Pixel-Accurate Registration of Photogrammetric Images and LiDAR in a Hybrid Airborne Oblique Imaging System

Deyan Deng[1], Rongjun Qin[1,2]*, Elisa Mariarosaria Farella[3], Fabio Remondino[3]

[1] Department of Civil, Environmental and Geodetic Engineering, The Ohio State University, Columbus, USA
[2] Department of Electrical and Computer Engineering, The Ohio State University, Columbus, USA
[3] 3D Optical Metrology (3DOM) unit, Bruno Kessler Foundation (FBK), Trento, Italy
{deng.1069, qin.324}@osu.edu, {elifarella, remondino}@fbk.eu

Keywords: Data Fusion, LiDAR, Photogrammetry, Depth Maps, View-Dependent Registration, Multi-View Stereo.

Abstract

Hybrid airborne imaging systems integrating oblique photogrammetric cameras and LiDAR sensors provide complementary geometric and radiometric information for high-fidelity 3D reconstruction. While LiDAR offers stable global geometric accuracy, photogrammetric reconstructions provide dense surface detail and rich façade information through oblique imaging. However, photogrammetric models often exhibit locally varying geometric drift caused by error accumulation during bundle adjustment, preventing pixel-accurate alignment between images and LiDAR data. Conventional global registration approaches are insufficient to correct such spatially varying misalignments. In this paper, we propose a view-dependent fusion framework for pixel-accurate registration between photogrammetric images and LiDAR data. The global alignment problem is decomposed into localized rigid registrations performed independently for each image. For each view, a local photogrammetric point cloud reconstructed from depth maps is rigidly aligned to a corresponding LiDAR subset using fixed-scale point-to-plane ICP. The estimated transformation is used to project LiDAR points into the image domain, where depth-consistency checks enforce visibility constraints and remove inconsistent measurements. Valid LiDAR points are colorized and fused with the photogrammetric reconstruction to generate anchored view-dependent point clouds. Finally, all view-dependent reconstructions are aggregated using voxel-grid fusion with median-based point selection. Experiments on airborne oblique imaging datasets demonstrate improved reconstruction completeness and pixel-level consistency between images and LiDAR geometry. The proposed framework provides a practical and robust solution for high-accuracy multi-modal 3D reconstruction in complex urban environments.

1. Introduction

Hybrid airborne imaging systems with oblique cameras and LiDAR (Toschi et al., 2018, Bacher, 2022, Farella et al., 2025) have received increasing attention nowadays, primarily due to their advantages in a number of scenarios where geometric accuracy and texture/semantic richness are required, such as in surveying very complex objects and facilities like powerplants, pylons, as well as for accurate geometric urban modeling (e.g., LoD (level of detail) models). More importantly, such a hybrid system can generate temporally consistent data of both modalities (image and LiDAR), serving as a valuable source of data. However, the registration of both sources at pixel level, despite the fact that they are in an integrated system, is still very challenging, due to the lack of supporting software/hardware infrastructure (calibration facilities, more nuanced technical considerations on deep integration of the camera, LiDAR, positional and control sensors) for such a system. Standard photogrammetric workflows suffer from non-linear cumulative drift, particularly in extended strips (Huang et al., 2024), causing misalignments when overlaid on LiDAR (Zeng, 2021). Crucially, conventional global rigid registration is insufficient to achieve pixel-level alignment. Since photogrammetric drift manifests as spatially varying deformations, a single "one-time" global transformation fails to correct local geometric deviations.

In this work, we propose a novel framework that synergizes the global fidelity of LiDAR with the local density of photogrammetry (Gressin, 2013). Unlike global registration methods that attempt to warp the entire model at once (Mandlburger et al., 2017, Glira et al., 2019, Bacher, 2021, Toschi et al., 2021, Qin et al., 2022, Xu et al., 2023), our approach is view-dependent. We treat the local geometry within each camera frustum as a rigid body. Building upon the initial georeferencing, we perform a fine-grained, local rigid registration ($SE(3)$) to "anchor" each MVS depth map to the LiDAR data. This decomposition allows us to achieve pixel-accurate registration within each local frame, effectively mitigating residual drift and interpolation errors.

* Corresponding author: qin.324@osu.edu

2. Related Work

2.1 LiDAR and Photogrammetry Registration

The integration of LiDAR and photogrammetric data has been extensively studied due to their complementary characteristics, where LiDAR provides accurate geometric measurements and images offer high-resolution texture and semantic information. Early approaches primarily relied on global rigid registration, often based on Iterative Closest Point (ICP) variants, to align photogrammetric point clouds with LiDAR reference data (Besl and McKay, 1992, Gressin, 2013, Yang and Chen, 2015, Haala et al., 2022, Alsadik et al., 2023, Glira et al., 2019). While effective for correcting global misalignments, such methods assume that residual errors can be adequately represented by a single rigid transformation, which is often violated in large-scale or strip-based photogrammetric reconstructions. More recent studies have explored non-rigid or piecewise registration strategies to account for spatially varying distortions between LiDAR and photogrammetric data (Xu et al., 2023). However, these approaches typically operate at the point cloud or surface level and do not explicitly consider pixel-level consistency between images and LiDAR geometry, which is critical for applications such as image-based analysis, texture mapping, and dense reconstruction refinement.

