

# ON THE BENEFIT OF CONCURRENT ADJUSTMENT OF ACTIVE AND PASSIVE OPTICAL SENSORS WITH GNSS & RAW INERTIAL DATA

K. Mouzakidou<sup>1\*</sup>, D. A. Cucci<sup>2</sup>, J. Skaloud<sup>1</sup>

<sup>1</sup> Geodetic Engineering Laboratory (TOPO), École Polytechnique Fédérale de Lausanne (EPFL), Switzerland  
(kyriaki.mouzakidou, jan.skaloud)@epfl.ch

<sup>2</sup> Geneva School of Economics and Management, University of Geneva, Switzerland  
davide.cucci@unige.ch

Commission I, WG I/9

**KEY WORDS:** Lidar, Geo-referencing, Photogrammetry, Sensor-fusion, UAVs

## ABSTRACT:

In airborne laser scanning a high-frequency trajectory solution is typically determined from inertial sensors and employed to directly geo-reference the acquired laser points. When low-cost MEMS inertial sensors are used, such as in lightweight unmanned aerial vehicles, non-negligible errors in the estimated trajectory project to the final point-cloud, resulting in unsatisfactory accuracy on the ground. There are different multi-sensor fusion approaches to correct the point-cloud errors caused by an imperfect trajectory determination. Mismatches between different optical observations and/or in the overlapping regions of the point-cloud can allow the correction of the final point-cloud, either directly, by means of rigid transformations, or indirectly, via improving the scanner trajectory estimation. In this work we propose to fuse lidar and cameras in a single adjustment based on dynamic networks, considering 2D tie-points from the imagery and 3D tie-points from overlapping point-cloud sections. On a challenging corridor mapping scenario, we show that considering either 2D or 3D tie-points, along with inertial and GNSS observations, results in a remarkably accurate point-cloud, even when low-cost inertial sensors are employed and in presence of challenging surface textures, such as high vegetation. Furthermore, since the distribution of the 2D and 3D tie-points is complementary, considering them together further increases the robustness of the adjustment due to higher redundancy. By employing the proposed approach within this controlled example, we were able to improve the final point-cloud accuracy by more than three times with respect to conventional geo-referencing methodology and to reduce the magnitude of the errors.

## 1. INTRODUCTION

Airborne Laser Scanning (ALS) is one of the most important geospatial data acquisition technologies, providing precise geometric products with reflectivity information, i.e., intensity values, and is currently used in several disciplines such as infrastructure monitoring, archaeology, agriculture and forestry. The constant effort to miniaturize Light Detection And Ranging (lidar) units reflects the need to use them in smaller, more flexible, lightweight and low-cost mapping platforms. When it comes to aerial mapping, Unmanned Aerial Vehicles (UAVs) have become of significant importance since they allow for prompt and flexible mobilization (Nex et al., 2022), while lidar units offer high spatial resolution and vegetation penetration due to their ability to detect multiple pulse echoes.

In kinematic laser scanning, due to the sequential measurement principle and the motion of the carrier platform, the exterior orientation parameters (position and attitude) of the scanner differ for every object point, or for some scanners, a line of sparse points. Hence, unlike in mapping with frame-based imagery, mobile lidar scanners rely on direct geo-referencing (Glenie, 2007) and therefore require appropriate GNSS and inertial hardware for trajectory determination. Restrictions on the weight and size (as well as the cost) of the payload on UAVs limit the accuracy of the navigation and optical sensors that can be employed. Consequently, the accuracy of laser scanning products depends heavily on the quality of the navigation

sensors used for the trajectory determination, and fails to meet the user expectations in certain cases (Schaer et al., 2009).

With modern high-end GNSS board-receivers occupying the size of a credit-card at the most, the trajectory limitations are mainly on the side of inertial sensors. Indeed, the residual attitude errors due to inertial sensor noise, and especially due to imperfections of platform initial orientation (so called “system misalignment errors” in navigation-terminology) are the predominant factors of trajectory quality. Additional systematic errors, or their projection, e.g., caused by an imperfect boresight between the lidar and the Inertial Measurement Unit (IMU), deteriorate further the final point-cloud accuracy. Plane-to-plane correspondences in overlapping parts of the point-cloud can correct and recover boresight and/or system misalignments and provide a refined point-cloud as shown in (Kager, 2004) and (Skaloud, Lichti, 2006). However, such approaches may fail due to under-fitting if planes of different slope and orientation are not present in the mapping area. Although such plane-to-plane constraints can be generalized when matching certain terrain features or objects (Kerstling et al., 2012), these conditions are usually weaker and thus the parameter recovery less precise.

Many approaches (apart from system calibration) have been proposed to correct errors of the final 3D point-clouds caused by an imperfect determination of the scanner trajectory. These can be divided into the following categories: i) approaches that act directly on the final 3D point-cloud (e.g., via rigid transformations between different sections), and ii) approaches that attempt to directly correct the scanner trajectory. All these tech-

\*Corresponding author

niques typically rely on mismatches between the measurements of different optical sensors and/or in the overlapping regions of the point-cloud. Further details are given in the sections below.

### 1.1 Point-cloud alignment methods

Here, we discuss methods that attempt to correct the point-cloud after it has been geo-referenced, by exploiting overlapping scans, based on rigid transformations that are able to globally correct the point-cloud, e.g., a shift and rotation of entire strips. As an example we refer to (Glira et al., 2015) where the ICP algorithm is applied in overlapping regions to improve point-cloud alignment. However, this approach is sensitive to poor approximation of the initial registration, such as those possibly caused by low-accuracy inertial units on UAVs. Given a coarsely geo-referenced point-cloud, extra information from image tie-points can improve the point-cloud geo-referencing accuracy in a hybrid adjustment as shown, e.g., in (Glira et al., 2019). When imagery information is considered, obtaining sufficient constraints on the point-cloud is highly dependent on the employed methodology, as described in (Cledat, Skaloud, 2020). Rasterizing and colorizing the processed lidar point-cloud using the elevation (Jayendra-Lakshman, Devarajan, 2013) or the intensity values (Hussnain et al., 2019) and performing 2D feature matching between overlapping regions can improve the point-cloud geo-referencing. Such approaches result in a certain degree of destruction of the original point-cloud information since 2D feature extraction is applied on the raster data and the point-cloud is treated solely in the 2D domain. Ongoing research in neural networks has shown that end-to-end *learned* approaches can improve the point-cloud registration accuracy. 3DSmoothNet (Gojcic et al., 2019), SuperGlue (Sarlin et al., 2019) and MDGAT-matcher (Shi et al., 2021) are some examples based on local feature matching, but to date these methods have not been evaluated for ALS.

