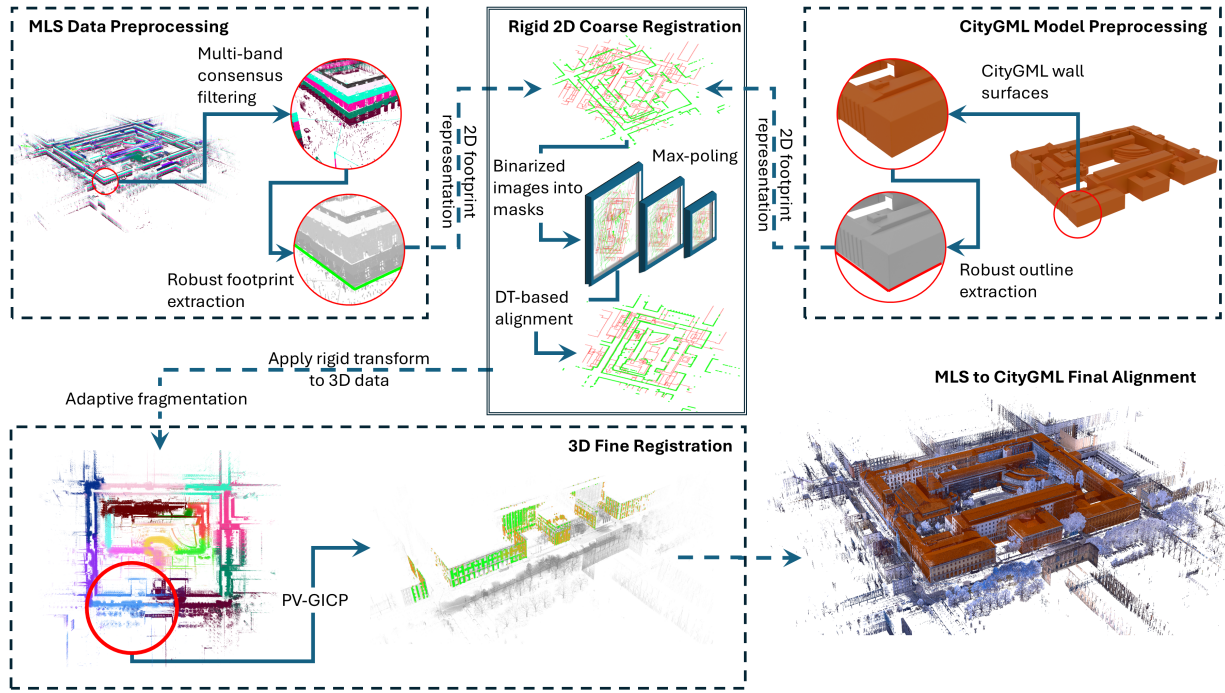


Figure 1. Overview of the proposed coarse-to-fine MLS-to-CityGML registration pipeline.

seed set and avoids over-representing locally dense areas. Candidates are retained only if they exhibit sufficient local neighborhood support, which reduces the influence of isolated outliers.

A plane  $\pi : ax + by + cz + d = 0$  is then fitted to the selected seed point set using RANSAC (Fischler and Bolles, 1987). The resulting plane parameters  $(a, b, c, d)$  define the reference ground and will be used for the subsequent band assignment.

### 3.1.2 MLS footprint extraction using multi-band consensus

Based on the fitted ground plane  $\pi$ , each point  $\mathbf{x}_i = (x_i, y_i, z_i)^\top$  is assigned a signed height above or below ground using the normalized point-to-plane distance

$$h(\mathbf{x}_i, \pi) = \frac{ax_i + by_i + cz_i + d}{\sqrt{a^2 + b^2 + c^2}}, \quad (1)$$

where the plane orientation is chosen such that  $h(\mathbf{x}, \pi) > 0$  denotes points above ground.

Based on this height measure, bands  $B$  parallel to the fitted ground plane are constructed with an initial height  $h_0$ , step size  $\Delta h$ , and band thickness  $w$ :

$$B_j = \{\mathbf{x}_i \mid h_0 + j\Delta h \leq h(\mathbf{x}_i) \leq h_0 + j\Delta h + w\}, \quad (2)$$

where  $j$  denotes the band index. Band generation stops once the number of points in a band falls below a minimum threshold  $N_{\min}$ , thereby avoiding unreliable bands at higher elevations.