2.2 Photogrammetric Drift and Local Refinement

Photogrammetric Structure-from-Motion (SfM) and Multi-View Stereo (MVS) pipelines are known to suffer from accumulated drift, particularly in extended image sequences or weakly constrained configurations (Triggs, 2000). Although bundle adjustment (BA) provides a globally optimal solution under ideal conditions, practical factors such as uneven image overlap, imperfect camera calibration, and limited control information can result in locally inconsistent geometry (Huang et al., 2024). Several works have addressed this issue through strip adjustment, block deformation models, or by introducing external constraints such as ground control points or LiDAR-derived references (Wehr and Lohr, 1999, Zeng, 2021). These methods generally aim to improve global geometric consistency but often remain insufficient for achieving pixel-accurate alignment, as local deviations persist even after global optimization.

2.3 View-Dependent and Depth-Based Fusion

To overcome the limitations of global registration, view-dependent processing strategies have been proposed, where geometric refinement is performed at the level of individual images or local neighborhoods (Gressin, 2013). In parallel, depth map–based representations have gained popularity in photogrammetry and multi-view reconstruction due to their close relationship with image formation and their suitability for per-view processing (Furukawa and Hernández, 2015). Depth-based fusion methods typically integrate information from multiple sources by enforcing consistency in the image domain or along viewing rays, enabling more precise handling of occlusions and local geometric variations (**?**). Nevertheless, existing approaches often focus on photogrammetry-only workflows or do not explicitly leverage LiDAR as a local geometric anchor. In contrast, our work emphasizes view-dependent rigid anchoring of photogrammetric depth maps to LiDAR geometry, enabling pixel-accurate alignment while preserving the metric integrity of the LiDAR reference.

3. Methodology

3.1 Overview of the Proposed Method

Hybrid airborne imaging systems combining oblique photogrammetric cameras and LiDAR sensors provide complementary geometric and radiometric information. LiDAR offers stable global geometric accuracy and robustness to vegetation penetration, while photogrammetry delivers dense, high-resolution surface reconstruction and rich façade details through oblique imaging. However, despite being acquired within an integrated system, photogrammetric reconstructions often exhibit locally varying geometric drift caused by error accumulation during bundle adjustment, particularly in extended flight strips or complex urban scenes. Such drift prevents pixel-accurate alignment between images and LiDAR data and cannot be sufficiently corrected using a single global rigid transformation. In this work, we address the problem of pixel-accurate registration between photogrammetric images and LiDAR data under the following practical assumptions, which are standard in airborne mapping scenarios. First, both modalities are expressed in a common metric coordinate system through sensor calibration and direct georeferencing, implying consistent scale. Second, the sensor platform is rigid, and reliable initial camera poses are available from photogrammetric processing. Third, LiDAR measurements are assumed to provide a geometrically stable reference, while residual photogrammetric errors are primarily local rather than global. These assumptions allow the registration problem to be formulated as a sequence of localized rigid alignments rather than a single global correction. Based on these considerations, we propose a view-dependent fusion framework that decomposes the global image–LiDAR registration problem into three stages. In the first stage, view-dependent registration is performed independently for each image. A local photogrammetric point cloud is reconstructed from the image and its depth map, and a spatially corresponding subset of the LiDAR point cloud is extracted within the camera frustum. A rigid transformation in $SE(3)$ is then estimated using a point-to-plane ICP formulation to locally anchor the photogrammetric geometry to the LiDAR reference, correcting residual drift while preserving the global metric scale. In the second stage, the estimated view-dependent transformation is applied to the LiDAR subset and projected into the image domain to generate a sparse LiDAR depth map. A depth-consistency check is performed against the photogrammetric depth map to filter out occluded or inconsistent LiDAR points. The remaining valid LiDAR points are assigned color information from the image and fused with the photogrammetric point cloud, resulting in an anchored and geometrically enhanced view-dependent reconstruction. In the final stage, all view-dependent point clouds from multiple images are aggregated into a common global coordinate system. A voxel-grid-based fusion strategy is employed to reduce redundancy and balance point density while preserving the geometric accuracy inherited from the LiDAR reference. The resulting point cloud achieves improved completeness, reduced local drift, and pixel-level consistency between image observations and 3D geometry. The workflow of the proposed framework is summarized in Figure 1.