### 1.2 Trajectory correction methods

At a trajectory level, (Glira et al., 2016) propose a spline trajectory correction model for a rigorous strip adjustment that handles the trajectory errors after the Kalman filtering of the INS/GNSS integration. (Mandlbürger et al., 2017) employ dense image matching to link image tie-points and lidar point-clouds in a hybrid strip adjustment, still relying on direct geo-referencing, separately for the lidar and imagery observations. (Li et al., 2019) exploit the advantage of a graph-based optimization algorithm to obtain an improved high-frequency trajectory via image matching and raw inertial observations, following (Cucci et al., 2017). Using the optimized image poses as reference, the authors register the lidar point-cloud by minimizing the depth discrepancy between the depth maps obtained from image matching and the raw laser scans. In this two-step methodology, lidar observations are considered only at the final stage. (Hussnain et al., 2021) propose a B-spline based 6 degrees-of-freedom trajectory adjustment with IMU observations and links between a terrestrial lidar and a reference (aerial) photogrammetric point-cloud. All these methods rely on multiple optical sensors and all related calibration parameters must be known accurately *a-priori*.

Recently, a novel approach has been proposed in (Brun et al., 2022) where the trajectory to be used in the final point-cloud geo-referencing is estimated considering raw inertial, GNSS and lidar observations in a single adjustment step. First, a coarse trajectory, obtained fusing raw inertial and GNSS observations only, is used to geo-reference the initial point-cloud.

Next, 3D matches, or tie-points are established within overlapping regions of such point-cloud by means of an automated 3D feature extraction and matching algorithm (Pham et al., 2019). These three-dimensional matches are then introduced as additional observations in a Dynamic Network (DN) adjustment (Cucci, Matteucci, 2014) along with GNSS positions and raw inertial data, namely specific forces and angular rates. This approach has been shown to drastically improve the final point-cloud registration. Further details are given in Section 2.

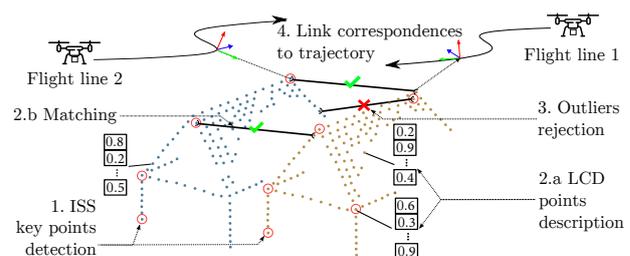
### 1.3 Contributions

In this work we extend the approach put forward in (Brun et al., 2022) by considering conventional tie-points, extracted from camera images acquired simultaneously with respect to the lidar points, together with 3D tie-points and raw inertial and GNSS observations in a single DN adjustment. We evaluate a challenging corridor mapping scenario observed simultaneously with industrial-grade and navigation-grade sensors, the latter providing the reference. We study the effect of passive (cameras) and active (lidar) sensor observations first separately and then jointly. The availability of a ground truth point-cloud with centimeter level accuracy allows us to precisely quantify the geo-referencing error in object space. The details on reference accuracy verifications are presented in (Vallet et al., 2020).

This work is organized as follows. In Section 2, we briefly review the approach put forward in (Brun et al., 2022) for establishing 3D tie-points between overlapping point-clouds. In Section 3, we describe the way these matching points are introduced in the DN adjustment, along with photogrammetric observations, and we stress the single-step integration of all the available information for trajectory determination. In Section 4, we present the design of experiments and in Section 5 the obtained results, before drawing the conclusions in Section 6.

## 2. 3D TIE-POINTS

This section summarizes the approach in (Brun et al., 2022). Apart from the rigorous formulation via DN in a global frame that builds on (Cucci et al., 2017), the approach put forward by the authors is based on the automatic extraction of matching three-dimensional features in overlapping regions of the point-cloud. This approach is comparable to the two-dimensional feature processing of any photogrammetric software that applies computer vision algorithms to detect 2D features on images and extract 2D tie-points based on feature matching. Similarly, in (Brun et al., 2022) 3D tie-point extraction was implemented to identify matching features to be used as additional information in the DN adjustment, as will be detailed in Section 3.



**Figure 1.** Steps of 3D tie-point extraction: (1) feature detection, (2a) description, (2b) matching and (3) filtering for point-cloud to point-cloud registration, after (Brun et al., 2022) .

The 3D tie-point extraction consists of four main steps, depicted in Fig. 1: detection, description, matching and outlier rejection of local features, that are then linked to the trajectory. In three-dimensional space, features correspond to spheroids and feature matching actually relates the centres of such spheroids. Automated 3D feature extraction has been studied extensively and there exist multiple 3D feature descriptors, handcrafted, e.g., Intrinsic Shape Signature (ISS) (Zhong, 2009) and Signature of Histograms of Orientations (SHOT) (Tombari et al., 2010), and learned ones, e.g., SpinNet (Ao et al., 2021) and LCD (Pham et al., 2019). We refer the reader to (Brun et al., 2022) for a comparison of many methods employed with ALS.