To extract robust MLS footprint features without relying on semantic classification, one band  $B_k$  is first selected as the master band. Only those points whose planimetric positions are supported by points in multiple additional bands are retained in this master band (i.e., multi-band consensus). Specifically, let  $\mathbf{p}_{xy}$  denote the  $(x, y)$  coordinates of a master-band point  $\mathbf{p} \in B_k$ . For each other band  $B_j$ , support is counted if there exists at least one point  $\mathbf{q} \in B_j$ , with planimetric coordinates  $\mathbf{q}_{xy}$ , within a

horizontal neighborhood radius  $r_{\text{cons}}$ :

$$\|\mathbf{p}_{xy} - \mathbf{q}_{xy}\|_2 \leq r_{\text{cons}}. \quad (3)$$

The number of *distinct* supporting bands is thus defined as

$$s(\mathbf{p}) = |\{j \neq k \mid \exists \mathbf{q} \in B_j : \|\mathbf{p}_{xy} - \mathbf{q}_{xy}\|_2 \leq r_{\text{cons}}\}|. \quad (4)$$

A master-band point is retained if  $s(\mathbf{p})$  is larger than a defined minimum number of supporting bands. The retained points are finally projected to a 2D plan view to form the MLS footprint:

$$\mathcal{P}^{2D} = \{\mathbf{p}_{xy} \mid \mathbf{p} \in B_k, s(\mathbf{p}) \geq S_{\min}\}. \quad (5)$$

Here,  $k$  is the selected master-band index,  $\|\cdot\|_2$  denotes the Euclidean norm in  $\mathbb{R}^2$ ,  $|\cdot|$  means the cardinality of a set, and  $S_{\min}$  is the minimum number of distinct supporting bands required to retain a master-band point.

Figure 2 illustrates the behaviour of the multi-band consensus filtering for three representative scenarios. The left example shows a building façade, where vertically persistent points are consistently detected across several bands and are therefore retained as MLS footprints. The corresponding top view shows the points accepted by the filter in green. The other two examples, including a traffic light and a tree, show typical non-building structures. Although these objects may also locally satisfy the multi-band consensus criterion, they are largely removed in the subsequent cleanup of small isolated clusters.

### 3.2 CityGML model preprocessing

2D CityGML footprint segments are derived from CityGML wall surfaces. To this end, the CityGML data structure is parsed and the posList coordinate triplets associated with WallSurface elements are converted into ordered 3D ring vertices as:

$$\mathcal{R} = \{\mathbf{v}_i = (x_i, y_i, z_i)\}_{i=1}^m \quad (6)$$



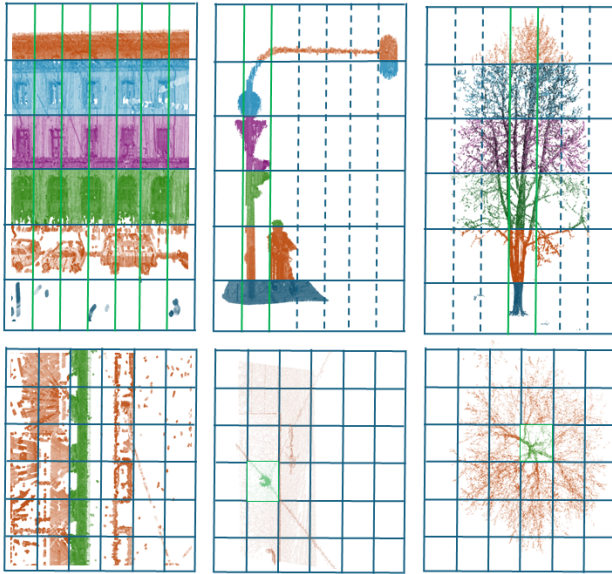


Figure 2. Examples of the height-band construction and multi-band consensus filtering for three representative scenarios. Left: building façade correctly retained as robust MLS footprint features. Middle and right: traffic light and tree as typical false positives, which are removed by the subsequent plan-view cleanup. The second row shows the corresponding top views.

For each ring, a local base height is estimated as  $z_{\min} = \min_i z_i$ , and consecutive edges are classified as ground-contact if both endpoints satisfy

$$z_i, z_{i+1} \leq z_{\min} + \varepsilon_g, \quad (7)$$

where  $\varepsilon_g$  is an empirical tolerance value. Each accepted edge is projected to the horizontal plane and stored as an undirected 2D segment  $(x_i, y_i, x_{i+1}, y_{i+1})$ .

Duplicate segments are then removed by endpoint quantization, and very short segments are also discarded. The resulting segment graph can provide corner candidates, defined as non-degree-two nodes and degree-two nodes with non-collinear incident directions, which are used for visualization and sanity checks.

### 3.3 Coarse-to-fine registration

A coarse 2D rigid alignment between MLS footprint features and CityGML footprint segments is first estimated by raster distance-transform matching (Sec. 3.3.1), and the result is then refined in 3D using PV-GICP (Sec. 3.3.2).

**3.3.1 Rigid 2D coarse registration** The MLS footprint set  $\mathcal{P}^{2D}$  and the CityGML footprint segments are rasterized into binary edge masks with identical image size and pixel resolution. Both datasets are mapped to pixel coordinates using the same pixel size, but each is centered individually at the image center. Internally, centered pixel coordinates  $(u, v)$  with a  $y$ -up convention are used:

$$u = x - c_x, \quad v = c_y - y, \quad (8)$$

where  $(x, y)$  denotes image pixel coordinates and  $(c_x, c_y)$  is the image center in pixel space. CityGML footprint segments are rendered as thickened lines, while the MLS footprint features

are drawn as small disks in order to produce connected edge responses. The resulting grayscale images are binarized into the masks  $M_{MLS}$  and  $M_{GML}$ .

