Extraction of Image-to-Lidar Correspondences and their Impact on Optimal Sensor Fusion

Kyriaki Mouzakidou *, Aurélien A. Brun, Jan Skaloud

Earth Sensing & Observation Laboratory (ESO), Ecole Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland
(kyriaki.mouzakidou, aurelien.brun, jan.skaloud)@epfl.ch

Keywords: Pixel-to-point correspondences, Hybrid sensor systems, Georeferencing, Dynamic networks

Abstract

This work extends our initial proof-of-concept via emulations on the benefits of relative spatial constraints between imagery and lidar point clouds in a factor graph based optimization with satellite positioning (GNSS) and raw inertial readings (Mouzakidou et al., 2025). Here, we demonstrate practically the automatic extraction and integration of 2D-3D correspondences established in the 3D domain within rough natural terrain flown over by an aircraft with sensors of high quality. We show that considering cross-domain (i.e. 2D-3D) constraints enables the calibration of internal camera parameters and its boresight on job, i.e. within mapping flight configurations, where conventional approaches fail. The common optimization of raw IMU data with such constraints improves the respective agreements between the lidar and image dense clouds, achieving consistency at ground resolution level, which is not the case for the conventional (standard) processing of acquired data.

1. Introduction

Increasing demands for (i) high spatial resolution, (ii) accurate georeferencing and (iii) rich texture and geometric detail in 3D models have led to the development of redundant, hybrid camera–lidar sensor systems (CityMapper, UltraCam Dragon). Despite sharing the same navigation sensor input, imagery and lidar data are typically processed separately in cascaded optimization workflows, often resulting in relative inconsistencies between the derived models (e.g. lidar vs photogrammetric point cloud). Consequently, tighter and more rigorous sensor fusion becomes not advantageous but essential.

Indeed, image and lidar observations are often fused together in a ”hybrid” adjustment to constrain imperfections in the INS/GNSS trajectory solution (Mandlburger et al., 2017, Pentek et al., 2020, Hussnain et al., 2021, Zhou et al., 2021, Haala et al., 2022, Jonassen et al., 2023). These approaches rely on different types of correspondences to link the imagery and lidar point clouds, either in the 2D or the 3D domains, with the goal of improving their co-registration. (Mouzakidou et al., 2025) proposed an image-to-lidar (2D-3D) observation model based on cross-domain correspondences. This spatial constraint can be introduced in single-step factor graph based optimizations, together with GNSS and raw inertial readings, to constrain the trajectory solution in a way that allows to mitigate certain errors directly at their source. The advantage in this approach is that although the factor graph formulation is solved through a linearized least-squares adjustment, there is no need to re-extract the observation constraints iteratively, unlike (Glira et al., 2019).

1.1 Definition of Correspondences

In optical datasets, correspondences are established by matching features that represent distinctive points, patterns, or structures across multiple images or modalities. This section clarifies the type of features considered as spatial constraints for optical observations. We differentiate between object-level and point-level features.

Object-level features are designed to capture high level semantic information, which is a more abstract, generalized perception of data, enriched with contextual understanding, e.g. lines (Hasheminasab et al., 2023), buildings (Wang et al., 2022) or trees (Lee et al., 2016). Such features represent a more complex and semantically rich structure than a single point, as they ”encode” the entire object instance or region (Kharroubi et al., 2025). They are primarily used in tasks such as detection, segmentation, or scene understanding, rather than in detailed geometric reconstruction.

Centered ”point-level” features, i.e. 2D pixel coordinates in images and 3D coordinates in point clouds, are better suited for precise geometric reconstruction as they are less dependent on semantic interpretation and viewing angle. Although feature detection considers a neighborhood around each center (a so-called patch), it is ultimately represented as a single 2D or 3D set of coordinates. As such, ”point-level” features enable more reliable generation of correspondences between overlapping images and are fundamental, e.g., for triangulation via the collinearity equation (Förstner and Wrobel, 2016). In lidar point clouds, they are also proven valuable for scan alignment via trajectory refinement (Brun et al., 2025). This work focuses on ”point-level” features stemming from different modalities.

1.2 Evolution of Extraction Approaches

Feature extraction can be categorized as single-domain or cross-domain. In single-domain extraction, both datasets belong in the same domain, either 2D, as in image-to-image matching in photogrammetry, or 3D, as in cloud-to-cloud matching for lidar data. In cross-domain extraction, the datasets lie in different domains, typically 2D and 3D. Methodologically and as discussed in the following, approaches are either hand-crafted, based on geometric, photometric, or physical principles and designed using expert knowledge, or learned, relying on neural networks trained to automatically extract and match features.

---

* Corresponding author

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
https://doi.org/10.5194/isprs-annals-XI-1-2026-175-2026 | © Author(s) 2026. CC BY 4.0 License.

1.2.1 Single-domain

In the 2D domain, feature matching based on hand-crafted methods is well-established, meaning well studied and explored, as reported, e.g. in (Lowe, 2004, Hartmann et al., 2016). However, recently, there has been a transition into learned methods, e.g. as reported in (Yi et al., 2016, Chen et al., 2021, Remondino et al., 2022). Learned methods have been proven more effective in difficult scenarios, like multi-temporal datasets (Zhang et al., 2021, Farella et al., 2022), multi-modal datasets, e.g. RGB to intensity (Li et al., 2020), varying lighting conditions (Verdie et al., 2015), or significant viewpoint differences (Sarlin et al., 2020, Lindenberger et al., 2023, Morelli et al., 2024a).

In the 3D domain, feature matching was also initially relied on hand-crafted approaches, e.g. (Rusu et al., 2009, Salti et al., 2014), as reviewed in (Guo et al., 2016). These approaches exploit geometric properties such as surface normals, curvature, or edge structures to establish repeatable features across different viewpoints. However, similar to the evolution observed in image features, there has been a notable shift toward learned 3D methods, which showed higher efficiency, particularly under variations in point density, viewpoint, and occlusion, as demonstrated in (Charles et al., 2017, Deng et al., 2018, Gojcic et al., 2019, Spezialetti et al., 2020, Brun, 2021, Cheng et al., 2025).