## 2.1 Point-cloud preparation

An initial rough trajectory, estimated via a Kalman Filter (KF) followed by Recursive Smoother (RS), is required to coarsely geo-reference the point-cloud and thus identify overlapping areas. The overlapping regions are then split into rectangular tiles and automatic feature extraction is performed on each tile separately. The necessary steps are discussed in the following.

## 2.2 3D feature detection

This step, depicted in Fig. 1 (point 1), is implemented using the *handcrafted* ISS algorithm (Zhong, 2009), which was originally conceived as a descriptor. (Brun et al., 2022) employed only the first parts of ISS, which correspond to keypoint detection. Thus, keypoint candidates for each point-cloud are selected based on the eigenvalues of the weighted covariance matrices that are computed for the spherical neighbourhood of each point.

## 2.3 3D feature description

This step, shown in Fig. 1 (point 2a) is based on a *learned* approach, the recently proposed Learned Cross-Domain descriptor (LCD) algorithm (Pham et al., 2019). As showcased in (Brun et al., 2022), LCD has better performance when compared with other competing approaches, at least in the case of ALS over a mixed natural-built environment. By encoding the geometric structure of spherical neighbourhoods for each keypoint, LCD outputs feature descriptors so that similar geometries have similar descriptors.

## 2.4 3D feature matching

In this step, depicted in Fig. 1 (point 2b), keypoints from the first cloud are compared with all points in the overlapping one, as opposed to the conventional keypoint matching workflows where only keypoints of the two clouds are considered. Considering all points in the query point-cloud helps to compensate for potentially weak detector repeatability (Salti et al., 2012). The matching is based on searching, for every keypoint in the first point-cloud, its nearest neighbour point in the overlapping one in terms of L2-norm of the difference of the feature vectors.

## 2.5 Outlier rejection

Due to the comparison of keypoints from one tile with all the points of another tile, the matching process is prone to a large number of outliers, i.e., incorrect matching points. To eliminate false matches, as shown in Fig. 1 (point 3), a variant of the Random Sample Consensus (RANSAC) algorithm (Fischler, Bolles, 1981) is used, relying on the assumption that errors in the navigation solution yield similar geo-referencing errors for the entire tile. This assumption is reasonable provided

that the tiles are small enough, i.e., were acquired within a time interval of a few seconds.

The final output of this four-step feature extraction workflow is a set of 3D homologous points in the two overlapping point-clouds. For each matching pair of points, the surroundings of the two different points (in each respective point-cloud) have similar structure in the object space, and thus probably identify the same point on the ground. In this way, couples of raw lidar measurements are created, for which geometric constraints can be formulated (as is discussed in Section 3), similar to the use of epipolar constraints for image coordinates in photogrammetric workflow.

## 3. MULTI SENSOR INFORMATION FUSION IN DYNAMIC NETWORKS

Dynamic Networks (DNs), first introduced in (Colomina, Blazquez, 2004), are an extension of conventional geodetic networks and have many applications in multi-sensor information fusion for trajectory determination, navigation and photogrammetry. In DNs, the unknowns are samples of the platform position and orientation, 3D coordinates of points in object space (such as tie-points) and optionally system calibration parameters, such as boresights, lens distortion coefficients, etc. Each raw sensor observation, such as the image coordinates of a tie-point in a given image, or the GNSS position observation at a certain time, forms a constraint between one or more unknowns. Minimizing the squared error associated with each constraint, weighted by the uncertainty of the sensor measurements, allows the determination of the maximum-likelihood estimate for all unknowns. An in-depth discussion of the approach is beyond the scope of this work, for which we refer the reader to the original publications, and to the ones referenced below.

In this contribution, we employ DNs to determine the trajectory of an airborne mapping system by fusing the following types of sensor measurements:

1. raw specific forces and angular velocities as measured by an IMU, the observation models being those presented in (Cucci, Skaloud, 2019),
2. GNSS positions,
3. (optional) image coordinates of 2D tie-points, as proposed in (Cucci et al., 2017),
4. (optional) raw laser ranges (vectors in lidar sensor frames) corresponding to 3D tie-points, as proposed in (Brun et al., 2022).

As an example, a 3D tie-point allows us to formulate the following constraint on the trajectory:

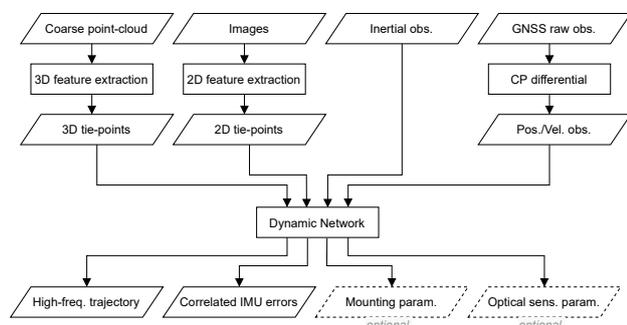
$$\Gamma_{b,t_1}^n \Gamma_L^b v_1^L = \Gamma_{b,t_2}^n \Gamma_L^b v_2^L + \xi, \quad (1)$$

where  $\Gamma_{b,t_1}^n$  and  $\Gamma_{b,t_2}^n$  are the transformation matrices encoding the position and orientation of the body frame  $b$  with respect to the global or navigation frame  $n$  at times  $t_1$  and  $t_2$ ,  $\Gamma_L^b$  encodes the lidar boresight and lever-arm and  $v_1^L$  and  $v_2^L$  are the two *raw* lidar measurements (points in lidar frame  $L$ ) that have been matched by the algorithm presented in the previous section.  $\xi$  is a zero-mean Gaussian error which identifies the uncertainty associated with the constraint and is related to the Ground Sampling Distance (GSD) of the point-cloud. For more

details please see (Brun et al., 2022). More informally, Equation 1 states that if two points in the laser frame are matched to form a 3D tie-point, the same coordinates should be obtained once  $v_1^L$  and  $v_2^L$  are geo-referenced (translated and rotated to the global frame  $n$ ) using the corresponding samples of the body frame position and orientation.