Next, the exact Euclidean distance transform of the CityGML mask is computed (Felzenszwalb and Huttenlocher, 2012), yielding squared distances  $D_{GML}^2(\cdot)$  in pixel units. A 2D rigid pose  $(\theta, \mathbf{t})$ , with translation  $\mathbf{t} = (t_x, t_y)^\top$ , is estimated by transforming MLS edge pixels and sampling their capped squared distances to the nearest CityGML edge:

$$E(\theta, \mathbf{t}) = \frac{1}{|\mathcal{P}|} \sum_{\mathbf{p} \in \mathcal{P}} \min(D_{GML}^2(\mathbf{R}(\theta)\mathbf{p} + \mathbf{t}), d_{\max}^2), \quad (9)$$

where  $\mathcal{P}$  denotes a subsampled set of MLS edge pixels in centered coordinates,  $\mathbf{p} \in \mathbb{R}^2$  is a 2D pixel location,  $\mathbf{R}(\theta)$  is the planar rotation matrix,  $|\mathcal{P}|$  denotes the cardinality of  $\mathcal{P}$ , and  $d_{\max}$  is the distance cap in pixels. The cap limits the influence of out-of-bounds points and weakly overlapping regions.

To reduce runtime and improve robustness, the objective  $E$  is optimized on an image pyramid from coarse to fine. At each pyramid level, masks are downsampled by max-pooling in order to preserve thin edges. The search evaluates a discrete grid over  $\theta$  within a user-defined rotation interval and translations within a bounded window around the current best estimate. At finer levels, the translation and rotation step sizes are reduced and the number of sampled MLS edge pixels is increased. This yields a progressively refined 2D alignment.

**3.3.2 3D fine registration using PV-GICP** In this step, the coarse 2D pose is elevated to a 3D rigid initialization for the following fine registration. Since the coarse alignment is estimated solely in the plan view, it provides only horizontal transformation components. The missing vertical component is obtained separately utilizing ground-level information. To accomplish this, the target point cloud is generated from both LoD2 building models and Digital Terrain Model (DTM), while the original MLS point cloud is initially fragmented using Sphere Check (SSC) and subsequently initialized with the previously estimated coarse transformation (Ortiz Rincón et al., 2026). The vertical translation is calculated from the difference between source and target ground levels, both estimated in accordance with the methodology described in Sec. 3.1.1.

The resulting 3D initialization is then refined using PV-GICP (Ortiz Rincón et al., 2026). PV-GICP is designed for urban MLS scenes and exhibits high efficiency by selecting planar surfaces within voxel partitions before performing a GICP-style optimization. By focusing the correspondence search and the local surface statistics on planar structures, the refinement becomes more robust to clutter and varying point density. The refined transformation is taken as the final MLS-to-CityGML registration parameters for each fragment.

## 4. Experimental evaluation

This section presents the experimental evaluation of the proposed workflow. It first describes the datasets used in this study (Sec. 4.1) and the evaluation criteria adopted to assess the registration quality (Sec. 4.2). It then reports the results of the multi-band consensus filtering (Sec. 4.3), the LoD2 footprint extraction (Sec. 4.4), the coarse registration (Sec. 4.5), the fine registration (Sec. 4.6), and the drift analysis (Sec. 4.7).



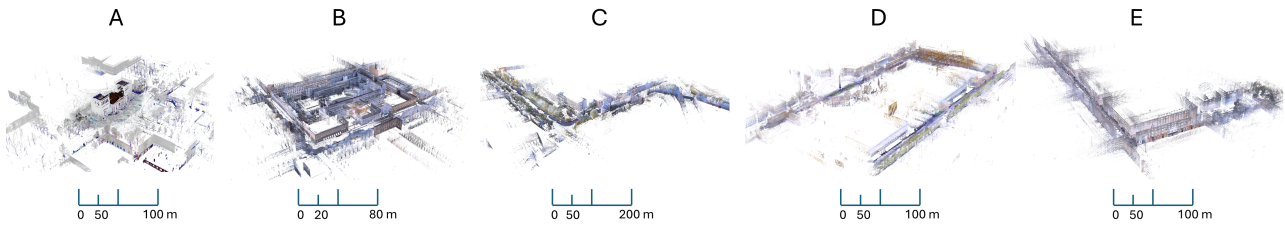


Figure 3. Overview of the acquired MLS source data from five different scenarios (A–E) for experimental evaluation.