1.2.2 Cross-domain

Cross-domain feature matching has achieved limited success with hand-crafted methods, as these typically rely on projecting 3D data into 2D representations (Hussnain et al., 2019, Hasheminasab et al., 2023) or on converting to single-domain, for example, generating photogrammetric point clouds for 3D coregistration (Morelli et al., 2024b, Wang et al., 2025) or ”rasterizing” lidar clouds for 2D matching (Zhu et al., 2021). Projection-based approaches are inherently sensitive to trajectory errors or inaccuracies in system parameters, which can lead to significant georeferencing and subsequent reprojection errors. Converting 3D data into the 2D domain inevitably results in a loss of dimension information. Finally, approaches that rely on photogrammetric point clouds introduce a cascade of adjustments, starting with Assisted Aerial Triangulation (AAT), i.e. aerial triangulation supported by GNSS/INS navigation data, followed by dense image matching to generate the point cloud, which is then matched to the lidar cloud and jointly adjusted to produce a refined final output.

The fundamentally different nature of the two datasets, i.e. texture in images versus geometry in point clouds, has recently evolved towards (direct) learned feature extraction techniques. This is further detailed in [(Mouzakidou et al., 2025) - Sec. 3].

1.3 Contributions

This work advances our preliminary investigations on the perspectives of using image-to-lidar (2D-3D) constraints with GNSS and raw inertial readings in the single-step factor graph based optimization of Dynamic Network (DN) (Mouzakidou et al., 2025). In that first study, we introduced the cross-domain observation model and implemented the proof-of-concept using emulated 2D-3D correspondences. In this work, we demonstrate the automatic extraction and integration of 2D-3D correspondences, established in the 3D domain using the approach introduced in (Brun et al., 2025). We study the effects of introducing the cross-domain constraints against the industry standard approaches in a long-range Airborne Laser Scanning (ALS) dataset with validation over rough natural terrain, where geometrical primitives, such as straight lines or planes, are absent.

The rest of the paper is organized as follows. In Sec. 2, we describe the workflow for extracting 2D-3D correspondences in the 3D domain. In Sec. 3, we briefly review the observation model proposed in (Mouzakidou et al., 2025) and its integration in the single-step DN adjustment. Sec. 4 outlines the experimental evaluation including the long-range ALS dataset and the related case studies. In Sec. 5, we present and analyze the obtained results, before drawing the conclusions in Sec. 6.

2. 2D-3D Correspondence Extraction

The proposed extraction of 2D-3D correspondences is based on feature matching in the 3D domain, i.e. point-to-point pairs, following the approach of (Brun et al., 2025) with the difference of using one photogrammetric point cloud. The steps we follow to obtain the final 2D-3D pixel-to-point pairs used in the sensor fusion are the following:

1. Coarsely georeference the lidar point cloud using the INS/GNSS trajectory and available lidar mounting parameters (lever-arm and boresight angles), while keeping track of acquisition timestamps and 3D laser vectors for each generated 3D point, e.g using LIEO (Skaloud et al., 2010). Note that the term ’laser vector’ refers to the 3D spatial vector expressed in the lidar sensor frame that defines an acquired 3D point.

2. Generate the photogrammetric point cloud after triangulation (AAT), in the same coordinate system as the lidar data, using initial estimates of camera interior orientation (IO) parameters and mounting configuration (lever-arm and boresight angles, if available).

3. Apply the advanced point-to-point extraction workflow of (Brun et al., 2025), to obtain 3D point pairs between the lidar and photogrammetric point clouds. The quality of the extracted 3D point pairs (i.e. the distribution of pairwise distances) is stored and used as *a-priori* uncertainty for the 2D-3D constraints in the subsequent DN adjustment. This uncertainty is related to the GSD of the imagery as discussed in 5.1.

4. Project the 3D counterparts from the photogrammetric point cloud onto all available images to obtain the corresponding 2D coordinates and, ideally, if prior terrain model is available, handle the occlusions (in our implementation, this will be addressed in the future).

5. Pair the 2D pixel coordinates from step 4 with the corresponding 3D laser vectors from the lidar point cloud, forming new observations to be introduced into the optimization via dynamic network (Sec. 3).

3. 2D-3D Constraint in Sensor Fusion

This section summarizes the approach introduced in (Mouzakidou et al., 2025). The proposed cross-domain observation model (Eq. 1) represents a spatial constraint between a 2D pixel on the imagery and a 3D point on the point cloud and is suitable for integration in a general form of factor graph based optimizations, like e.g. he dynamic networks (Colomina and Blázquez, 2004, Rouzaud and Skaloud, 2011, Cucci et al., 2017) (Fig. 1), with other spatiotemporal constraints summarized in [(Mouzakidou et al., 2024) - Tab. 1].

$$\ell_{cl} + v_{cl} = \underbrace{\ell^{c_i}_{(p)}}_{\substack{\text{image}\\ \text{observation}}} - \underbrace{\Xi\left[\pi\left[K\widetilde{\Pi}\left[\widetilde{\Gamma}^m_{b(t_i)}\Gamma^b_c\right]^T\widetilde{\Gamma}^m_{b(t_j)}\Gamma^b_l\ell^{l_j}_{(P)}\right]\right]}_{\substack{\text{lidar observation}\\ \text{projected to image coordinates}}} \quad (1)$$

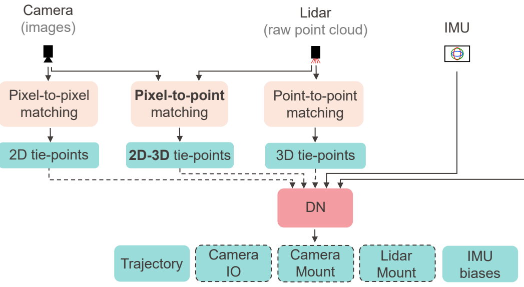

Figure 1. Integration of the new type of optical 2D-3D constraints in the adjustment. The optional input measurements and readjusted parameters are indicated with dashed line.

