A flexible targetless LiDAR–GNSS/INS–camera calibration method for UAV platforms

Quentin Pentek, Pol Kennel, Tristan Allouis, Christophe Fiorio, Olivier Strauss

A Flexible Targetless LiDAR–GNSS/INS–Camera Calibration Method for UAV Platforms

Quentin Pentek[a,b,*], Pol Kennel[b], Tristan Allouis[b], Christophe Fiorio[a], Olivier Strauss[a]

[a] *LIRMM, Univ. Montpellier, CNRS, 860 rue de St Priest, 34095 Montpellier, France*
[b] *YellowScan, 1 chemin de Fescau, 34980 Montferrier-sur-Lez, France*

Abstract

There is a growing need for 3D colored maps acquired from multi-sensor moving platforms. Accurate multi-sensor data alignment is an important prerequisite for the construction of 3D colored maps derived from simultaneously acquired camera and Light Detection and Ranging (LiDAR) data. However, current alignment methods are hampered by low automation, heavy computational costs or tedious system calibration set-ups. In this paper, we consider a LiDAR–global navigation satellite system (GNSS)/inertial navigation system (INS)–camera system mounted on an Unmanned Aerial Vehicle (UAV) platform. We present a detailed literature review of existing calibration methods for such systems. We propose a new versatile automatic and targetless calibration method of this system. This method involves estimating the calibration parameters by optimizing the correspondence between pairs of conjugate image points extracted from overlapping images and the projection of these points onto the georeferenced LiDAR point cloud. Experiments on actual data show the suitability of this method for the construction of 3D colored point clouds. Quantitative calibration results using checkpoints indicate that the obtained calibration accuracy is compatible with the accuracy of the georeferenced LiDAR point cloud, i.e. 5 cm. Further experiments on simulated data show the robustness of this approach to initial calibration parameters and low sensitivity to LiDAR point cloud density and noise. As this method is quite flexible, we believe it is more suitable for 3D color map generation than other methods proposed in the literature.

Keywords: Alignment, automatic, laser scanner, multi-sensor, optical imagery, point cloud colorization

1. Introduction

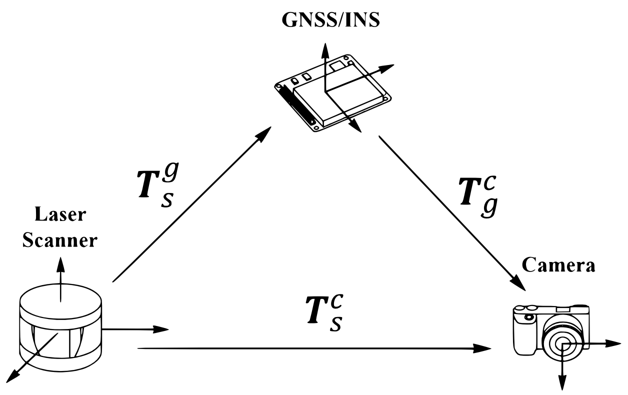


Figure 1: Mutli-sensor configuration.

Light-weight Light Detection and Ranging (LiDAR) sensors, originally developed for automotive applications, have enabled the development of mapping and surveying applications from unmanned aerial vehicles (UAV) over the past decade [1]. Today, UAV platforms equipped with LiDAR and global navigation satellite system (GNSS)/inertial navigation system (INS) sensors provide a lower-cost alternative to airborne LiDAR for quick and efficient generation of 3D maps at the square kilometer scale. Though LiDAR might be seen as a competitor to photogrammetry, LiDAR and imagery are two complementary technologies. Whereas LiDAR natively delivers 3D point clouds, imagery provides 2D to 2.5D rasters of an object's spectral properties. Matching the color on the geometry leads to colored point clouds or a color-draped digital surface model (DSM) which generates valuable information for the model, e.g. for classification [2]. However, in order to take advantage of their complementarity, those two heterogeneous data must be properly aligned.