#### 4.1 Data description

This study uses two complementary datasets: (I) five MLS point clouds serving as the source data (Figure 3), and (II) LoD2 city models functioning as the target map reference. The source data captures scene geometry at the street level, acquired with the Z+F FlexScan 22 system<sup>1</sup>. The target dataset consists of city-wide building representations in LoD2 models, provided by the Bavarian State Mapping Agency (LDBV) and made publicly available through the Bavarian open data portal<sup>2</sup>.

#### 4.2 Evaluation criteria

Since the ground-truth transformations for the datasets are unavailable, the registration accuracy is evaluated through 1) direct visual inspection and 2) calculating the spatial distances within stable regions between registered point clouds. The second approach relies on locally planar regions extracted by the voxel-based partition in PV-GICP (Ortiz Rincón et al., 2026).

For quantitative evaluation, the estimated coarse and fine transformations are consecutively applied to the MLS point clouds and compared to a reference point cloud sampled from the LoD2 models. Registration quality is assessed using M3C2 distances computed independently on the extracted planar sections. Planes with vertical normals are used to assess the horizontal alignment accuracy, while horizontally oriented planes are used to evaluate the vertical components. The resulting residuals are summarized using separate means horizontally and vertically.

#### 4.3 Results of multi-band consensus filtering

As mentioned in Section 4.1, we evaluated the proposed workflow on five representative urban scenarios. All MLS point clouds were uniformly subsampled to a spatial resolution of 0.02 m. We then applied the height-band preprocessing described in Section 3.1, using an initial offset of  $h_0 = 1.0$  m above the estimated ground plane, a band step of  $\Delta h = 0.5$  m, and a band thickness of  $w = 0.5$  m. Band generation stops once the next band contains fewer than  $N_{\min} = 500$  points. For the multi-band consensus filter, we set the planimetric neighborhood radius to  $r_{\text{cons}} = 0.05$  m and retained a master-band point only if it is supported by at least 10 additional bands.

We finally applied a lightweight 2D cleanup to remove small isolated point clusters that are unlikely to belong to building footprints. Nearby points are grouped in plan view using a distance tolerance of 0.30 m; groups with fewer than 100 points or with a horizontal extent below 2.0 m are discarded.

Figure 4 shows a qualitative comparison of three MLS footprint variants across all five scenarios. The single-band projection

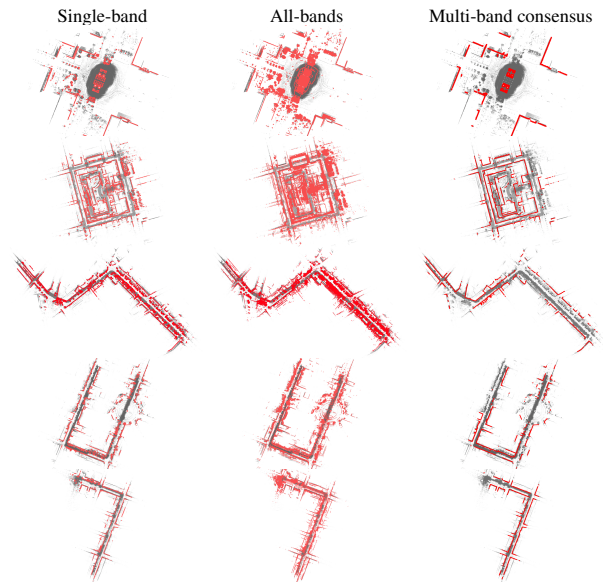


Figure 4. Qualitative comparison of three MLS footprint feature extraction strategies across the five evaluated scenarios. Gray parts represent the original MLS point clouds, and the red lines denote the extracted footprints.

can capture dominant façade structures, but often remains fragmented or locally incomplete. In contrast, the naive all-bands projection increases spatial coverage but also introduces substantial clutter and broadens the footprint response due to vertically distributed non-building structures. The proposed multi-band consensus filter provides the most balanced representation: it preserves the main façade runs and corner layouts while suppressing many isolated or weakly supported structures. The resulting cleaned 2D point set is therefore used for rasterization and subsequent coarse registration.

#### 4.4 Results of LoD2 footprint extraction

Across all five scenes, the extracted CityGML model footprints provide a clean and geometrically expressive plan-view reference. With a ground-contact tolerance of  $\varepsilon_g = 0.10$  m and a sampling step of  $s = 0.20$  m, the resulting segment set preserves long façade runs as well as finer structures such as inner courtyard boundaries and curved elements, while remaining dense enough for stable rasterization. Duplicate edges caused by shared walls and tile overlaps are effectively removed by endpoint quantization on a 2 cm grid, and short spurious fragments are suppressed using a minimum segment length of 0.50 m. Figure 6 illustrates a representative example, showing the LoD2 building model (left) and the corresponding extracted 2D CityGML footprint segments (right) used as target input for the subsequent MLS-to-CityGML registration.