The advantage of DNs lies in their flexibility and generality: many types of constraints such as the one presented in eq. 1 can be formulated to fuse all information from multiple heterogeneous and partly redundant sensors in a single adjustment step (Fig. 2). This has several advantages, such as rigorous sensor modeling, consistent uncertainty quantification, better observability and easier de-correlation of calibration parameters.

In the following we will demonstrate how DNs allow the exploitation of different types of optical sensors, i.e., lidar sensors and cameras, along with inertial and positioning sensors, to obtain a high-frequency trajectory suitable for accurate geo-referencing.



**Figure 2.** Integration of all the available sensor information in a single step. The 3D tie-point extraction is performed on a point-cloud that has previously been geo-referenced by means of an INS/GNSS trajectory obtained via recursive smoothing.

#### 4. EXPERIMENTAL EVALUATION

In this section, we present the design of experiments to evaluate the concurrent adjustment of passive and active sensors. We describe the dataset used and its adequacy for the specific evaluation, as well as the different scenarios tested to investigate the significance of each set of optical observations.

##### 4.1 Data



**Figure 3.** Overview of the study area and the image footprints for the flight lines (FL) studied. Both urban and rural areas are enclosed in the mapping area.

The experimental evaluation of this work is performed on the controlled airborne dataset presented in (Vallet et al., 2020).

This dataset includes measurements from optical and navigation sensors of high and low accuracy in terms of orientation and mapping performance. It was gathered over different terrain types to include many possible mapping features in urban and rural areas, such as low and high vegetation, a part of a railway and power lines. All sensors were rigidly mounted in a vibration dampened configuration on a helicopter. The helicopter flew at a constant speed of 12 m/s to mimic a small UAV.

In this contribution, we consider only some of the available sensors. We consider the pre-calibrated, IXAR180 (PhaseOne) camera with 80 megapixels and 42 mm lens and the medium-range lidar sensor VQ-480 (Riegl). We also consider the the Navchip v1 (Thales) IMU with 500 Hz sampling frequency, the performances of which are similar to the popular APX15 (Applanix) (also present on board but not considered here), and the Delta TRE-3 (Javad) multi-frequency receiver, post-processed in PPK mode with Grafnav (Novatel). The optimal recursive smoothing of PPK with a navigation-grade AIRINS (iXblue) IMU, provides the reference trajectory (T REF in Tab. 1).

We focus on two successive flight lines (FL) depicted in Fig. 3. They include 66 images with 80% forward and 40% – 60% lateral overlap and a GSD of approximately 3 cm. The swath width of the lidar is about 180 m and the side overlap is close to 40%. The field-of-view of the lidar unit is almost the same as that of the camera ( $\approx 60^\circ$ ). Thus, the lidar swath coincides with the image footprints in Fig. 3. The helicopter was flown at 230 m above ground level which resulted in a point-cloud density of 35 pts/m<sup>2</sup> to 70 pts/m<sup>2</sup> (lidar GSD between 10 and 20 cm).

##### 4.2 Study cases

The system and optical-sensor calibration parameters, including lever-arms, boresights, and camera-intrinsic parameters, were obtained by different techniques before the examined flight, including: (i) laboratory calibration of lever-arm proposed in (Vallet, Skaloud, 2004), (ii) in-flight recovery of boresight<sup>1</sup> between IMU and optical sensors as described in (Skaloud, Lichti, 2006), (iii) in-flight determination of camera additional parameters as detailed in (Lichti et al., 2008). These are considered as known. The time-correlated errors of the industrial-grade IMU are treated as unknowns in the dynamic network.

While the GNSS-derived position and velocity information is identical to that of the reference, we examine four trajectory determination approaches (summarized in Tab. 1) using different subsets of the considered sensors (see Section 4.1).

**4.2.1 Reference trajectory (T REF):** We consider the trajectory computed with the navigation-grade IMU as ground-truth. We use the APPS software (iXblue) to integrate the AIRINS IMU reference data with the GNSS position and velocity (PPK). The trajectory integrated with the AIRINS IMU has attitude errors smaller than  $< 0.003^\circ$ , which is roughly one order of magnitude lower than those based on MEMS IMUs. This implies that, at ranges  $< 250$  m, the maximum error due to reference attitude is lower than the laser ranging precision, which is  $\approx 2$  cm ( $1\sigma$ ). The boresight between the AIRINS and the Navchip IMUs is known. Since the two units operate under the same conditions, we can safely consider the AIRINS-trajectory as ground-truth for a fair comparison.

<sup>1</sup>The boresight between the reference and the MEMS-IMU was determined as a mean value that minimises the attitude differences between the two trajectories obtained by applying the recursive smoother (RS) to data from both IMUs over almost 1 hour.

**4.2.2 Recursive Smoothing (T<sub>0</sub>):** In this approach the trajectory is generated via the loosely coupled integration of the Navchip v1 IMU readings with the GNSS position and velocity solution in a recursive smoother. We use Posproc (Applanix) with internally designed filters tuned for the Navchip v1 IMU.

**4.2.3 DN with 3D tie-points (T<sub>1-3D</sub>):** This approach corresponds to the one proposed in (Brun et al., 2022) in which 3D tie-points between overlapping point-clouds, raw inertial and GNSS observations are tightly fused in a DN. The automatic 3D tie-point extraction is performed as described in Section 2.

**4.2.4 DN with 2D tie-points (T<sub>2-2D</sub>):** This approach corresponds to the one proposed in (Cucci et al., 2017) according to which 2D image tie-points, raw inertial and GNSS readings are fused simultaneously in a DN. This trajectory can be obtained with the freely available tools presented in (Cucci, 2022). The automatic 2D tie-point extraction is achieved using Metashape (Agisoft), where after initial alignment using camera poses derived from T<sub>0</sub> (and before dense matching) we extract the image coordinates of the tie-points per photo. The image observations are used in the DN along with the angular velocities, the specific forces and the GNSS position observations. The camera intrinsic calibration and the boresight with respect to the Navchip v1 IMU were previously estimated and are considered as known.