LiDAR and cameras have different acquisition principles and therefore capture different features; LiDAR data is able to represent complex 3D objects with poor spatial resolution and color information, while cameras capture 2D dense and colored information. These differences in modalities make it very challenging to solve the LiDAR–camera calibration automatically in natural environments without markers or precise initialization of the calibration parameters. Many research studies have addressed

*Corresponding author

Email addresses: quentin.pentek@gmail.com (Quentin Pentek), pol.kennel@yellowscan-lidar.com (Pol Kennel), tristan.allouis@yellowscan-lidar.com (Tristan Allouis), christophe.fiorio@lirmm.com (Christophe Fiorio), olivier.strauss@lirmm.fr (Olivier Strauss)

the alignment problem regarding camera and LiDAR data in the past decade [3–7], but no versatile solution has been found yet [8, 9]. Recent studies, such as [10–12], corroborate the fact that efficient camera and LiDAR integration on moving platforms is still an unsolved research problem.

In this paper we propose a novel approach for camera and LiDAR data alignment by avoiding the above mentioned limitation. The proposed alignment method intends to jointly calibrate 3 rigidly mounted sensors: a laser scanner (s), a camera (c), and a GNSS-Aided INS (g). As in [11], we rely on time-synchronized camera and LiDAR datasets with respect to the GNSS/INS measurements.

The proposed method

- is fully automatic
- requires no calibration markers or target
- is suitable for natural environments as it does not require the capture of remarkable geometric features, e.g. over urban scenes
- does not require LiDAR intensity data
- does not require initial calibration parameters close to their optimum values.

Besides the above-mentioned advantages of our approach, the limitations are summarized as follows:

- The method relies on the availability of sufficiently accurate GNNS/INS data.
- Our approach does not estimate the camera intrinsic parameters but requires them to be pre-calibrated.

Calibrating a multi-sensor system, whose individual sensors have been pre-calibrated, consists of estimating the six mounting parameters; three rotational parameters, called boresight angles (roll, pitch, heading), and three translational parameters called lever-arm. Those parameters can be represented by a rigid-body homogeneous transformation matrix T_a^b which depicts the position and orientation of frame b of sensor B with respect to frame a of sensor A. Assuming that each sensor is rigidly mounted, estimating the calibration parameters of T_g^c and T_s^g is equivalent to estimating T_s^c, because $T_g^c T_s^g = T_s^c$, as illustrated in Fig. 1.

In this paper we propose a new method that solves the full LiDAR–GNSS/INS–camera calibration problem. Our approach successively computes T_s^g and T_g^c in a way that ensures data consistency, and thus accurate LiDAR–camera sensor alignment.

The paper is structured as follows. In section 2, a detailed overview of state-of-the-art methods for LiDAR-camera, LiDAR–GNSS/INS, GNSS/INS–camera, joint LiDAR–GNSS/INS–camera data alignment, and point cloud colorization is presented. Section 4 describes the study site and the datasets used for the proposed calibration method and its evaluation. Section 3 introduces the proposed calibration method and section 5 describes the experiments carried out to evaluate the performance of the proposed calibration method. The results are presented in section 6 and we discuss the results and some limitations of our methods in section 7. This work is finally concluded in section 8.

2. Related Work

In this section, we review major state-of-the-art methods that address the problem of aligning data captured from a multi-sensor system composed of a LiDAR, a camera and a GNSS/INS. Only a few previous studies addressed the joint calibration problem for such a multi-sensor system. Most studies only consider a pairwise calibration between two of the three sensors. Therefore, we present methods that align data from either a LiDAR and a camera, a LiDAR and a GNSS/INS, a camera and a GNSS/INS and a full multi-sensor system. The presented methods are applied to airborne platforms as well as to UAV-borne and terrestrial platforms.

2.1. LiDAR–Camera Alignment

Extensive studies have been conducted on matching LiDAR and camera data and the different methods can be classified depending on the acquisition platform (airborne, UAV-borne, or terrestrial), dimension of conjugate features (2D-2D, 2D-3D, 3D-3D), human-interaction level (manual, semi-automatic, fully automatic), observed scene (urban versus natural environment), operation mode (online versus offline), etc.