<sup>1</sup> <https://www.zofre.de/en/laser-scanners/3d-laser-scanner/z-f-flexscan-22>

<sup>2</sup> <https://geodaten.bayern.de/openeodata/>

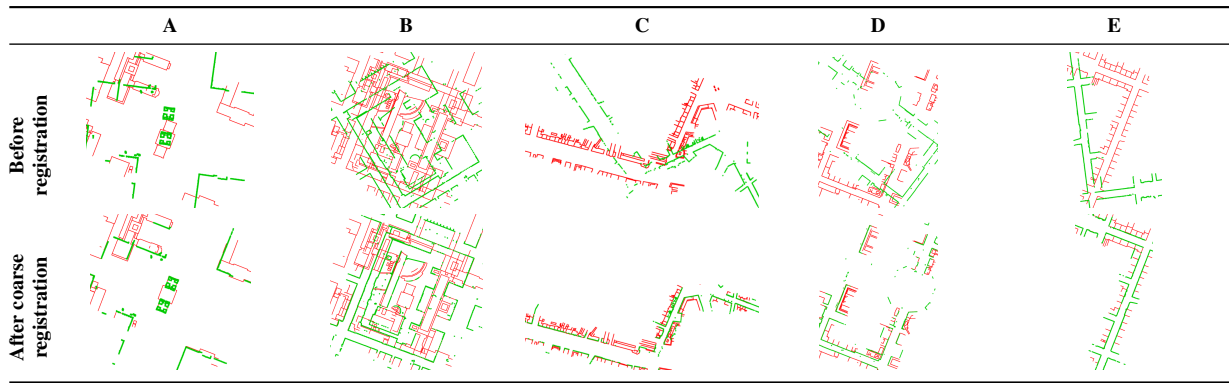


Figure 5. Coarse plan-view registration across five scenarios using multi-band-filtered MLS footprint features. Each panel overlays the CityGML footprint segments (red) and the MLS footprint features (green). Top: before registration. Bottom: after applying the estimated coarse rigid transform (2D translation and yaw rotation).

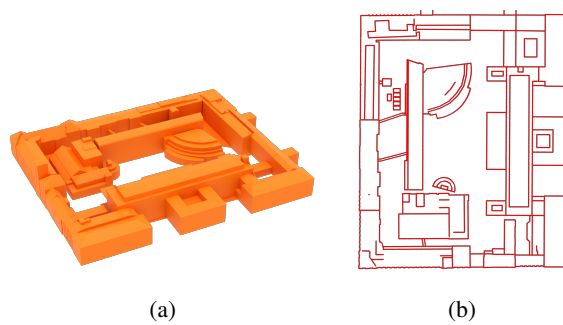


Figure 6. CityGML footprint segments generation of Scenario B: (a) LoD2 model and (b) extracted plan-view footprint segments.

#### 4.5 Results of coarse registration

The coarse alignment stage estimates an initial 2D rigid transformation between the MLS footprint features and the CityGML footprint segments using the distance-transform matching strategy described in Section 3.3. The resulting 2D pose is then transferred back to the original 3D MLS data by embedding the estimated vertical rotation and horizontal translation into a  $4 \times 4$  transformation matrix. The transformed MLS point cloud is subsequently evaluated against a reference point cloud sampled from the LoD2 model as described in Section 4.2.

Table 1 quantifies the influence of three MLS footprints variants on this initialization. The proposed multi-band consensus filter yields the lowest horizontal residual in all five scenarios, with values between 0.025 m and 0.053 m and a mean of 0.041 m. In contrast, the single-band baseline produces substantially larger residuals, ranging from 0.129 m to 0.582 m, with a mean of 0.397 m. The naive all-bands projection does not provide a usable coarse alignment in four scenarios, and in the only successful case still results in a large residual of 0.643 m. These results indicate that the proposed multi-band consensus filtering provides the most distinctive and reliable footprint features for coarse registration.

Figure 5 qualitatively confirms the performance of our method. Before alignment, the MLS footprint features and the CityGML target exhibit clear planimetric offsets, whereas after applying the estimated coarse transformation the dominant façade outlines and corner structures are brought into close agreement in all five scenarios.

| Scenario | Single-band | All-bands | Multi-band consensus |
|----------|-------------|-----------|----------------------|
| A        | 0.385       | –         | 0.044                |
| B        | 0.582       | –         | 0.032                |
| C        | 0.389       | –         | 0.052                |
| D        | 0.502       | –         | 0.025                |
| E        | 0.129       | 0.643     | 0.053                |
| Mean     | 0.397       | –         | 0.041                |

Table 1. Coarse-registration residuals [Unit: m] using three different MLS footprint extraction strategies. Mean horizontal residuals are reported after the raster-based coarse alignment.

#### 4.6 Results of fine registration

After the coarse initialization described in Section 3.3.1, the alignment is refined in 3D using the PV-GICP algorithm presented in Section 3.3.2. In contrast to the raster-based coarse stage, this refinement operates directly on the 3D point clouds and uses locally planar surface structure to improve the alignment between MLS fragments and the LoD2-based reference geometry. Figure 7 shows examples of the planar regions extracted during the PV-GICP preprocessing step. These regions correspond mainly to geometrically stable façade and roof surfaces and form the basis for the subsequent correspondence search.