**4.2.5 DN with 3D and 2D tie-points (T<sub>3-3D-2D</sub>):** This approach combines lidar and camera observations as proposed in Section 3. Both 2D and 3D tie-points are jointly integrated with raw inertial and GNSS measurements in a DN.

All approaches allow us to obtain a high-frequency trajectory for the IMU reference frame. These trajectories are employed to geo-reference the lidar ranges by means of the in-house software LIEO (Skaloud, 2017), and obtain the final point-cloud, (the lidar boresight has been previously estimated). The accuracy of each of the four approaches is evaluated based on deviations of each point in the point-cloud with respect to the reference (obtained by geo-referencing the same lidar ranges using the reference trajectory, T<sub>REF</sub>).

Traj. #	Est. method	3D tie-points	2D tie-points
T <sub>REF</sub>	RS	no	no
T <sub>0</sub>	RS	no	no
T <sub>1-3D</sub>	DN	yes	no
T <sub>2-2D</sub>	DN	no	yes
T <sub>3-3D-2D</sub>	DN	yes	yes

**Table 1.** Overview of the examined trajectories and their characteristics, i.e., the estimator used to compute them and the inclusion of tie-points with the IMU and GNSS observations. RS: recursive smoother, DN: dynamic network.

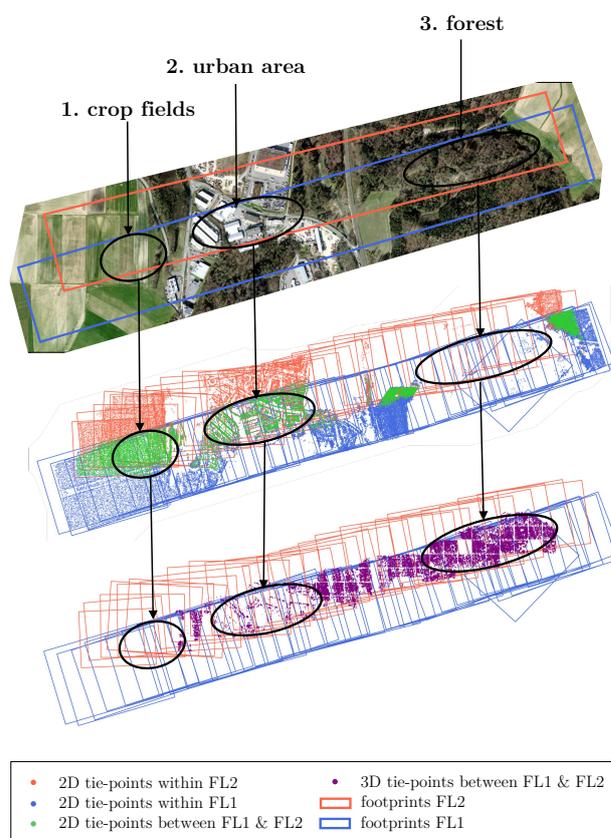
## 5. RESULTS

### 5.1 Distribution of 2D and 3D tie-points

An interesting property is observed in the spatial distribution of 2D and 3D tie-points. 2D tie-points are mainly detected in the urban areas or areas with low vegetation. For example, feature extraction is adequate in parts of the imagery that correspond to agricultural land, since the different colors of each crop and the harvesting traces create sufficient texture for the automated

2D feature matching (Fig. 4 - area 1). 2D feature detection is sensitive to surface homogeneity present in, e.g., vegetated, rocky or snow-covered areas. As a consequence, in parts of the imagery dominated by forest or high vegetation (Fig. 4 - area 3), few 2D tie-points are found. This results not only in sparse regions in the 3D point-cloud that can be obtained from photogrammetry only (e.g., by dense matching) and holes in the final digital surface model, but also in less controlled trajectory determination for such areas.

The opposite situation occurs with the automatic 3D tie-point extraction. As also reported in (Brun et al., 2022), in the forest area (Fig. 4 - area 3) there are almost five times more 3D tie-points than in the built-up regions. This is due to the ISS algorithm that detects more features in high-vegetation. In the same figure, one can also distinguish the 50 × 50 m tiles in which the point-cloud was split for 3D tie-point extraction.



**Figure 4.** Distribution of 2D and 3D tie-points along the two flight lines and the mapping surface they correspond to.

It appears that each automatic feature extraction workflow is complementary, i.e., their combination allows to establish tie-points where the other cannot and this offers continuous control (conditioning) along the trajectory when 2D and 3D tie-points are considered together (T<sub>3-3D-2D</sub>). This was somewhat unexpected as it is known that it is notoriously difficult to detect features in high vegetation surfaces, e.g., trees, due to the high degree of homogeneity and the possible moving branches. This fact is better examined in the following section.

### 5.2 Trajectory comparison

In the following, we present the orientation error statistics of the different trajectories with respect to T<sub>REF</sub>. These are sum-

marized in Tab. 2. We don't report position error statistics since they are similar for all methods. Indeed, the estimated position is largely driven by the GNSS measurements, resulting in negligible, sub-cm, differences for all methods.

We note that all methods for trajectory correction, i.e., including 2D, 3D or both types of tie-points in the DN adjustment, lead to more accurate orientation estimates with respect to the one obtained via optimal recursive smoothing (T 0). Such improvement is clearly visible in the mean error, where all methods achieve roll and pitch accuracy comparable to the reference ( $\approx 0.003^\circ$ ), but also in standard deviation, where an improvement by a factor of 2–3 with respect to T 0 is observed for all methods. A residual bias is visible in the yaw for all methods, but still  $\approx 4$  times smaller than what is obtained with T 0. The higher accuracy of the estimated orientation directly translates to a higher accuracy point-cloud once the estimated trajectories are employed for direct geo-referencing; this will be discussed in more detail in Section 5.3.