In the literature review hereafter, we have distinguished target-based from targetless methods. Among targetless approaches, we segmented 3D-3D alignment, motion-based and feature-based methods, with the latter being further divided into feature-based and dependence-based methods. The advantages and limitations of each approach are discussed.

2.1.1. Target-based Methods

Target-based methods utilize a calibration object, or calibrated markers, which can be identified in both sensor modalities. A geometric constraint is then found to estimate the best rigid-body transformation that aligns the extracted calibration object or features from both modalities in a common space.

To our knowledge, the first published method that addressed the calibration problem regarding a LiDAR–camera system in a robotic context was in [13]. By using a V-shaped target, the authors define a geometric relationship between the laser range finder and the camera. They estimate the mounting parameters through a nonlinear least-square function minimization. In [14] and [15], the authors calibrate a laser range finder and a camera by using camera and LiDAR observations of a checkerboard viewed from multiple angles. Laser points lying on the checkerboard pattern and the normal vector of the calibration plane estimated in the camera reference frame provide constraints on the calibration parameters. These parameters are then estimated by minimizing the distance from the laser points lying on the checkerboard pattern to the corresponding plane observed on the image. Thereafter, several modifications of the previous methods were proposed using single or multiple checkerboards [16–19] but also different calibration patterns, such as a circular target [20], a spherical target [21] or a trihedral calibration object [22]. Apriltags were also used in [23] to calibrate a multi-camera and multi-LiDAR system.

The main advantage of target-based calibration methods is their ability to provide accurate calibration without requiring any initial pose estimation. However, as a shortcoming, target-based calibration is often conducted indoors in non-operating conditions. As the target is closer to the multi-sensor system than the objects of interest, the calibration would most likely be biased. Moreover, target-based methods are often performed once based on the assumption that the calibration parameters will not be altered after several task repetitions. This may be a valid assumption for static platforms, but it might be not true for mobile platforms.

2.1.2. 3D-3D Alignment Methods

Another way to align LiDAR and camera data is via 3D-3D alignment methods which are based on cloud-to-cloud registration. A 3D point cloud representing the captured scene is usually reconstructed from multiple images in a bundle adjustment framework [24] and then aligned with the recorded LiDAR point cloud of the same scene. For cloud-to-cloud registration, the iterative closest point (ICP) algorithm [25, 26] is a method of choice with several advantages. ICP has been proven to converge and is straightforward to implement. Moreover, many variants tailored for specific tasks at hand have been developed (see [27] for a comprehensive review). The study in [28] proposes an automatic video-to-3D registration framework using aerial oblique video images and LiDAR data. The relative camera poses are retrieved by frame tracking and alignment. Motion stereo is used to compute a dense 3D point cloud from the video, which is then aligned with LiDAR data using ICP. A coarse-to-fine registration method that aligns UAV-borne LiDAR and camera data is proposed in [7]. Coarse registration is performed by extracting and matching building outlines in a LiDAR point cloud and images. Fine alignment is then achieved using ICP on the point cloud and a dense 3D photogrammetric model reconstructed using structure-from-motion (SfM) and multi-view-stereo algorithms. In a recent study [10], the authors propose a coarse-to-fine registration of LiDAR and camera data acquired from a low-cost UAV. Coarse registration is performed using a GNSS/INS-aided SfM, which aims to correct the GNSS/INS trajectory. Fine registration is then carried out by iteratively minimizing the difference between the depth maps derived from SfM reconstructed point clouds and the projected laser points. In this work, the cloud-to-cloud registration is achieved in the image space.

The main advantage of 3D-3D alignment methods is to allow precise registration using a reconstructed geometry of the captured scene from image and LiDAR data. However, the huge computing cost underlying those methods is a major drawback, as dense reconstruction from multiple images is often required for accurate results. Furthermore, ICP requires good initial estimates to converge to the global optimum.