Assuming that a 2D-3D correspondence is established following the steps under Sec. 2 between a 2D pixel p and a 3D lidar point P, Eq. 1 links the 2D image observation $\ell^{c_i}_{(p)}$, in pixel units, with the homologous lidar observation, i.e. the 3D laser vector $\ell^{l_j}_{(P)}$ in meters. The laser vector is projected to image coordinates given (i) the camera pose at which the image is taken ($\widetilde{\Gamma}^m_{b(t_i)}$[1]), (ii) the lidar pose at which the laser vector is acquired ($\widetilde{\Gamma}^m_{b(t_j)}$[1]), (iii) camera internal parameters (K[2] and $\Xi(\cdot)$[3]), (iv) camera and (v) lidar mounting parameters (Γ^b_c and Γ^b_l respectively)[4], i.e. boresight matrices and lever arms. Term ℓ_{cl} comprises a so-called *zero*-observation edge, or a *pseudo*-observation in the network terminology, corresponding to no actual sensor measurement, v_{cl} is a $[2 \times 1]$ vector of zero mean Gaussian noise representing the reprojection residuals in pixels, $\pi(\cdot)$ is the projection function[5], and $\widetilde{\Pi}$ is an auxiliary matrix that handles the homogeneous coordinates[6]. Following the factor graph formulation of DN described in (Cucci et al., 2017), and as shown schematically in Fig. 1, these constraints are introduced into the optimization as the camera-to-lidar edges connecting two pose nodes and optionally, the camera and lidar boresight nodes. Thus, the uncertainty of the extracted correspondences stochastically conditions the trajectory solution through the improved determination of navigation sensor errors (e.g. inertial biases) and system and sensor parameters.

4. Experimental Evaluation

This section presents the experimental design used to evaluate the impact of integrating 2D–3D constraints, extracted in the 3D domain, into the DN. We describe the employed hybrid long-range aerial dataset and outline the test scenarios designed to assess the contribution of these new constraints.

4.1 Dataset

The dataset employed in this work is publicly available in the EPFL-AddLidar repository[7] and was acquired using a fixed-wing aircraft, in collaboration with the Swiss Flight Services SA. The aircraft flew 500-1000 m above the mountain ridges and 1000-2500 m above the deepest areas in the valleys, in the canton of Valais, Switzerland. Elevation differences across the study area exceed 1500 m (Fig. 2).

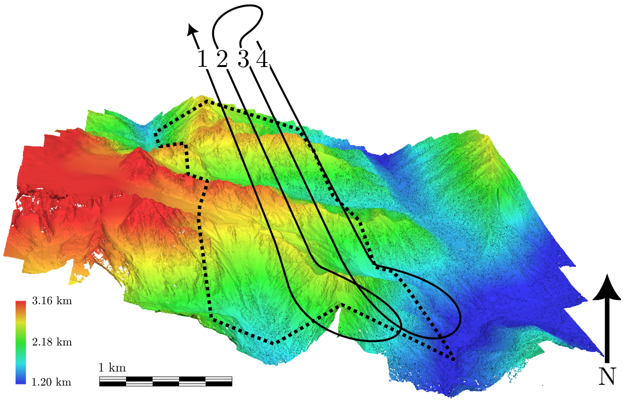

Figure 2. Terrain relief of the study area, colored by elevation. Imagery footprint and trajectory of the aircraft along the four flight lines used in this work are depicted in dashed and plain black, respectively.

Sensors: The hybrid sensor VQ-1560II (Riegl) (Fig. 3) was used for data capture, mounted on a stabilized platform (GSM-400) commanded by the real-time attitude output of the AV-610 (Trimble-Applanix) embedded INS/GNSS system, comprising the navigation-grade IMU-57 with 200 Hz sampling frequency and the Trimble Maxwell-6 GNSS mutli-frequency receiver. The GNSS antenna is placed on the aircraft fuselage, introducing a time-varying lever-arm ($\tilde{\boldsymbol{\alpha}}^b_{GNSS}$ in Fig. 3) in the body frame axes due to the stabilization motion, which is accounted for in processing. The optical sensors are the 150 megapixel IXU-1000RS (PhaseOne) camera with 50 mm lens and the dual

---

[1] Given the $[3 \times 3]$ attitude matrix $R^m_{b(t_i)}$ and the $[3 \times 1]$ position vector $T^m_{b(t_i)}$ of the *body* (b) frame in the *mapping* (m) frame at time t_i, $\widetilde{\Gamma}^m_{b(t_i)} = \begin{bmatrix} R^m_{b(t_i)} & T^m_{b(t_i)} \\ \mathbf{0}_{1\times 3} & 1 \end{bmatrix}$.

[2] K represents the $[3 \times 3]$ camera matrix. Given the principal distance c and the principal point $[ppx, ppy]$ expressed in pixels, $K = \begin{bmatrix} c & 0 & ppx \\ 0 & c & ppy \\ 0 & 0 & 1 \end{bmatrix}$.

[3] $\Xi(\cdot)$ models the lens distortions coefficients, that relate distorted and undistorted image coordinates, e.g. the Brown-Conrady distortion model (Brown, 1971). Given the undistorted image coordinates $[x_u, y_u]$ resulted from the projection function, the distorted image coordinates $[x_d, y_d]$ are determined as a function of the radial $(k_1, k_2, \ldots)$ and tangential $(p_1, p_2, \ldots)$ distortion coefficients as:

$$\Xi: \begin{cases} x_d = x_u\left(1 + k_1 \cdot r^2 + k_2 \cdot r^4\right) \\ \qquad + p_1\left(r^2 + 2 \cdot {x_u}^2\right) + 2 \cdot p_2 \cdot x_u \cdot y_u \\ y_d = y_u\left(1 + k_1 \cdot r^2 + k_2 \cdot r^4\right) \\ \qquad + p_2\left(r^2 + 2 \cdot {y_u}^2\right) + 2 \cdot p_1 \cdot x_u \cdot y_u \end{cases}$$

where $r^2 = {x_u}^2 + {y_u}^2$.

[4] For a given sensor s, i.e. camera (c) or lidar (l), mounting information that comprises the $[3 \times 3]$ boresight matrix R^b_s and the $[3 \times 1]$ lever-arm vector α^b_s of the sensor (s) in the *body* (b) frame are summarized in the matrix $\Gamma^b_s = \begin{bmatrix} R^b_s & -\alpha^b_s \\ \mathbf{0}_{1\times 3} & 1 \end{bmatrix}$.

[5] $\pi : \mathbb{R}^3 \rightarrow \mathbb{R}^2 : \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} \rightarrow \frac{1}{Z}\begin{bmatrix} X \\ Y \end{bmatrix}$

[6] $\widetilde{\Pi} = [\boldsymbol{I}_{3\times 3} \quad \mathbf{0}_{3\times 1}]$

[7] `https://addlidar.epfl.ch/viewer/0002_Val_dArpette`

https://doi.org/10.5194/isprs-annals-XI-1-2026-175-2026 | © Author(s) 2026. CC BY 4.0 License.