3.2 View-Dependent Registration

For each image I_i, the corresponding Multi-View Stereo (MVS) depth map is back-projected using the initial camera pose obtained from photogrammetric bundle adjustment to generate a local photogrammetric point cloud $\mathcal{P}_i$. This image-centric representation preserves the direct association between 3D points and image pixels and provides a suitable basis for pixel-level refinement. Although the photogrammetric reconstruction and the LiDAR point cloud are globally aligned, residual local misalignments remain due to accumulated photogrammetric drift, which motivates a localized registration strategy. To restrict the alignment problem to the spatial region relevant to each view, an oriented bounding box (OBB) is computed for $\mathcal{P}_i$ based on the camera pose and viewing direction. The use of an OBB accounts for the oblique imaging geometry and limits the registration to the volume observed by the camera, thereby avoiding the inclusion of distant or occluded LiDAR points. The global LiDAR point cloud is then spatially queried to extract a local subset $\mathcal{L}_i$ contained within the OBB, which serves as the geometric reference for view-dependent alignment. The alignment between $\mathcal{P}_i$ and $\mathcal{L}_i$ is formulated as a rigid registration prob-

https://doi.org/10.5194/isprs-archives-XLIX-B2-2026-143-2026

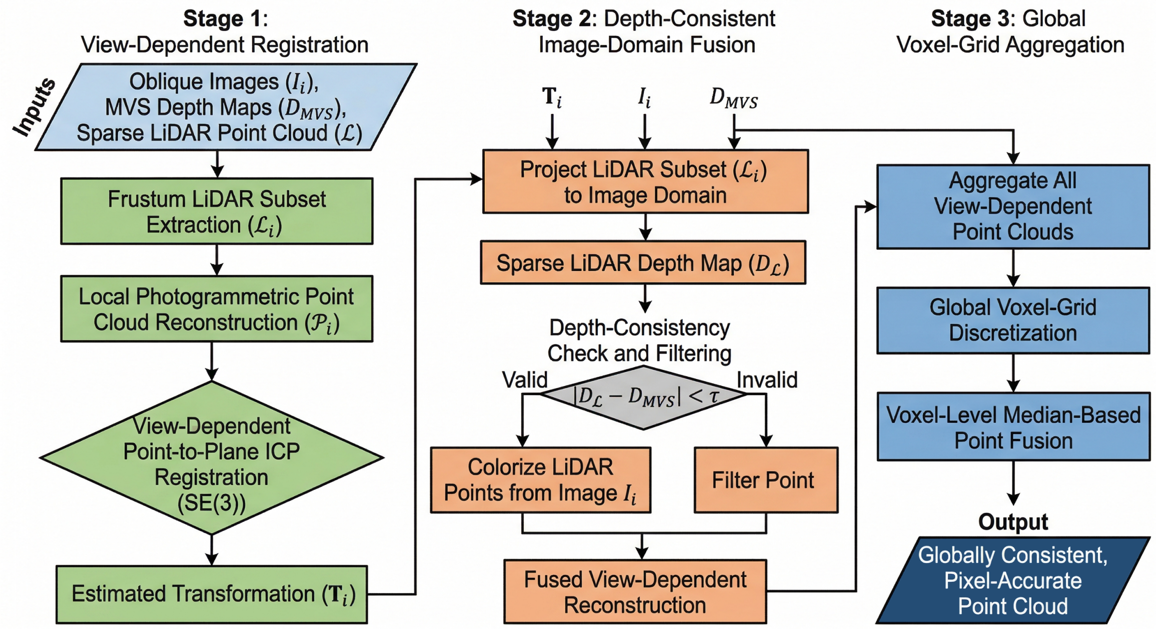

Figure 1. The workflow of the proposed framework.

lem. Although Iterative Closest Point (ICP) algorithms can be extended to estimate similarity transformations in $Sim(3)$, including an additional scale parameter, such an extension is unnecessary in the present context. Both photogrammetric and LiDAR data are already expressed in a common metric scale through sensor calibration and direct georeferencing. Estimating scale would therefore introduce unnecessary degrees of freedom and may lead to instability, particularly when the registration is performed within a spatially limited camera frustum. Consequently, the transformation is constrained to a rigid-body motion in $SE(3)$, with the scale implicitly fixed to unity. A point-to-plane ICP formulation is employed to estimate the rigid transformation $\mathbf{T}_i = (\mathbf{R}_i, \mathbf{t}_i)$ that aligns $\mathcal{P}_i$ with $\mathcal{L}_i$. This choice is motivated by the distinct sampling characteristics of the two modalities. LiDAR point clouds typically exhibit lower but more uniform spatial resolution, whereas photogrammetric reconstructions are significantly denser and may contain locally varying noise. Direct point-to-point ICP can therefore bias the optimization toward densely sampled photogrammetric regions and degrade geometric fidelity. In contrast, the point-to-plane metric minimizes the distance between photogrammetric points and locally approximated tangent planes of the LiDAR surface, thereby reducing sensitivity to point density discrepancies and better preserving the underlying geometric structure. Formally, given corresponding points $p_j \in \mathcal{P}_i$ and $q_j \in \mathcal{L}_i$, together with surface normals $\mathbf{n}_j$ estimated from the LiDAR data, the optimization problem is expressed as