2.1.3. Motion-based Methods

Motion-based methods exploit the motion of rigidly mounted sensors on a moving platform to estimate the sensors mounting parameters. These methods are closely related to the hand-eye calibration problem [29] addressed by the robotic community, where a camera (“eye”) is rigidly mounted on a robot gripper (“hand”). The aim of the hand-eye calibration is to estimate the unknown transformation between the camera and the gripper coordinate frames based on the motions undergone by the gripper and camera, with the latter being estimated from captured images. In early studies [30, 31], the main limitation of motion-based methods was that calibrated markers were required to estimate the camera motion. More recently, [32–35] make use of visual odometry and SfM techniques to overcome this limitation. In [36], the authors extend the hand-eye calibration framework to initialize the mounting parameters of a 3D LiDAR unit and a camera mounted on a moving robot. The motion each sensor undergoes is estimated independently. LiDAR motion is estimated using the ICP algorithm, while camera motion is computed using standard image feature point tracking.

The advantage of motion-based techniques is that they do not require precise sensor pose initialization. In addition, no sensor overlap is required. Nevertheless, these methods are limited by the specific drawbacks of the techniques used to estimate the motion of each sensor. Moreover, a major limitation of motion-based methods is that they require large range of motion to give accurate calibration results. Finally, precise temporal registration between sensor motions is required.

2.1.4. Feature-based Methods

Feature-based methods retrieve the best calibration parameters by extracting and matching conjugate features from LiDAR and camera data. According to [37], finding conjugate features follows typically one or a combination of the following approaches:

- extract 3D features from LiDAR data and stereo images or a reconstructed photogrammetric model (2.5D/3D)
- extract 3D features from LiDAR data and 2D features from images
- create a synthetic 2D image from a LiDAR point cloud and then extract 2D features from both datasets.

Low-level feature techniques extract and match features such as edges and corners to usually determine the alignment between LiDAR derived intensity and optical images [38–40]. High-level feature techniques utilize regions, building contours or line-segment-based features extracted from images and from point clouds to align LiDAR data and optical imagery [41, 42]. The authors in [43] propose a new feature composed of connected line segments to compute 3D-2D correspondences between aerial LiDAR data and aerial images. A two-level random sample consesus (RANSAC) scheme [44] is then used to achieve a robust estimation of the calibration. The method described in [6] combines corner-based and segment-based features to match airborne LiDAR and optical imagery depth maps. A two-level RANSAC is used to overcome outlier problems for camera pose estimation.

Feature-based methods have the advantage of being well suited in situations where distinctive structural details are present within both data modalities, such as urban or man-made environments. But automatic feature-based methods are not very effective in natural environments [45]. Furthermore, optical images and LiDAR data capture different feature characteristics, and even between intensity and visible images, low-level features may not have correspondences, thus leading to failure. Finally, automatic high-level feature extraction methods generally require manual supervision or intervention for the calibration to be accurate. The alignment accuracy strongly depends on the feature extraction quality.

2.1.5. Dependence-based Methods

In dependence-based methods, calibration is performed by maximizing a dependence metric expressing a signal similarity between LiDAR data and optical imagery. Usually the signal is expressed in a two-dimensional space, so a synthetic LiDAR image must be created by projecting and interpolating LiDAR points on the image grid. The method is based on the assumption that two specific signals, i.e. one extracted from LiDAR data and the other from image data, are somehow correlated. The calibration parameters are estimated by maximizing a dependence index in an optimization framework. Two major similarity metrics are utilized in dependence-based methods: χ^2 statistics and the mutual information (MI).

The authors of [46] propose to maximize χ^2 statistics between a LiDAR reflectance image and a gray-scale derived RGB image in order to estimate the mounting parameters of the LiDAR-camera system. Assuming two random variables X and Y, the χ^2 statistic gives a measure of how close the observed X-Y joint distribution would be to the distribution obtained by assuming that X and Y are statistically independent. In [46], X and Y represent the probability densities of the laser-derived reflectance image and the optical gray-scale image respectively. The authors experiments show that the method requires a good initial calibration value because the χ^2 statistic's global maximum does not always correspond to the desired alignment.