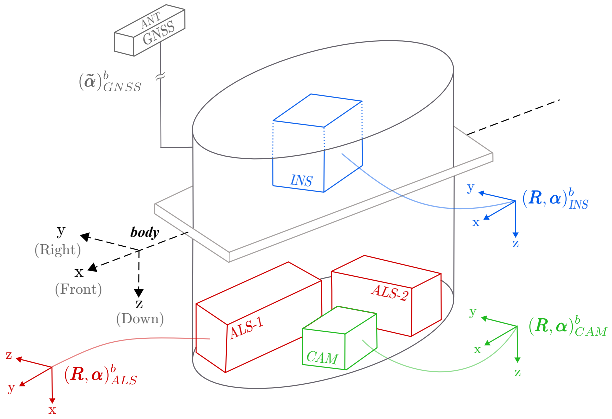

Figure 3. Approximate configuration of the navigation and optical sensors in the VQ-1560II hybrid sensor.

long-range lidar sensor (Riegl). In this work, observations from both laser units are utilized by transforming the laser vectors of the second scanner into the reference frame of the first using the known mounting parameters. Subsequently, only the primary sensor is introduced into the adjustment.

Calibrations: The lever-arm details and the lidar boresight calibration are provided by the flight service provider and the VQ-1560II (Riegl) calibration certificate. The camera IO parameters were determined nearly one year prior to the studied mission. The calibration flight was conducted at approximately half the flying altitude of the studied mission, which can affect the precision of the estimated parameters due to changes in pressure and temperature. At the same time, no information was provided for the camera boresight angles.

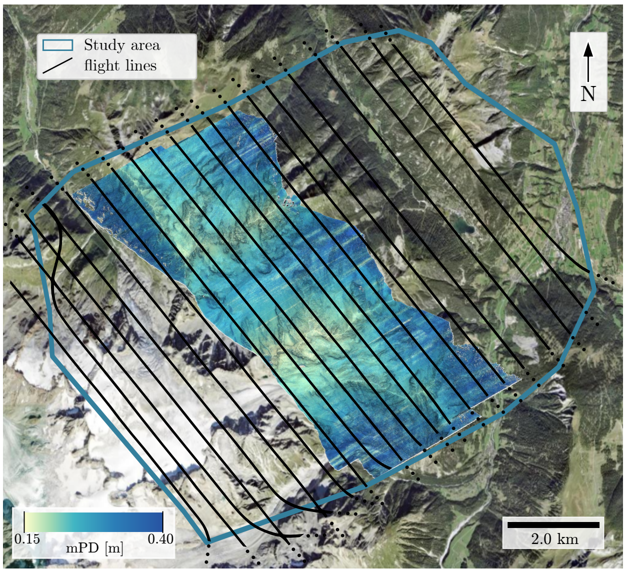

Figure 4. Overview of the study area with the 21 flight lines and the metric point density (mPD) of the analyzed region.

Mapping details: From all 21 overlapping flight lines of this mission (Fig. 4), this work focuses on 4 consecutive flight lines of the plane trajectory (Fig. 2), which span approximately 30 min. The flight lines have an overlap of 80-90% forward and 50-70% lateral overlap. There are 80 images in total with a GSD ranging from 6 to 20 cm, in the highest and lowest parts of the valley, respectively. The lidar sensor swath coincides with the image footprints on the ground since both sensors have a similar field of view (≈ 55°). For the lidar point cloud, we use the metric point density (mPD), which represents the average metric distance between a point and its closest neighbor. In this case, the mPD varies from 13 to 40 cm (which is roughly equivalent to points ground density of 60 to 6 pts/m^2), in the highest and lowest parts of the valley, respectively. These large variation of the imagery GSD and the lidar mPD result from the significant elevation changes across the steep terrain (Fig. 2). The mapped area consists of natural terrain, i.e. high and low vegetation, low boulders, rocks, and snow (or glacier) covered areas, with less than 10% built structures located in the lower valley. The four selected flight lines enclose all these terrain types and are therefore representative and sufficient to support the experiments.

4.2 Optimization cases

Based on the type of spatial constraints used together with the GNSS and raw inertial observations in DN, we study the following trajectory determination cases to inspect the impact and significance of the new type of cross-domain constraints:

- **Baseline:** Trajectory generated with the navigation grade IMU-57 which has high georeferencing accuracy with attitude noise smaller than 0.003°. Through that and by direct georeferencing, we obtain the baseline lidar point cloud using the formerly calibrated lidar boresight. We then use the trajectory in an AAT to obtain the baseline photogrammetric point cloud from the imagery. We use the whole block of 21 flight lines to determine less known parameters such as camera boresight or job-related variations in camera constant and perspective point
- **Case A - [IMU + GNSS +] 2D-2D:** DN computed trajectory following the proposed approach of integrating GNSS and raw inertial readings with single-domain 2D-2D constraints only.
- **Case B - [IMU + GNSS +] 2D-2D + 2D-3D:** DN computed trajectory following the proposed approach of integrating GNSS and raw inertial readings with single-domain optical constraints from the imagery, i.e. image-to-image (2D-2D), and cross-domain optical constraints, i.e. image-to-lidar (2D-3D).
- **Reference:** There are no external control features or areas determined independently with higher accuracy. The factory strip adjustment was performed using all flight-lines with insignificant (i.e. sub-mPD) changes on the georeferenced point cloud. Thus, we perform relative comparisons as analyzed in the following.

5. Results & Analysis

5.1 Quality of 3D-3D Extraction

It is first possible to assess the quality of the extracted 3D-3D correspondences by analyzing the pairwise correspondence distances. As shown in Fig. 5, the distance distributions across the four flight lines exhibit similar characteristics.

The average offset (i.e. mean values) provides a rough estimate of the initial mapping accuracy via the agreement of the lidar and photogrammetric point clouds georeferenced with the baseline solution. As indicated, there is a systematic bias

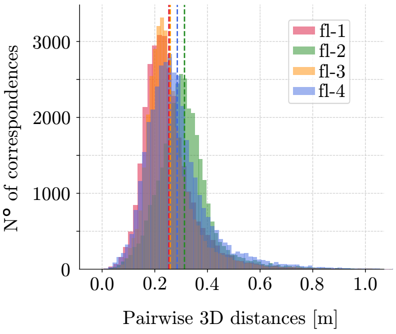

Figure 5. Pairwise distance distribution of the extracted 3D–3D features from the lidar and photogrammetric clouds, per flight line. Mean of each distribution is indicated with a vertical line.

between the two point clouds, of approximately 25-30 cm, which is 2-3 times larger than the mean imagery and lidar mPD values within these flight lines.