	Traj. #	T 0	T 1•3D	T 2•2D	T 3•3D•2D
Roll ( $^\circ$ )	MEAN	0.018	-0.004	0.003	<b>-0.001</b>
	STD	0.034	<b>0.018</b>	0.024	0.020
	RMSE	0.038	<b>0.019</b>	0.024	0.020
Pitch ( $^\circ$ )	MEAN	-0.031	<b>-0.001</b>	-0.012	0.002
	STD	0.042	0.019	0.028	<b>0.018</b>
	RMSE	0.052	0.019	0.030	<b>0.018</b>
Yaw ( $^\circ$ )	MEAN	-0.080	<b>-0.021</b>	-0.039	-0.037
	STD	0.172	0.048	0.062	<b>0.034</b>
	RMSE	0.189	0.053	0.073	<b>0.050</b>

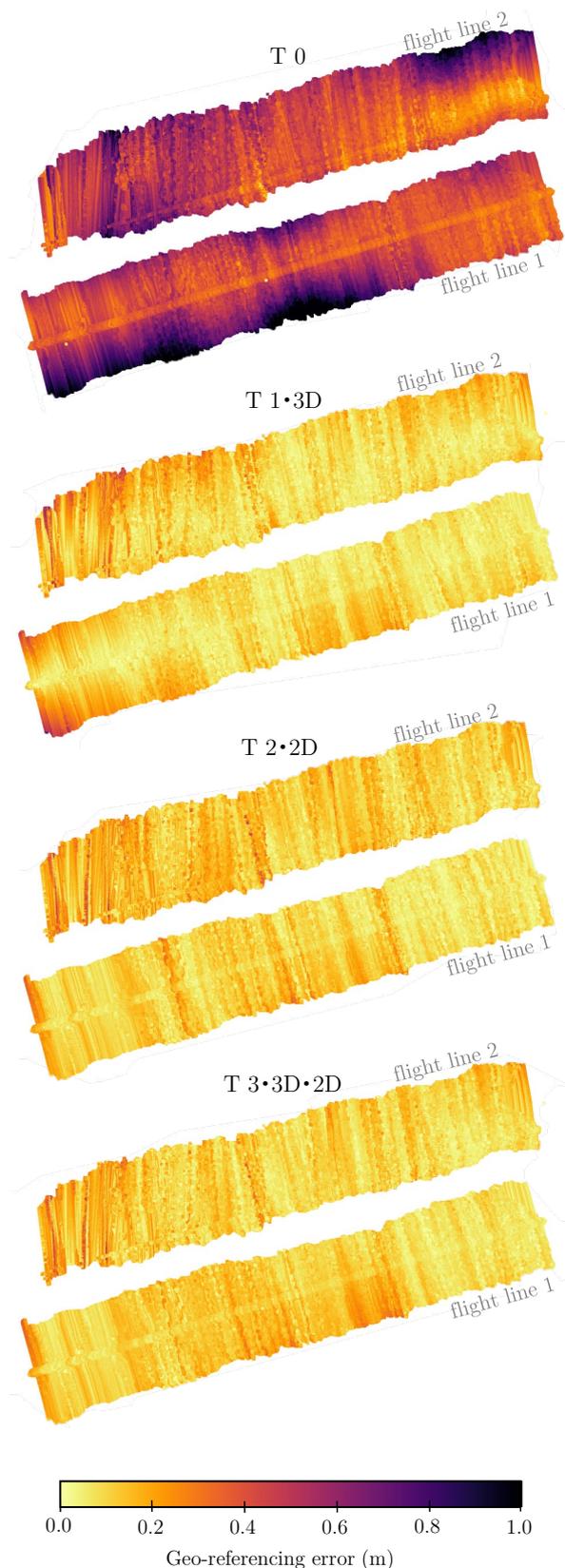
**Table 2.** Error statistics of the estimated trajectories. The best trajectory for each metric (line) is highlighted in bold.

We also note that 3D tie-points appear to be responsible for most of the improvement in the orientation error. In general, photogrammetry (T 2•2D) exhibits higher residual biases in pitch and yaw, probably because of the weak geometry of the flight and the non-uniform distribution of 2D tie-points over the surveyed area. Note that no ground control points were used; residuals on the check-points are presented together with the employed DN solver in (Cucci, 2022) (their RMS remains at sub-pixel level). Nevertheless, the use of both types of tie-points together (T 3•3D•2D) slightly reduces error statistics in most of the cases, especially in yaw (STD and RMS) with a slightly higher mean (bias). It becomes clear that in order to take full advantage of both types of tie-points, all sensor boresights need to be resolved better than  $0.01^\circ$ . Given the weak geometry of the examined flight, it is not recommended to refine such parameters within the DN adjustment, for more details please see (Brun et al., 2022).

### 5.3 Point-cloud comparison

An accurately estimated trajectory is reflected in the geo-referencing error of the lidar point-cloud. In the considered scenario, the attitude errors are responsible for over 90% of the geo-referencing error budget. As described in Section 4.2, a point-cloud is generated for each one of the four computed trajectories using the system parameters and the raw lidar measurements. The accuracy of each output is evaluated by comparing each point-cloud with the one generated from the reference trajectory. The difference of point coordinates with respect to

the reference corresponds to the geo-referencing error which is shown in Fig. 5 while Tab. 3 presents the error statistics.



**Figure 5.** Comparison of the point-cloud geo-referencing errors generated from each trajectory.

Low-accuracy IMU observations (Fig. 5 - T 0) result in a poorly geo-referenced 3D model due to imperfect trajectory determination. The mean error in this case (0.51 m in FL1 and 0.46 m in FL2) is 2–3 times larger than the mean value of the point-cloud GSD ( $\approx 0.15$  m). A significant improvement factor of 4–5 in terms of lower mean error (0.09 m and 0.12 m) is obtained when 3D tie-points are inserted in the DN estimator (Fig. 5 - T 1:3D). As is also reported in (Brun et al., 2022), when 3D tie-points are considered, the mean geo-referencing error is smaller than the mean point-cloud GSD.