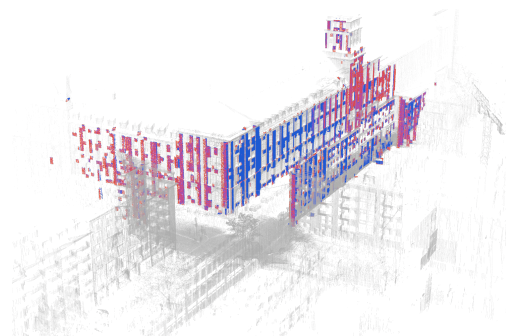


Figure 7. Planar surface regions identified by PV-GICP. Blue and red indicate target and source data, respectively.

As described in Section 4.2, the evaluation is performed separately for horizontal and vertical planar sections. Across all five scenarios, the fine registration reduces the residuals in both

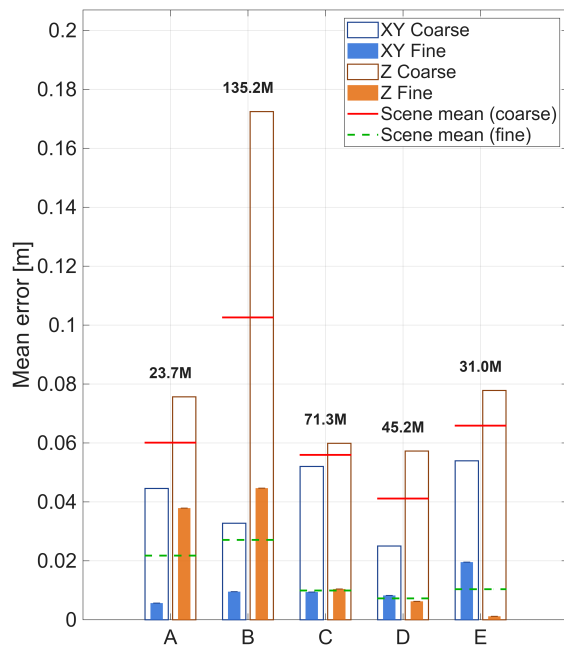


Figure 8. Mean residual errors after coarse and fine registration for the five evaluated scenarios, separated into horizontal ( $XY$ ) and vertical ( $Z$ ) components.

components. Figure 8 provides a graphical summary of the final registration accuracy. The nested bars illustrate the change in mean residuals between the coarse and fine stages for both horizontal and vertical components.

On average, the horizontal mean residual decreases from 0.041 m to 0.010 m, while the vertical mean residual decreases from 0.089 m to 0.020 m, corresponding to reductions of 74.8% and 77.3%, respectively. The greatest horizontal improvement is observed in Scene C, where the residual decreases from 0.052 m to 0.009 m. In the vertical direction, the largest improvement is obtained in Scene B, where the residual drops from 0.172 m to 0.045 m, while Scene E is reduced from 0.078 m to 0.001 m. Considering both components, the mean residual across all scenes decreases from 0.065 m after coarse registration to 0.015 m after fine registration. These results confirm that PV-GICP consistently improves the coarse initialization.

#### 4.7 Drift analysis

After the fine registration of all fragments, the consistency of the resulting transformations was evaluated across all scenarios. Table 2 summarizes this analysis in terms of local start-to-end (loop) deviations, normalized by trajectory length, together with the mean absolute rotational differences in roll, pitch, and yaw between consecutive fragment transformations.

Overall, the results indicate small horizontal loop deviations and generally moderate rotational differences. Scenario B, which has the longest trajectory (1893.1 m), shows the smallest normalized horizontal deviation with only 2 mm/100 m, indicating a highly consistent sequence of fragment transformations. In contrast, Scenario E exhibits the largest normalized horizontal deviation (25 mm/100 m), although its rotational differences remain comparatively small. The largest angular inconsistencies are observed in Scenario D, where the mean absolute differences reach  $0.444^\circ$  in roll,  $0.589^\circ$  in pitch, and  $0.287^\circ$  in yaw. Scenario A was excluded from the consistency analysis because it consists of only a single fragment.

| Scen. | Frag. | $L$<br>[m] | $\delta_{xyz}^{\text{loop}}$<br>[m] | $\delta_{xyz}^{100}$<br>[mm/100 m] | $ \Delta r $<br>[ $^\circ$ ] | $ \Delta p $<br>[ $^\circ$ ] | $ \Delta y $<br>[ $^\circ$ ] |
|-------|-------|------------|-------------------------------------|------------------------------------|------------------------------|------------------------------|------------------------------|
| A     | 1     | —          | —                                   | —                                  | —                            | —                            | —                            |
| B     | 15    | 1893.1     | 0.039                               | 2                                  | 0.266                        | 0.235                        | 0.133                        |
| C     | 3     | 926.8      | 0.065                               | 7                                  | 0.170                        | 0.133                        | 0.018                        |
| D     | 6     | 1089.3     | 0.050                               | 4                                  | 0.444                        | 0.589                        | 0.287                        |
| E     | 4     | 518.5      | 0.129                               | 25                                 | 0.045                        | 0.151                        | 0.027                        |