The width of the distributions (i.e. 1σ spread) reflects the internal accuracy of the extracted correspondences, as discussed in (Brun et al., 2025). This correspondence accuracy is around 10-15 cm, corresponding to roughly half of the lidar mPD, which is consistent with the expected behavior discussed in (Brun et al., 2025). This value is therefore adopted to define the *a-priori* uncertainty (weight) of the 2D–3D constraints in the subsequent DN adjustment, after conversion to the equivalent value in pixels, given the imagery GSD.

5.2 Quality of the Adjustment

5.2.1 Camera Calibration

The quality of the adjustment is first evaluated on the potential to self-calibrate the camera IO and boresight parameters when using the new type of cross-domain constraints (2D-3D) and without special calibration patterns. In order to account for potential adjustment instabilities, we evaluate different self-calibration scenarios of case studies A and B. In all cases, the parameters are ”relaxed” (free) for self-calibration are initialized with the same values, and the lens distortions are kept fixed. The calibration scenarios are:

- All fixed → both the IO and the boresight parameters stay fixed to the calibrated values provided
- All free → both the IO and the boresight parameters stay ”free” for self-calibration
- $[R^b_{CAM}]$ free → only the boresight parameters are set ”free” for self-calibration
- $[c + pp]$ free → only the IO parameters are set ”free” for self-calibration

Fig. 6 illustrates the residuals of the image-to-image (2D-2D) constraints for each case and calibration scenario, compared against their *a-priori* uncertainty in the adjustment. When both the IO and boresight parameters are set ”free”, the 2D residuals reach their lowest values, showing a first indication for improved adjustment consistency relative to the other cases. At the same time, when introducing the 2D-3D constraints, the residuals in the ”All free” scenario, nearly double. However, as they remain below the 3σ threshold, they are considered acceptable, suggesting that the overall consistency of the adjustment is preserved.

The magnitudes of residuals alone are not informative without examining the convergence values of the free parameters. Tab. 1 presents the estimated camera IO parameters, while Fig. 7 shows the estimated boresight parameters for each optimization case. The DN estimated values are compared with those from the AAT solution based on the baseline trajectory. We use as benchmark the AAT self-calibration solution, computed from all 21 flight lines of the dataset. The large relative altitude differences (steep topography) of the larger study area increase the confidence on the estimated camera parameters, especially the principal distance, due to the different scales present in the imagery.

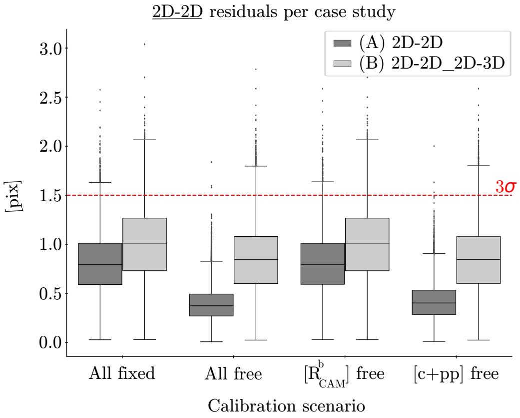

Figure 6. Adjusted residuals of the 2D-2D optical constraints after the DN adjustment, for each case study.

	Baseline	**Estimated Camera IO Parameters (fixed lens, 4fls)**		
	AAT (free lens, 21 fls)	**AAT**	**(A) DN 2D-2D**	**(B) DN 2D-2D_2D-3D**
c [pix]	11192.5	11166.4	11184.3	11187.0
pp_x [pix]	14.8	4.6	4.4	12.0
pp_y [pix]	-2.2	-1.3	-3.0	-4.2

Table 1. Estimated camera IO parameters for comparison with the factory calibration. In all optimization cases, the camera parameters are initialized with the same values.

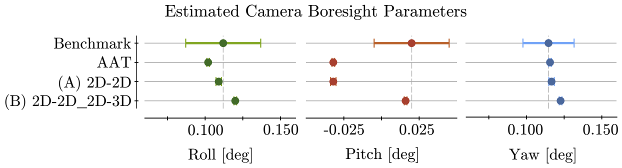

Figure 7. Estimated camera boresight parameters for comparison with the benchmark values. In all optimization cases, the components are initialized with the same values.

The use of 2D-3D constraints, between the imagery and the lidar point cloud (case B: DN_2D-2D_2D-3D) increases the adjustment consistency in scenarios with poor camera calibration information. This improvement is first evident in the estimated principal distance: while the AAT solution diverges by ≈ 25 pixels from the benchmark, the DN adjustment reduces this difference to 5-8 pixels. Similarly, the principal point, particularly its x component, deviates by about 10 pixels in the AAT solution but only by 2 pixels when cross-domain spatial constraints are incorporated in DN. Such deviations in the baseline (AAT) solution lead to a georeferencing error of ≈ 2 m in the photogrammetric point cloud. This trend is also reflected in the estimated boresight values: due to the correlation between the x

component of the principal point and the boresight pitch angle, the latter compensates for the former’s deviation, converging to a value beyond the 3σ threshold of the benchmark. In contrast, the inclusion of 2D–3D constraints effectively stabilizes the camera parameters and preserves internal consistency.

5.2.2 Check constraints

Next, we evaluate the adjustment quality based on the behavior of the remaining optical constraints: the lidar-to-lidar (3D-3D) correspondences, extracted between overlapping lidar point cloud sections, and the newly introduced image-to-lidar (2D-3D) correspondences. We characterize them as ”check”, since they are used in the adjustment with large uncertainties, i.e. $1\sigma = 12$ m for the 3D–3D (10× the lidar mPD), and $1\sigma = 100$ pixels for the 2D–3D (10× their uncertainty defined in 5.1). In that way, they do not influence the adjustment, but we can report their residuals as relative quality indicators of the internal consistency between observations.