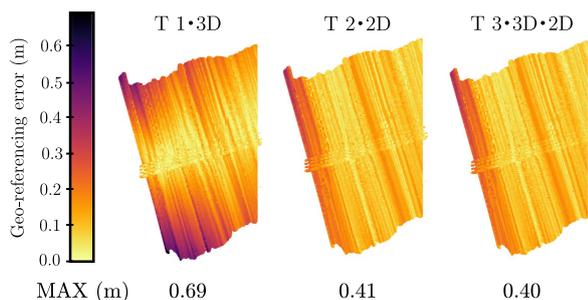
In terms of global statistics the influence of the 2D tie-points (Tab. 3 - T 2:2D) on the geo-referencing accuracy is very similar to that for the 3D tie-points. When combining the observations from both optical sensors, no global geo-referencing accuracy improvement is observed over the whole flight lines, compared to the use of 3D tie-points alone. However, there are important local differences highlighted for example in Fig. 6.

Traj. #	Flight line 2				
	MEAN (m)		RMSE (m)		
	norm	STD (m)	E	N	h
T 0	0.46	0.15	0.28	0.38	0.13
T 1:3D	0.12	0.08	<b>0.12</b>	<b>0.07</b>	<b>0.03</b>
T 2:2D	0.15	0.07	0.14	0.09	<b>0.03</b>
T 3:3D:2D	0.13	0.07	<b>0.12</b>	0.08	<b>0.03</b>

Traj. #	Flight line 1				
	MEAN (m)		RMSE (m)		
	norm	STD (m)	E	N	h
T 0	0.51	0.17	0.30	0.43	0.11
T 1:3D	0.09	0.06	0.09	<b>0.06</b>	<b>0.02</b>
T 2:2D	0.10	0.05	<b>0.08</b>	<b>0.06</b>	<b>0.02</b>
T 3:3D:2D	0.11	0.05	0.10	<b>0.06</b>	0.03

**Table 3.** Statistics of the geo-referencing error of each computed point-cloud. The best trajectory for each RMSE metric (column) is highlighted in bold.

Indeed, in the (T 3:3D:2D) trajectory, the complementary nature of the 2D and 3D matches on different surface textures gives an even distribution of tie-points along the flight lines. An example of this fact is shown in Fig. 6 that focuses on the extreme western side of flight line 1. The lack of 3D tie-points in this area results in considerable larger geo-referencing deviations between the T1 and T3 adjustments. The use of 2D tie-points in the T3 case reduces the *maximum* geo-referencing error within this area by  $\approx 0.3$  m, which is significant (Fig. 6).



**Figure 6.** Western part of the point-cloud of flight line 1.

## 6. CONCLUSIONS

In this work we have evaluated the joint contribution of 2D tie-points from camera observations and 3D tie-points from a lidar point-cloud when employed along with GNSS position and raw inertial observations in a single dynamic network adjustment. We have shown that:

1. 2D tie-points are mostly detected in urban and low-vegetation areas, while 3D tie-points are mostly detected in high-vegetation areas, where 2D feature matching is challenging; this implies that the two types of tie-points are complementary for mixed terrain types,
2. 3D tie-points are responsible for most of the improvement in the trajectory orientation error,
3. the use of 2D tie-points results in higher residual biases in pitch and yaw, probably because of the very weak flight geometry (corridor),
4. the use of both type of tie-points together yields slightly better trajectory error statistics in most cases except for a higher residual yaw bias,
5. in this specific corridor mapping case, the inclusion of image tie-points in the adjustment does not allow the further improvement of the geo-referencing accuracy of the point-cloud when evaluated over whole flight lines; in other words, the mean absolute geo-referencing error of the point-cloud is equivalent to the one including 2D tie-points only and the one using 3D tie-points only,
6. in all cases the mean error in the point-cloud coordinates improves by 4–5 times with respect to direct geo-referencing and its value is (1.2–1.5 $\times$ ) smaller than the mean GSD value of the point-cloud while the standard deviation (norm) is reduced by a factor of 2–3,
7. when both optical observations are considered, the maximum geo-referencing errors are reduced significantly where 3D tie-points are sparse or absent.

This presented use case provides limited initial, yet encouraging, indications of the practical benefit of tie-points detected in both sets of optical observations within the adjustment involving raw inertial observations via the dynamic network paradigm. More insights are expected from future more detailed studies.

## ACKNOWLEDGEMENTS

The helicopter flight and the reference sensors were provided by Helimap System. This contribution was partially supported by the Swiss National Foundation (SNF), project n. 200021\_182072, by the European H2020 project GAMMS, EC n. 101004255 and by the Swiss Innovation Agency (Innosuisse), project n. 37308.1.

## REFERENCES

- Ao, S., Hu, Q., Yang, B., Markham, A., Guo, Y., 2021. SpinNet: Learning a General Surface Descriptor for 3D Point Cloud Registration. *arXiv (preprint)*. <https://arxiv.org/abs/2011.12149>.
- Brun, A., Cucci, D. A., Skaloud, J., 2022. LiDAR point-to-point correspondences for rigorous registration of kinematic scanning in dynamic networks. *arXiv (preprint)*. <https://arxiv.org/abs/2201.00596>.