Table 2. Registration-consistency measures across the evaluated scenarios.  $L$  denotes the trajectory length,  $\delta_{xyz}^{\text{loop}}$  the local horizontal start-to-end deviation, and  $\delta_{xyz}^{100}$  the corresponding drift normalized by trajectory length. Rotational values are reported as mean absolute Euler-angle differences between consecutive fragment transformations. Scenario A contains only a single fragment and is therefore excluded from the drift analysis.

## 5. Discussion

**Strengths.** The results show that the proposed coarse-to-fine workflow is suitable for MLS-to-CityGML registration in structured urban scenes. In particular, the ablation study demonstrates that the quality of the coarse alignment depends strongly on the type of MLS footprint features. The single-band baseline often remains fragmented, while the naive all-bands projection introduces substantial clutter and does not yield a usable coarse alignment in most scenarios. In contrast, the proposed multi-band consensus filter provides the most distinctive and reliable footprint features, leading to the lowest coarse-registration residuals in all evaluated scenes. The PV-GICP refinement consistently reduces residuals in both components, confirming that footprint features reliably initialize 3D alignment.

**Limitations.** The quality of the registration still depends on the reliability of the extracted building-related features, since local false positives may remain in heavily cluttered scenes. In addition, the LoD2 reference geometry is mainly a simplified representation of the building objects and infrastructures and does not model facade details or street furniture. As a result, some residual discrepancies remain even after 3D refinement. For long streets with simple or repetitive geometry—such as one-sided facades or no buildings—registration may fail through ambiguous footprint matches or unstable fragment transformations.

## 6. Conclusions

This paper presented a coarse-to-fine workflow for the registration of MLS point clouds to CityGML LoD2 city models in urban scenes. The approach combines a learning-free multi-band consensus filter for extracting robust footprint features from MLS fragments, and realizes 2D coarse alignment using a raster-based distance-transform matching, further refined by PV-GICP. The results show that the proposed footprint feature extraction is critical for robust initialization: compared with a single-band baseline and a naive all-bands projection, the multi-band consensus filter provides the most reliable coarse registration across all evaluated scenarios. The subsequent fine registration further reduces the remaining horizontal and vertical residuals, achieving a final 3D registration accuracy of a few centimeters.

## References

Arroyo Otori, K., Biljecki, F., Kumar, K., Ledoux, H., Stoter, J., 2018. *Modeling Cities and Landscapes in 3D with CityGML*. Springer International Publishing, Cham, 199–215.