$$\min_{\mathbf{R}_i, \mathbf{t}_i} \sum_j \left((\mathbf{R}_i p_j + \mathbf{t}_i - q_j) \cdot \mathbf{n}_j \right)^2 \tag{1}$$

where $\mathbf{R}_i$ is constrained to be a valid rotation matrix, i.e., $\mathbf{R}_i \in SO(3)$. The optimization is initialized using the identity transformation, relying on the globally consistent georeferencing as a strong prior. Due to the relatively small residual misalignments present after photogrammetric processing, convergence is typically achieved within a limited number of iterations. The resulting transformation $\mathbf{T}_i$ corrects local photogrammetric drift within the camera frustum while preserving the global metric consistency of the LiDAR reference. This view-dependent rigid alignment forms the foundation for subsequent pixel-level fusion in the image domain.

3.3 Depth-Consistent Image-Domain Fusion

After view-dependent registration, the estimated rigid transformation $\mathbf{T}_i$ aligns the local LiDAR subset $\mathcal{L}_i$ with the photogrammetric point cloud $\mathcal{P}_i$ within the coordinate frame of image I_i. While this alignment corrects local geometric drift in 3D space, achieving pixel-accurate consistency between image observations and LiDAR geometry requires reasoning directly in the image domain. To this end, we perform a depth-consistent fusion process that integrates LiDAR information at the pixel level. The transformed LiDAR points are first projected into the image plane using the camera intrinsic parameters and the registered pose. For each LiDAR point $\ell_k \in \mathcal{L}_i$, its image projection (u_k, v_k) and corresponding depth value $D_\mathcal{L}(u_k, v_k)$ are computed. This procedure yields a sparse LiDAR depth map that is directly comparable to the dense photogrammetric depth map D_{MVS} generated from Multi-View Stereo. Due to differences in viewpoint, sampling density, and surface visibility, not all projected LiDAR points are consistent with the photogrammetric reconstruction. In particular, LiDAR points may correspond to surfaces that are occluded in the image or lie outside the effective depth support of the photogrammetric model. To address this issue, a depth-consistency check is applied by comparing LiDAR and photogrammetric depths at each projected pixel location. A LiDAR point is considered valid if

$$|D_\mathcal{L}(u_k, v_k) - D_{MVS}(u_k, v_k)| < \tau, \tag{2}$$

where τ is a tolerance threshold that accounts for measurement noise and residual registration error. This filtering step effectively removes LiDAR points that violate image-domain visibility constraints, such as those originating from background structures or surfaces occluded by foreground geometry. As a result, only LiDAR points that are geometrically and photometrically consistent with the image observation are retained for fusion. Compared to purely 3D-based filtering strategies, this image-domain validation ensures that the retained LiDAR points are directly supported by image evidence, which is essential for pixel-accurate registration. For each valid LiDAR point, color information is sampled from the image at the corresponding pixel location and assigned to the 3D point. The colored LiDAR points are then merged with the photogrammetric point cloud $\mathcal{P}_i$ to form an anchored and enhanced view-dependent reconstruction. In regions where photogrammetric reconstruction is sparse or incomplete, such as textureless surfaces or areas affected by poor image overlap, LiDAR points contribute to improved completeness. Conversely, in regions with rich photogrammetric detail, the dense MVS points dominate the reconstruction while remaining geometrically anchored by the LiDAR reference. Through this depth-consistent image-domain fusion, each view-dependent point cloud achieves improved geometric accuracy, reduced local drift, and enhanced completeness, while maintaining strict pixel-level consistency between 3D geometry and image observations. These view-dependent reconstructions form the input for the subsequent global aggregation stage.