Fig. 8 shows the residual distributions of such check-constraints (dashed lines), for the self-calibration scenario ”All free”, including the lens distortions. The left plot reveals a remaining disagreement between overlapping lidar point clouds, indicating that using 3D–3D constraints in the adjustment could help compensate for these inconsistencies within the lidar data. The right plot highlights the contribution of the 2D-3D constraints. When excluded from the adjustment (case A: DN_2D-2D), their residuals exceed the 3σ threshold. On the contrary, their inclusion (case B: DN_2D-2D_2D-3D), reduces the residuals below the observation level (1σ), without increasing other residuals. Combined with the acceptably estimated camera IO and boresight parameters from the previous analysis, this confirms the enhanced internal consistency of the adjustment and the positive contribution of the cross-domain constraints.

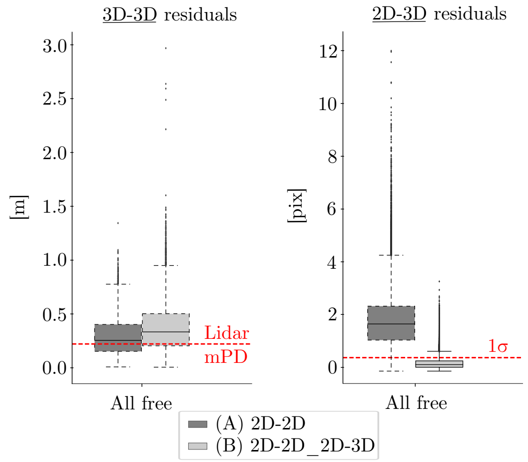

Figure 8. Adjusted residuals of the 3D-3D (left) and 2D-3D (right) correspondences used as ”check constraints” in the DN adjustment, for each optimization case. The dashed lines indicate the use of these constraints as check-constraints.

5.3 Relative Cross-Sensor Accuracy

The contribution of the 2D-3D constraints becomes even more valuable when it is reflected in the georeferenced lidar and photogrammetric point clouds. Since we have no absolute reference point cloud of higher accuracy, we assess the relative agreement of two types of point clouds (imagery versus lidar), using the M3C2 tool of Cloud-Compare software for a cloud-to-cloud comparison (Lague et al., 2013).

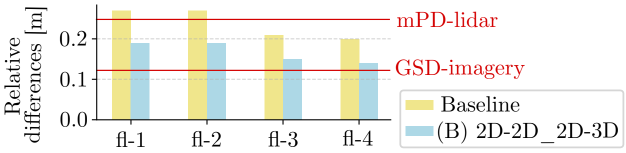

Figure 9. Mean difference per flight line between the photogrammetric and lidar clouds for baseline and (B) cases.

After applying the M3C2 tool, in Fig. 9, we show per flight line the mean difference between the lidar and photogrammetric point clouds when both are georeferenced with the baseline solution (all lines, separate process) or the DN_2D-2D_2D-3D (case B) trajectory and respective estimated parameters. The results are compared with the GSD and mPD of the optical data for a fair comparison. As observed, for all flight lines, a constant improvement of ≈ 30% is achieved on the relative agreement of the lidar and photogrammetric point clouds with respect to the baseline, when introducing the 2D-3D constraints. This confirms our initial investigations with emulated correspondences (Mouzakidou et al., 2025) that brought an improvement of 30-50% on absolute point cloud georeferencing.

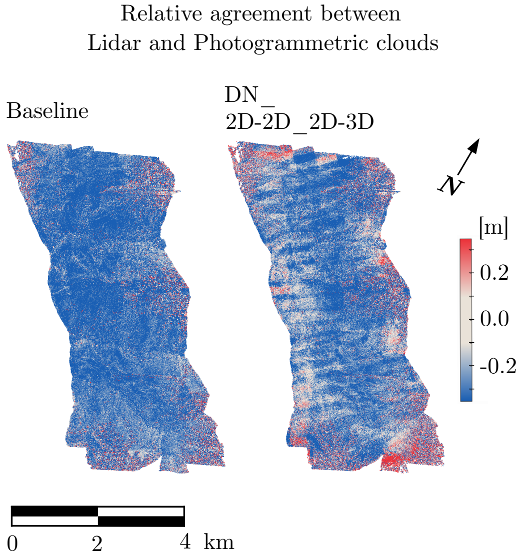

Figure 10. Relative discrepancy between the lidar and photogrammetric clouds when georeferenced using the baseline or the DN computed trajectory with the new type of 2D-3D constraints (case B), for the extent of the 4 flight lines.

The spatial distribution of the relative differences across the study area is shown in Fig. 10. When the point clouds are georeferenced using the baseline solution, a nearly constant displacement (bias) of approximately 30 cm is observed over the whole area. This is reduced to below lidar mPD level (15-20 cm) when considering the cross-domain constraints in the adjustment. Nevertheless, some residual bias remains, as also reflected in the distribution of relative differences for flight lines 1 and 4 in Fig. 11 that represent the borders. This bias is within the magnitude of the camera GSD and lidar mPD and may well disappear if other flight lines are included. Nevertheless, further analysis using the lidar-to-lidar constraints is required to assess the potential for additional improvement when jointly integrating all available optical constraints with GNSS and raw inertial

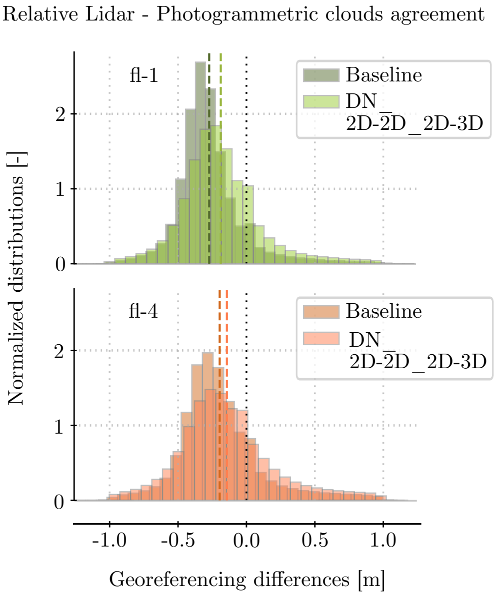

Figure 11. Distribution of relative discrepancy between the geo-referenced lidar and dense image-matching clouds for the first and last flight line (borders) of the block.

observations in the DN adjustment.