- Besl, P., McKay, N. D., 1992. A method for registration of 3-D shapes. *IEEE Trans. Pattern Anal. Mach. Intell.*, 14(2), 239–256.
- Borgefors, G., 1986. Distance transformations in digital images. *Computer Vision, Graphics, and Image Processing*, 34(3), 344–371.
- Borgefors, G., 1988. Hierarchical chamfer matching: a parametric edge matching algorithm. *IEEE Trans. Pattern Anal. Mach. Intell.*, 10(6), 849–865.
- Chen, Y., Medioni, G., 1991. Object modeling by registration of multiple range images. *Proceedings. 1991 IEEE International Conference on Robotics and Automation*, 2724–2729 vol.3.
- Cheng, L., Chen, S., Liu, X., Xu, H., Wu, Y., Li, M., Chen, Y., 2018. Registration of Laser Scanning Point Clouds: A Review. *Sensors*, 18(5).
- Eriksson, H., Harrie, L., 2021. Versioning of 3D City Models for Municipality Applications: Needs, Obstacles and Recommendations. *ISPRS Int. J. Geo-Inf.*, 10(2).
- Felzenszwalb, P. F., Huttenlocher, D. P., 2012. Distance Transforms of Sampled Functions. *Theory Comput.*, 8, 415–428.
- Fischler, M. A., Bolles, R. C., 1987. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. M. A. Fischler, O. Firschein (eds), *Readings in Computer Vision*, Morgan Kaufmann, San Francisco (CA), 726–740.
- Floros, G., van der Zander, B., Leibe, B., 2013. Openstreetslam: Global vehicle localization using openstreetmaps. *2013 IEEE International Conference on Robotics and Automation*, 1054–1059.
- Gressin, A., Cannelle, B., Mallet, C., Papellard, J.-P., 2012. Trajectory-based registration of 3d lidar point clouds acquired with a mobile mapping system. *ISPRS Ann. Photogramm. Remote Sens. Spatial Inf. Sci.*, I-3, 117–122.
- Hammoudi, K., Dornaika, F., Paparoditis, N., 2009. Extracting building footprints from 3d point clouds using terrestrial laser scanning at street level. *Int. Arch. Photogram Remote Sens. Spat. Inf. Sci.*, XXXVIII number 3/W4, 65–70.
- Kalenjuk, S., Lienhart, W., 2022. A Method for Efficient Quality Control and Enhancement of Mobile Laser Scanning Data. *Remote Sensing*, 14(4).
- Kolbe, T. H., 2009. *Representing and Exchanging 3D City Models with CityGML*. Springer Berlin Heidelberg, Berlin, Heidelberg, 15–31.
- Lucks, L., Klingbeil, L., Plümer, L., Dehbi, Y., 2021. Improving trajectory estimation using 3D city models and kinematic point clouds. *Transactions in GIS*, 25(1), 238–260.
- Ma, K., Yan, F., Li, S., Huang, G., Jia, X., Wang, F., Chen, L., 2025. Low-Overlap Registration of Multi-Source LiDAR Point Clouds in Urban Scenes Through Dual-Stage Feature Pruning and Progressive Hierarchical Methods. *Remote Sensing*, 17(17).
- Ortiz Rincón, M. A., Yang, Y., Holst, C., 2026. SSC-Registration. <https://github.com/TUM-Engineering-Geodesy/SSC-Registration/releases/tag/v1.0.4>. Version 1.0.4.
- Ortiz Rincón, M. A., Yang, Y., Holst, C., 2026. Quality-controlled registration of urban MLS point clouds reducing drift effects by adaptive fragmentation. *Int. J. Appl. Earth Obs. Geoinf.*, 149, 105272.
- Pomerleau, F., Colas, F., Siegwart, R., 2015. A Review of Point Cloud Registration Algorithms for Mobile Robotics. *Foundations and Trends in Robotics*, 4(1), 1–104.
- Puente, I., González-Jorge, H., Arias, P., Armesto, J., 2011. Land-based mobile laser scanning systems: a review. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, XXXVIII-5/W12, 163–168.
- Rottmann, P., Haunert, J.-H., Dehbi, Y., 2022. Automatic building footprint extraction from 3d laserscans. *ISPRS Ann. Photogramm. Remote Sens. Spatial Inf. Sci.*, X-4/W2-2022, 233–240.
- Rusu, R. B., Blodow, N., Beetz, M., 2009. Fast point feature histograms (fpfh) for 3d registration. *2009 IEEE International Conference on Robotics and Automation*, 3212–3217.
- Segal, A. V., Hähnel, D., Thrun, S., 2009. Generalized-icp. *Robotics: Science and Systems*.
- Stilla, U., Xu, Y., 2023. Change detection of urban objects using 3D point clouds: A review. *ISPRS J. Photogramm. Remote Sens.*, 197, 228–255.
- Takai, S., Date, H., Kanai, S., Niina, Y., Oda, K., Ikeda, T., 2013. Accurate registration of MMS point clouds of urban areas using trajectory. *ISPRS Ann. Photogramm. Remote Sens. Spatial Inf. Sci.*, II-5/W2, 277–282.
- Wang, C., Wen, C., Dai, Y., Yu, S., Liu, M., 2020. Urban 3D modeling using mobile laser scanning: a review. *Virtual Reality & Intelligent Hardware*, 2(3), 175–212. 3D Visual Processing and Reconstruction Special Issue.
- Wysocki, O., Xu, Y., Stilla, U., 2021. Unlocking point cloud potential: Fusing MLS point clouds with semantic 3D building models while considering uncertainty. *ISPRS Ann. Photogramm. Remote Sens. Spatial Inf. Sci.*, 8, 45–52.
- Xu, Z., Schwab, B., Yang, Y., Kolbe, T. H., Holst, C., 2026. L2M-Reg: Building-level uncertainty-aware registration of outdoor LiDAR point clouds and semantic 3D city models. *ISPRS J. Photogramm. Remote Sens.*, 233, 467–486.
- Yang, B., Wei, Z., Li, Q., Li, J., 2013. Semiautomated Building Facade Footprint Extraction From Mobile LiDAR Point Clouds. *IEEE Geosci. Remote Sens. Lett.*, 10(4), 766–770.
- Yang, Y., Holst, C., 2025. Piecewise-ICP: Efficient and robust registration for 4D point clouds in permanent laser scanning. *ISPRS J. Photogramm. Remote Sens.*, 227, 481–500.
- Zhou, Q.-Y., Park, J., Koltun, V., 2016. Fast global registration. B. Leibe, J. Matas, N. Sebe, M. Welling (eds), *Computer Vision – ECCV 2016*, Springer International Publishing, Cham, 766–782.
- Zhuo, X., Fraundorfer, F., Kurz, F., Reinartz, P., 2018. Optimization of OpenStreetMap Building Footprints Based on Semantic Information of Oblique UAV Images. *Remote Sensing*, 10(4).