3.4 Global Aggregation of View-Dependent Reconstructions

Following view-dependent registration and depth-consistent image-domain fusion, each image yields a locally anchored point cloud that is geometrically consistent with the LiDAR reference and pixel-aligned with the corresponding image observation. These view-dependent point clouds are expressed in a common global coordinate system through the use of georeferenced camera poses and locally estimated rigid transformations. The goal of the global aggregation stage is to integrate all view-dependent reconstructions into a single, coherent point cloud representation. A direct union of all view-dependent point clouds would introduce substantial redundancy due to overlapping observations across multiple views and would result in highly non-uniform point density. In particular, photogrammetric reconstructions are typically much denser than LiDAR measurements, which can lead to oversampling and increased noise if no consolidation strategy is applied. To address these issues, we adopt a voxel-grid-based aggregation approach that balances point density while preserving geometric fidelity. The global space is discretized into a regular voxel grid with a predefined resolution. Within each voxel, multiple candidate points originating from different views and sensing modalities may be present. The selection of a representative point is therefore non-trivial and can influence the stability of the aggregated reconstruction. Several voxel-level fusion strategies can be considered in this context, including arithmetic averaging, weighted averaging, geometry-based selection, and median-based fusion. An arithmetic averaging reduces random noise but is sensitive to outliers, particularly when combining dense photogrammetric points with sparser LiDAR measurements. Weighted averaging requires reliable per-point confidence measures, which are difficult to estimate consistently across heterogeneous data sources. Geometry-based selection strategies, such as choosing points associated with favorable intersection angles, may become unstable in regions with limited view diversity. Based on these considerations, median-based fusion is adopted in the proposed framework. By selecting the median position of points within each voxel, this strategy effectively suppresses outliers without requiring additional weighting parameters or confidence estimation. As a result, the aggregated point cloud preserves sharp structural features while maintaining robustness across regions with varying point density and modality contribution. Through voxel-grid aggregation with median-based fusion, the proposed framework produces a dense and globally consistent point cloud that combines the complementary strengths of photogrammetric and LiDAR data. This aggregated reconstruction serves as the final output for subsequent analysis and evaluation.

4. Experiments

We evaluated the proposed method on an airborne hybrid imaging dataset consisting of 160 oblique and nadir images and a co-registered LiDAR point cloud. The data were acquired using an UltraCam Dragon 4.1 system (Farella et al., 2025), which integrates high-resolution multi-view imagery with a LiDAR scanner operating at a pulse repetition rate of 2.4 MHz. The dataset covers a complex urban area characterized by large building facades, roof structures, and varying surface materials, providing a challenging test case for pixel-accurate multi-modal registration. Photogrammetric processing was performed using our in-house pipeline, MetricSfM (Elhashash and Qin, 2022, Remondino et al., 2022, Huang et al., 2022, Song et al., 2024, Chen et al., 2024, Huang and Qin, 2025), for aerial triangulation and camera pose estimation, followed by dense depth map generation. The LiDAR point cloud was treated as the geometric reference due to its stable global accuracy. All data were processed in a common georeferenced coordinate system, ensuring consistent metric scale between modalities.

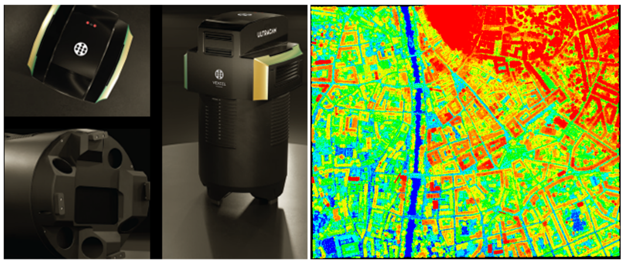

Figure 2. The hybrid UltraCam Dragon 4.1 system (left). The LiDAR point cloud used as the geometric anchor for the photogrammetric data (right).

4.1 Qualitative Evaluation

Figure 3 highlights the improvement in surface completeness. Prior to integration (Figure 3, top), the photogrammetric point cloud exhibits characteristic gaps on surfaces with low texture or unfavorable lighting, which are common in dense urban ”canyons.” By leveraging the registered LiDAR subset as a geometric scaffold, the proposed method populates these gaps with high-accuracy LiDAR measurements. These points are subsequently colorized using the depth-consistent image-domain fusion described in Section 3.3. The result (Figure 3, bottom) is a continuous, high-density surface that maintains the radiometric richness of the photography while filling the ”holes” typical of pure MVS pipelines. This synergy is particularly effective for oblique views of building facades, where photogrammetric density often fluctuates.

4.2 Quantitative Evaluation

Image–geometry consistency analysis. To rigorously assess the ”pixel-accurate” claim of the proposed framework, we analyze the geometric consistency between the photogrammetric reconstructions and the LiDAR reference through reprojection and profile analysis. Figure 4 provides a visual inspection of

Figure 3. One-view point cloud comparison: The **top** image shows the initial photogrammetry result, while the **bottom** image illustrates the improved completeness and consistency after registration and LiDAR integration.

LiDAR reprojection quality. In this visualization, the LiDAR points are color-coded based on surface normal clustering, where each color represents a distinct planar facet of the buildings. Prior to local anchoring, noticeable offsets are present where LiDAR points fail to align with strong image gradients and structural edges, such as rooflines and window frames. Following the view-dependent registration, the LiDAR reprojection aligns precisely with these features. The seamless alignment of these colored clusters with their corresponding image boundaries demonstrates that our method effectively achieves pixel-level consistency between the 2D image observations and the 3D geometry, which is essential for high-fidelity texture mapping. We further extracted height profiles along representative cross-sections of building roofs and facades (Figure 5). Before local anchoring, the photogrammetric reconstruction (Blue) exhibits systematic deviations from the LiDAR reference (Green), reflecting the residual local drift accumulated during bundle adjustment. After applying the proposed view-dependent registration, the corrected photogrammetric surface (Red) shows a significantly reduced deviation, accurately following the LiDAR-anchored reference across both planar surfaces and high-frequency structural features.