6. Conclusions & Outlook

In this work, we proposed an automatic extraction of image-to-lidar (2D-3D) correspondences established in the 3D domain (via same descriptors as in (Brun et al., 2025)) and evaluated their impact when integrated with GNSS and raw inertial readings in a single dynamic network adjustment. We showed that:

1. 2D-3D correspondences are reliably extracted between the lidar and photogrammetric point clouds, with accuracy consistent with the dataset GSD and mPD (in a natural rough-terrain limited-trees scenario),
2. 3D–3D correspondences, derived using the method of (Brun et al., 2025), provide a robust measure of relative accuracy between imagery and lidar point clouds,
3. the use of 2D-3D constraints enables camera calibration, both internal and boresight parameters, in a mapping pattern of few parallel lines, which is a weak-geometry configuration where the conventional (AAT) approach fails,
4. in this specific mapping scenario, the use of 3D-3D constraints between the lidar point clouds could further mitigate the residual inconsistencies observed within the lidar data using the two step factory approach (direct georeferencing followed by ”strip” adjustment),
5. when cross-domain spatial constraints are considered, the relative agreement of lidar and photogrammetric point clouds is improved by 30% compared to those produced by conventional separate approaches.

These investigations confirm our initial findings with emulated data (Mouzakidou et al., 2025), opening promising directions for cross-domain (image-to-lidar) matching followed by common optimization. More insights are expected from future more detailed studies, benefiting from additional datasets, possibly different systems, varying scales, textures, and flying configurations, as well as from refinements in the matching workflow.

7. Acknowledgments

The flight and data were provided by the Swiss Flight Services. This contribution is partially supported by the Swiss Innovation Agency project 119.293 IP-ENG. The dissemination of open data in this research was supported by the Open Research Data Program of the ETH Board.

References

Brown, D., 1971. Close-range camera calibration. https://api.semanticscholar.org/CorpusID:56490170.

Brun, A., 2021. LiDAR correspondences for Integrated Sensor Orientation. https://infoscience.epfl.ch/handle/20.500.14299/189153.

Brun, A., Kolecki, J., Xiao, M., Insolia, L., van der Zwan, E. V., Guerrier, S., Skaloud, J., 2025. Generalization of point-to-point matching for rigorous optimization in kinematic laser scanning. *ISPRS Journal of Photogr. and Rem. Sens.*, 229, 107-121.

Charles, R. Q., Su, H., Kaichun, M., Guibas, L. J., 2017. PointNet: Deep Learning on Point Sets for 3D Classification and Segmentation. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 77–85.

Chen, L., Franz, R., Heipke, C., 2021. Feature detection and description for image matching: from hand-crafted design to deep learning. *Geo-spatial Information Science*, 24(1), 58–74.

Cheng, P., Yu, H., Liu, C., Luo, K., Akhtar, N., Chen, X., 2025. RID-Net: A Hybrid MLP-Transformer Network for Robust Point Cloud Registration. *IEEE Robotics and Automation Letters*, 10(5), 5066–5073.

Colomina, I., Blázquez, M., 2004. A unified approach to static and dynamic modeling in Photogr. and Rem. Sens. *ISPRS Archives of Photogr., Rem. Sens. and Spatial Information Sciences*, 35-B1, 178-183.

Cucci, D. A., Rehak, M., Skaloud, J., 2017. Bundle adjustment with raw inertial observations in UAV applications. *ISPRS Journal of Photogr. Engineering and Rem. Sens.*, 130, 1-12.

Deng, H., Birdal, T., Ilic, S., 2018. PPFNet: Global Context Aware Local Features for Robust 3D Point Matching. *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 195–205.

Farella, E. M., Morelli, L., Remondino, F., Mills, J. P., Haala, N., Crompvoets, J., 2022. THE EUROSDR TIME BENCHMARK FOR HISTORICAL AERIAL IMAGES. *Int. Arch. Photogramm. Rem. Sens. Spatial Inf. Sci.*, XLIII-B2-2022, 1175–1182.

Förstner, W., Wrobel, B. P., 2016. *Photogrammetric Computer Vision: Statistics, Geometry, Orientation and Reconstruction*. Geometry and Computing, Springer Int. Publishing.

Glira, F., Pfeifer, N., Mandelburger, G., 2019. Hybrid orientation of airborne lidard point clouds and aerial images. *The Int. Archives of the Photogr., Rem. Sens. and Spatial Information Sciences*, IV-2/W5, 567-574.

Gojcic, Z., Zhou, C., Wegner, J. D., Wieser, A., 2019. The perfect match: 3D point cloud matching with smoothed densities. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2019-June, 5540–5549.

Guo, Y., Bennamoun, M., Sohel, F., Lu, M., Wan, J., Kwok, N. M., 2016. A Comprehensive Performance Evaluation of 3D Local Feature Descriptors. *Int. Journal of Computer Vision*, 116(1), 66–89.

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 Photogr. and Rem. Sens.*, 4, 100014.

Hartmann, W., Havlena, M., Schindler, K., 2016. Recent developments in large-scale tie-point matching. *ISPRS Journal of Photogr. and Rem. Sens.*, 115, 47–62.

Hasheminasab, S. M., Zhou, T., Habib, A., 2023. Linear Feature-Based Image/LiDAR Integration for a Stockpile Monitoring and Reporting Technology. *IEEE Journal of Selected Topics in Applied Earth Observations and Rem. Sens.*, 16, 2605–2623.

Hussnain, Z., Elberink, S. O., Vosselman, G., 2019. Automatic extraction of accurate 3D tie points for trajectory adjustment of mobile laser scanners using aerial imagery. *ISPRS Journal of Photogr. and Rem. Sens.*, 154, 41-58.

Hussnain, Z., Elberink, S. O., Vosselman, G., 2021. Enhanced trajectory estimation of mobile laser scanners using aerial images. *ISPRS Journal of Photogr. and Rem. Sens.*, 173, 66-78.

Jonassen, V. O., Kjørsvik, N. S., Gjevestad, J. G. O., 2023. Scalable hybrid adjustment of images and LiDAR point clouds. *ISPRS Journal of Photogr. and Rem. Sens.*, 202, 652–662.

Kharroubi, A., Remondino, F., Ballouch, Z., Hajji, R., Billen, R., 2025. Semantic and Geometric Fusion for Object-Based 3D Change Detection in LiDAR Point Clouds. *Rem. Sens.*, 17(7), 1311.