- Cledat, E., Skaloud, J., 2020. Fusion of photo with airborne laser scanning. *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences*, V-1-2020, 173–180.
- Colomina, I., Blazquez, M., 2004. A unified approach to static and dynamic modeling in photogrammetry and remote sensing. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 35-B1, 178-183.
- Cucci, D. A., 2022. ODyN: An online dynamic network solver for photogrammetry and LiDAR geo-referencing. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*.
- Cucci, D. A., Matteucci, M., 2014. Position tracking and sensors self-calibration in autonomous mobile robots by gaussian optimization. *2014 IEEE International Conference on Robotics and Automation (ICRA)*, 1269–1275.
- Cucci, D. A., Rehak, M., Skaloud, J., 2017. Bundle adjustment with raw inertial observations in UAV applications. *ISPRS Journal of Photogrammetry Engineering and Remote Sensing*, 130, 1-12.
- Cucci, D. A., Skaloud, J., 2019. On Raw Inertial Measurements In Dynamic Networks. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, IV-2/W5, 549–557.
- Fischler, M. A., Bolles, R. C., 1981. Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography. *Commun. ACM*, 24(6), 381–395.
- Glennie, C. L., 2007. Rigorous 3D error analysis of kinematic scanning LIDAR systems. *Journal of Applied Geodesy*, 1, 147-157.
- Glira, P., Pfeifer, N., Briese, C., Ressel, C., 2015. A correspondence framework for ALS strip adjustments based on variants of the ICP algorithm. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, IV-2/W5, 567–574.
- Glira, P., Pfeifer, N., Mandlbürger, G., 2016. Rigorous Strip Adjustment of UAV-based Laserscanning Data Including Time-Dependent Correction of Trajectory Errors. *Photogrammetric Engineering and Remote Sensing*, 82, 945-954.
- Glira, P., Pfeifer, N., Mandlbürger, G., 2019. Hybrid orientation of airborne LiDAR point clouds and aerial images. *ISPRS Annals of the Photogrammetry, Remote Sensing 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. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2019-June, 5540–5549. Cited By :72.
- 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 Photogrammetry and Remote Sensing*, 154, 41-58.
- Hussnain, Z., Oude Elberink, S., Vosselman, G., 2021. Enhanced trajectory estimation of mobile laser scanners using aerial images. *ISPRS Journal of Photogrammetry and Remote Sensing*, 173, 66-78.
- Jayendra-Lakshman, M., Devarajan, V., 2013. A new feature descriptor for LiDAR image matching. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, II-2/W1, 157–162.
- Kager, H., 2004. Discrepancies between overlapping laser scanning strips - simultaneous fitting of aerial laser scanner strips. *Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 35-B1, 555–560.
- Kerstling, A., Habib, A., Bang, K.-I., Skaloud, J., 2012. Automated approach for rigorous light detection and ranging system calibration without preprocessing and strict terrain coverage requirements. *Optical Engineering*, 51(7), 076201.
- Li, J., Yang, B., Chen, C., Habib, A., 2019. NRLI-UAV: Non-rigid registration of sequential raw laser scans and images for low-cost UAV LiDAR point cloud quality improvement. *ISPRS Journal of Photogrammetry and Remote Sensing*, 158, 123-145.
- Lichti, D., Skaloud, J., Schaer, P., 2008. On the calibration strategy of medium format cameras for direct georeferencing. *International calibration and orientation workshop EuroCOW 2008*.
- Mandlbürger, 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*, IV-2/W4, 259–266.
- Nex, F., Armenakis, C., Cramer, M., Cucci, D., Gerke, M., Honkavaara, E., Kukko, A., Persello, C., Skaloud, J., 2022. UAV in the advent of the twenties: Where we stand and what is next. *ISPRS Journal of Photogrammetry and Remote Sensing*, 184, 215-242.
- Pham, Q.-H., Uy, M. A., Hua, B.-S., Nguyen, D. T., Roig, G., Yeung, S.-K., 2019. LCD: Learned cross-domain descriptors for 2d-3d matching.
- Salti, S., Petrelli, A., Tombari, F., Di Stefano, L., 2012. On the affinity between 3d detectors and descriptors. *2012 Second International Conference on 3D Imaging, Modeling, Processing, Visualization Transmission*, 424–431.
- Sarlin, P., DeTone, D., Malisiewicz, T., Rabinovich, A., 2019. SuperGlue: Learning Feature Matching with Graph Neural Networks. *CoRR*, abs/1911.11763. <http://arxiv.org/abs/1911.11763>.
- Schaer, P., Skaloud, J., Stebler, Y., Tome, P., Stengele, R., 2009. Airborne LiDAR in-flight accuracy estimation. *GPS World*, 20(8), 37-41. <https://infoscience.epfl.ch/record/140564?ln=en>.
- Shi, C., Chen, X., Huang, K., Xiao, J., Lu, H., Stachniss, C., 2021. Keypoint Matching for Point Cloud Registration Using Multiplex Dynamic Graph Attention Networks. *IEEE Robotics and Automation Letters*, 6(4), 8221-8228.
- Skaloud, J., 2017. Lieo : Lidar orientation and point cloud generation, software description n°6.0704.
- Skaloud, J., Lichti, D., 2006. Rigorous approach to bore-sight self calibration in airborne laser scanning. *ISPRS Journal of Photogrammetry and Remote Sensing*, 61, 47-59.
- Tombari, F., Salti, S., Di Stefano, L., 2010. Unique signatures of histograms for local surface description. K. Daniilidis, P. Maragos, N. Paragios (eds), *Computer Vision – ECCV 2010*, 6313, Springer Berlin Heidelberg, 356–369.
- Vallet, J., Gressin, A., Clausen, P., Skaloud, J., 2020. Airborne and mobile LiDAR, which sensors for which application? *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLIII-B1-2020, 397–405.
- Vallet, J., Skaloud, J., 2004. Development and experiences with a fully-digital handheld mapping system operated from a helicopter. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XXXV, Istanbul, Part B, Commission 5.
- Zhong, Y., 2009. Intrinsic shape signatures: A shape descriptor for 3d object recognition. *2009 IEEE 12th International Conference on Computer Vision Workshops, ICCV Workshops*, 689–696.