Analysis of view-dependent residuals. While profile and reprojection analyses demonstrate improved alignment after registration, they do not explicitly characterize the spatial structure of the residual misalignment. To this end, we further analyze per-view geometric residuals before and after view-dependent registration. For each image I_i, point-to-plane distances are computed between the local photogrammetric point cloud $\mathcal{P}_i$ and its corresponding LiDAR subset $\mathcal{L}_i$ under the initial georeferencing. The residuals are summarized per view using robust statistics, including the median absolute distance and interquartile range (IQR), as reported in Table 1. The resulting per-view residual statistics exhibit substantial variation across views prior to registration, indicating that photogrammetric drift is not spatially uniform over the scene. This heterogeneity implies that the residual misalignment cannot be adequately modeled or corrected by a single global rigid transformation. After applying the proposed view-dependent registration, the per-view residual distributions consistently contract, with reduced median error and variance across views. These results quantitatively confirm that localized rigid anchoring is necessary to achieve pixel-level consistency between images and LiDAR geometry.

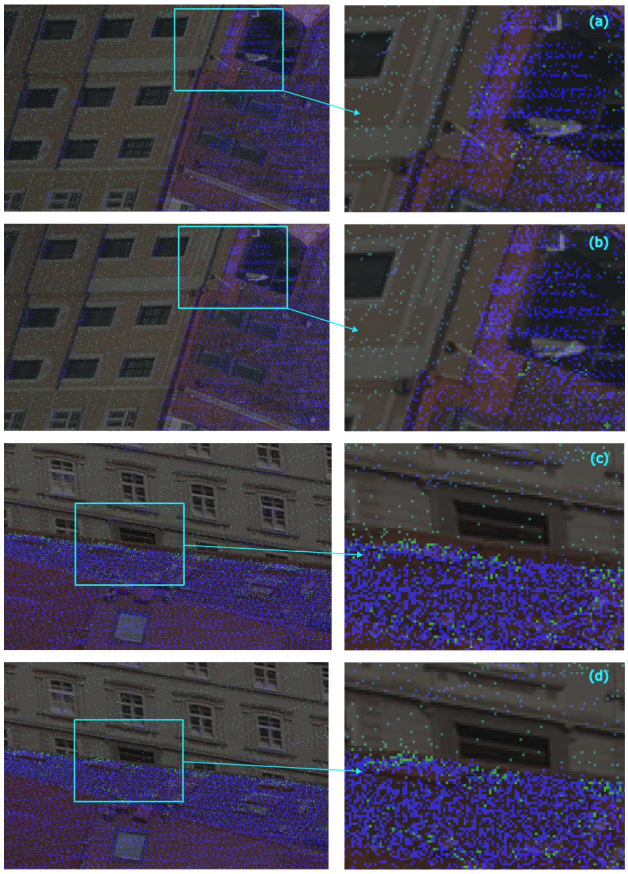

Figure 4. Visual inspection of LiDAR reprojection before (a, c) and after (b, d) registration. The point cloud is clustered based on surface normals, with different colors representing distinct clusters to highlight the alignment of planar facets with structural edges (building roofs).

Table 1. Statistics of view-dependent point-to-plane residuals [m] before and after view-dependent registration. Residuals are computed per view and summarized using robust statistics.

Method	Median	IQR	P90
Before registration	0.35	0.18	0.54
After registration	0.20	0.08	0.33

Global aggregation strategy evaluation. After view-dependent registration, multiple locally anchored point clouds must be aggregated into a single global reconstruction. We therefore evaluate several voxel-level fusion strategies for combining view-dependent results.

https://doi.org/10.5194/isprs-archives-XLIX-B2-2026-143-2026

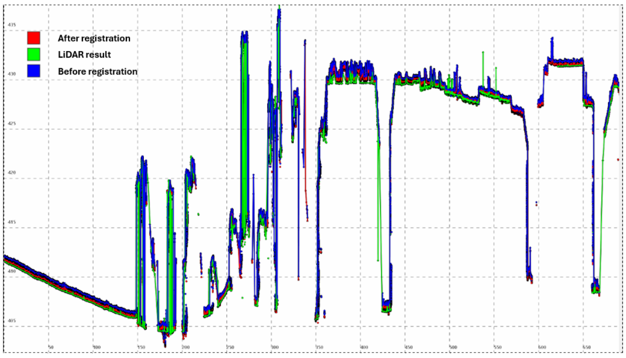

Figure 5. Vertical profile analysis demonstrating that the registered elevation (Red) aligns closely with the LiDAR reference (Green), significantly reducing the initial geometric drift (Blue).