Lague, D., Brodu, N., Leroux, J., 2013. Accurate 3D comparison of complex topography with terrestrial laser scanner: Application to the Rangitikei canyon (N-Z). *ISPRS Journal of Photogr. and Rem. Sens.*, 82, 10–26.

Lee, J.-H., Biging, G. S., Fisher, J. B., 2016. An Individual Tree-Based Automated Registration of Aerial Images to LiDAR Data in a Forested Area. *Photogrammetric Engineering & Rem. Sens.*, 82(9), 699–710.

Li, J., Hu, Q., Ai, M., 2020. RIFT: Multi-Modal Image Matching Based on Radiation-Variation Insensitive Feature Transform. *IEEE Trans. on Image Processing*, 29, 3296–3310.

Lindenberger, P., Sarlin, P.-E., Pollefeys, M., 2023. LightGlue: Local Feature Matching at Light Speed. *2023 IEEE Int. Conference on Computer Vision (ICCV)*, 17581–17592.

Lowe, D. G., 2004. Distinctive Image Features from Scale-Invariant Keypoints. *Int. Journal of Computer Vision*, 60(2), 91–110.

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 Photogr., Rem. Sens. and Spatial Information Sciences*, IV-2/W4, 259–266.

Morelli, L., Ioli, F., Maiwald, F., Mazzacca, G., Menna, F., Remondino, F., 2024a. DEEP-IMAGE-MATCHING: A TOOLBOX FOR MULTIVIEW IMAGE MATCHING OF COMPLEX SCENARIOS. *Int. Arch. Photogramm. Rem. Sens. Spatial Inf. Sci.*, XLVIII-2-W4-2024, 309–316.

Morelli, L., Perda, G., Ioli, F., Trybała, P., Sterpin, A., Rigon, S., Sutherland, N., Medici, M., Remondino, F., Vitti, A., 2024b. Co-registering Laser Scanning Point Clouds and Photogrammetric Images with Deep Learning Multi-Modal Matching. *Int. Arch. Photogramm. Rem. Sens. Spatial Inf. Sci.*, XLVIII-2-W8-2024, 335–342.

Mouzakidou, K., Brun, A., Cucci, D. A., Skaloud, J., 2024. Airborne sensor fusion: Expected accuracy and behavior of a concurrent adjustment. *ISPRS Open Journal of Photogr. and Rem. Sens.*, 12, 100057.

Mouzakidou, K., Stoltz, T., Jospin, L. V., Cucci, D. A., Skaloud, J., 2025. On the Perspectives of Image-to-Lidar Constraints in Dynamic Network Optimisation. *Int. Arch. Photogramm. Rem. Sens. Spatial Inf. Sci.*, XLVIII-M-6-2025, 213–220.

Pentek, Q., Kennel, P., Allouis, T., Fiorio, C., Strauss, O., 2020. A flexible targetless LiDAR–GNSS/INS–camera calibration method for UAV platforms. *ISPRS Journal of Photogr. and Rem. Sens.*, 166, 294–307.

Remondino, F., Morelli, L., Stathopoulou, E., Elhashash, M., Qin, R., 2022. AERIAL TRIANGULATION WITH LEARNING-BASED TIE POINTS. *The Int. Archives of the Photogr., Rem. Sens. and Spatial Information Sciences*, XLIII-B2-2022, 77–84.

Rouzaud, D., Skaloud, J., 2011. Rigorous Integration of Inertial Navigation with Optical Sensors by Dynamic Networks. *NAVIGATION*, 58(2), 141–152.

Rusu, R. B., Blodow, N., Beetz, M., 2009. Fast Point Feature Histograms (FPFH) for 3D registration. *2009 IEEE Int. Conference on Robotics and Automation*, 3212–3217. ISSN: 1050-4729.

Salti, S., Tombari, F., Di Stefano, L., 2014. SHOT: Unique signatures of histograms for surface and texture description. *Computer Vision and Image Understanding*, 125, 251–264.

Sarlin, P.-E., DeTone, D., Malisiewicz, T., Rabinovich, A., 2020. SuperGlue: Learning Feature Matching With Graph Neural Networks. *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, Seattle, WA, USA, 4937–4946.

Skaloud, J., Schaer, P., Stebler, Y., Tome, P., 2010. Real-time registration of airborne laser data with sub-decimeter accuracy. *ISPRS Journal of Photogr. Engineering and Rem. Sens.*, 65(2), 208-217.

Spezialetti, R., Salti, S., Di Stefano, L., Tombari, F., 2020. 3D Local Descriptors—from Handcrafted to Learned. Y. Liu, N. Pears, P. L. Rosin, P. Huber (eds), *3D Imaging, Analysis and Applications*, Springer Int. Publishing, Cham, 319–352.

Verdie, Y., Yi, K. M., Fua, P., Lepetit, V., 2015. TILDE: A Temporally Invariant Learned DEtector. *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 5279–5288.

Wang, C., Gu, Y., Li, X., 2025. LPRnet: A Self-Supervised Registration Network for LiDAR and Photogrammetric Point Clouds. *IEEE Trans. on Geoscience and Rem. Sens.*, 63, 1–12.

Wang, X., Xu, B., Zhao, S., Li, X., 2022. Automatic registration of optical image and airborne LiDAR data based on centers and corner points of building boundaries. *Int. Journal of Rem. Sens.*, 43(18), 6644–6668.

Yi, K. M., Verdie, Y., Fua, P., Lepetit, V., 2016. Learning to Assign Orientations to Feature Points. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 107–116.

Zhang, L., Rupnik, E., Pierrot-Deseilligny, M., 2021. Feature matching for multi-epoch historical aerial images. *ISPRS Journal of Photogr. and Rem. Sens.*, 182, 176–189.

Zhou, T., Hasheminasab, S. M., Habib, A., 2021. Tightly-coupled camera/LiDAR integration for point cloud generation from GNSS/INS-assisted UAV mapping systems. *ISPRS Journal of Photogr. and Rem. Sens.*, 180, 336–356.

Zhu, B., Ye, Y., Zhou, L., Li, Z., Yin, G., 2021. Robust registration of aerial images and LiDAR data using spatial constraints and Gabor structural features. *ISPRS Journal of Photogr. and Rem. Sens.*, 181, 129–147.