- **Mean-based voxel fusion:** The representative point is computed as the arithmetic mean of all point coordinates within a voxel. This approach reduces random noise by minimizing overall deviation but is sensitive to outliers and uneven point density.
- **Maximum intersection angle selection:** The representative point is selected based on geometric reliability, favoring the observation associated with the largest intersection angle, which typically corresponds to stronger triangulation geometry.
- **Intersection-angle–weighted fusion:** Points within a voxel are combined using weights derived from their intersection angles, allowing contributions from multiple views while emphasizing geometrically favorable observations.
- **Median-based voxel fusion:** The representative point is determined by the component-wise median of point coordinates within a voxel. This strategy effectively suppresses outliers without requiring additional confidence measures or weighting parameters.

Figure 6 shows a representative height profile extracted from the globally aggregated point clouds produced using different voxel-level fusion strategies. The profile highlights the impact of fusion strategy selection on geometric stability and structural preservation. Mean-based fusion exhibits noticeable smoothing effects, particularly around sharp transitions, due to averaging across mixed-density observations. Fusion strategies based on intersection geometry show increased variability in regions with changing view configurations, reflecting their sensitivity to local viewing conditions.

In contrast, median-based voxel fusion closely follows the LiDAR-anchored reference profile while maintaining sharp structural features and reduced variance along the profile. By suppressing outliers without introducing smoothing artifacts, median-based fusion provides a stable and robust aggregation result across regions with heterogeneous point density and modality contributions.

4.3 Summary of Results

The experimental results demonstrate that the proposed view-dependent fusion framework effectively mitigates local photogrammetric drift and achieves improved geometric consistency with the LiDAR reference. By enforcing depth consistency in

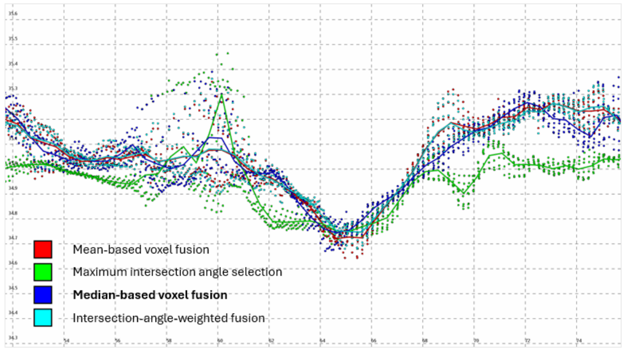

Figure 6. Height profile comparison of global aggregation results using different voxel-level fusion strategies. Median-based voxel fusion produces the most stable profile and preserves sharp geometric features, while mean-based and geometry-based fusion strategies exhibit increased smoothing or local instability.

the image domain and aggregating view-dependent reconstructions through voxel-based fusion, the method produces dense, globally consistent point clouds with enhanced completeness and pixel-level alignment. These results confirm the suitability of the proposed approach for high-fidelity multi-modal 3D reconstruction in complex urban environments.

5. Discussion and Conclusions

In this paper, we presented a view-dependent fusion framework for pixel-accurate registration of photogrammetric images and LiDAR data acquired from a hybrid airborne oblique imaging system. By decomposing the global image–LiDAR registration problem into a sequence of localized rigid alignments, the proposed method effectively mitigates residual photogrammetric drift that cannot be corrected using conventional global registration approaches. The integration of fixed-scale, point-to-plane ICP with depth-consistent image-domain fusion enables accurate anchoring of dense photogrammetric reconstructions to the geometrically stable LiDAR reference. Experimental results confirm that this localized strategy significantly improves geometric accuracy and reconstruction completeness, particularly along complex urban building facades. While the approach slightly relaxes global epipolar consistency to prioritize local alignment, the resulting reduction in drift and improved pixel-level fidelity provides a superior outcome for high-fidelity surface modeling. Future research will focus on reconciling the localized registration results with global geometric constraints. Specifically, we aim to incorporate Pose Graph Optimization (PGO) to resolve the discrepancies introduced by independent per-view adjustments. By treating the view-dependent $SE(3)$ transformations as observations and the relative camera-to-camera epipolar constraints as edges in a pose graph, we can minimize the ambiguity of relative poses and restore global consistency across the image block. Additionally, we plan to extend this framework to environments with sparse LiDAR coverage while incorporating adaptive, per-point confidence measures to further enhance robustness in challenging mapping scenarios.

Acknowledgements

This work is supported by the Office of Naval Research (Award No. N000142312670) and Intelligence Advanced Research

Projects Activity (IARPA) via Department of Interior/Interior Business Center (DOI/IBC) contract number 140D0423C0075.

References

Alsadik, B., Nex, F., Yadav, Y., 2023. Hybrid adjustment of uas-based lidar and image data. *5th Geospatial Week 2023, GSW 2023*, Copernicus, 633–640.

Bacher, U., 2021. 3D Content Generation using Hybrid Aerial Sensor Data. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 43, 297–303.

Bacher, U., 2022. Hynrid aerial sensor data as basis for a geospatial digital twin. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLIII-B4-2022, 653–659.

Besl, P., McKay, N., 1992. A method for registration of 3d shapes. *PAMI*.

Chen, G., Wu, J., Chen, H., Teng, W., Gao, Z., Feng, A., Qin, R., Zhao, Y., 2024. Geometry-aware feature matching for large-scale structure from motion. *International Conference on 3D Vision*.

Elhashash, M., Qin, R., 2022. Investigating spherical epipolar rectification for multi-view stereo 3D reconstruction. *ISPRS Annals Photogramm. Remote Sens. Spatial Inf. Sci*, V-2-2022, 47–52.

Farella, E. M., Morelli, L., Remondino, F., Qin, R., Schachinger, B., Legat, K., 2025. Investigating the New Ultracam Dragon Hybrid Aerial Mapping System. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLVIII-M-6-2025, 117-125.

Furukawa, Y., Hernández, C., 2015. Multi-View Stereo: A Tutorial. *Foundations and Trends® in Computer Graphics and Vision*, 9(1–2), 1–148.

Glira, P., Pfeifer, N., Mandlburger, G., 2019. Hybrid orientation of airborne lidar point clouds and aerial images. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 4, 567–574.

Gressin, A. e. a., 2013. Semiglobal Registration of Point Clouds. *ISPRS Journal*.

Haala, N., Kölle, M., Cramer, M., Laupheimer, D., Zimmermann, F., 2022. Hybrid georeferencing of images and LiDAR data for UAV-based point cloud collection at millimetre accuracy. *ISPRS Open Journal of Photogrammetry and Remote Sensing*, 4, 100014.

Huang, D., Qin, R., 2025. Uncertainty estimation for photogrammetric point clouds of UAV imagery. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences (ISPRS Geospatial Week)*.

Huang, D., Qin, R., Elhashash, M., 2024. Bundle Adjustment with Motion Constraints for Uncalibrated Multi-Camera Systems at The Ground Level. *ISPRS Journal of Photogrammetry and Remote Sensing*, 211, 452–464.

Huang, D., Tang, Y., Qin, R., 2022. An evaluation of PlanetScope images for 3D reconstruction and change detection – experimental validations with case studies. *GIScience & Remote Sensing Journal*, 59(1), 744–761.

Mandlburger, G., Wenzel, K., Spitzer, A., Haala, N., Glira, P., Pfeifer, N., 2017. Improved topographic models via concurrent airborne LiDAR and dense image matching. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 4(2/W4), 259–266.

Qin, R., Ling, X., Farella, E. M., Remondino, F., 2022. Uncertainty-guided Depth Fusion from Multi-view Satellite Images to Improve the Accuracy in Large Scale DSM Generation. *Remote Sensing*, 14(6), 1309.

Remondino, F., Morelli, L., Stathopoulou, E. K., Elhashash, M., Qin, R., 2022. Aerial triangulation with learning-based tie points. *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLIII-B2-2022, 77–84.

Song, S., Morelli, L., Wu, X., Qin, R., Albanwan, H., Remondino, F., 2024. Deep learning meets satellite images – an evaluation on handcrafted and learning-based features for multi-date satellite stereo images. *2nd Workshop on Traditional Computer Vision in the Age of Deep Learning (TradiCV), in European Conference on Computer Vision (ECCV)*.

Toschi, I., Farella, E. M., Welponer, M., Remondino, F., 2021. Quality-based registration refinement of airborne LiDAR and photogrammetric point clouds. *ISPRS Journal of Photogrammetry and Remote Sensing*, 172, 160–170.

Toschi, I., Remondino, F., Rothe, R., Klimek, K., 2018. Combining airborne oblique camera and lidar sensors: investigation and new perspectives. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 42, 437–444.

Triggs, B. e. a., 2000. Bundle Adjustment—A Modern Synthesis. *Vision Algorithms*.

Wehr, A., Lohr, U., 1999. *Airborne Laser Scanning—An Introduction and Overview*.

Xu, N., Qin, R., Song, S., 2023. Point Cloud Registration for LiDAR and Photogrammetric Data: a Critical Synthesis and Performance Analysis on Classic and Deep Learning Algorithms. *ISPRS Open Journal of Photogrammetry and Remote Sensing*, 100032.

Yang, B., Chen, C., 2015. Automatic registration of UAV-borne sequent images and LiDAR data. *ISPRS Journal of Photogrammetry and Remote Sensing*, 101, 262–274.

Zeng, L. e. a., 2021. A Review of LiDAR–Photogrammetry Fusion Techniques. *ISPRS Journal*.