The MI similarity metric is used in [47] to match aerial imagery of an urban scene on LiDAR point clouds by minimizing the joint entropy, which is equivalent to maximizing the MI, between the grayscale-encoded LiDAR elevation, LiDAR return intensity and optical imagery. A correlation between the LiDAR elevation and image luminance is assumed, which is suited for urban scenes with high buildings. The authors of [48] introduce the combined mutual information (CMI) method for multivariable statistical similarity using matched aerial LiDAR DSM and intensity values with aerial optical imagery. The study shows that CMI techniques improve registration accuracy and robustness compared to conventional MI. In [49] and [4], the authors manage to align LiDAR intensity images and and gray-scale images by maximizing the MI. The method proposed in [50] uses normalized MI and particle swarm optimization to compute the rigid body transformation and the camera focal length between a synthetic LiDAR image and an optical image. Depending on the application, different LiDAR features such as intensity return values, or estimated surface normals are used to generate the LiDAR image. The authors of [3] propose a new dependence metric, i.e. the gradient orientation measure (GOM), which computes how well gradient orientations are aligned between a LiDAR intensity image and optical images. In [51], the authors propose a coarse-to-fine method to register large-scale urban terrestrial LiDAR data and optical images. A new dependence metric is introduced and an alternative method in case of nonavailable LiDAR reflectance values is proposed. The authors of [52] use a dependence metric for online calibration of an automotive LiDAR-camera system. The metric is based on the idea that LiDAR depth discontinuities should correspond to an edge in the images. The method involves projecting LiDAR points corresponding to depth discontinuities in an edge-processed image. A score is then computed, that rewards LiDAR points falling on an image edge, assuming that the score is maximal when the two sensors are correctly aligned.

Dependence-based methods allow generally fast, automatic and fine sensor calibration. However, the optimization problem is usually highly non-convex. These methods require a highly constrained search space and are thus solely limited to locally-optimized calibration refinement. Moreover, for dependence metrics using intensity, nonuniform lighting (e.g. shadows), can play a critical role in drastically decreasing LiDAR reflectivity and image intensity correlation.

2.2. LiDAR–GNSS/INS Calibration

In this section, we review methods that address the problem of estimating calibration parameters between a GNSS/INS and a laser scanner. Incorrect LiDAR–GNSS/INS data alignment is the main source of error in data produced by airborne and terrestrial mobile mapping laser systems. Various methods have been proposed for correcting LiDAR–GNSS/INS calibration parameters by eliminating discrepancies in overlapping point cloud areas. Extensive research revealed two main categories of calibration approaches that should be mentioned, i.e. approximate and (quasi-)rigorous methods.

Approximate methods, such as those described in [53–55], are data-driven methods that solely use the positional information of georeferenced LiDAR points to reduce strip-to-strip discrepancies. Such methods do not require raw data like GNSS/INS measurements, which are not always available for end users. However, those methods are considered non-rigorous because all the biases related to point cloud georeferencing cannot be compensated by arbitrary transformations.

On the other hand, rigorous methods, such as [56–58] fully or partially (e.g. in [55]) model the LiDAR–GNSS/INS georeferencing principle in order to eliminate systematic errors in the calibration parameters causing misalignment between LiDAR and GNSS/INS data. Such methods express strip-to-strip discrepancies as a function of the calibration parameters and by integrating LiDAR and GNSS/INS measurements, such as the range and scan angle of the laser scanner, and position and orientation observations of the GNSS/INS unit.

In order to quantify strip-to-strip discrepancies, conjugate tie points, planar patches and/or modeled surfaces are usually matched in overlapping LiDAR point clouds from survey strips. Those discrepancies are then used to determine a misalignment criterion generally defined as a distance (e.g. [59]). Due to the nature of LiDAR data, automatic identification of conjugate points is not reliable [53] and should therefore be performed manually. Furthermore, the identification and selection of conjugate surfaces requires fastidious pre-processing steps (e.g. region growing, principle component analysis or RANSAC) and adapted sites, e.g. urban environments [54, 56, 58, 60]. However, point-to-patch matching methods [61, 62] advantageously present direct and automated correspondences.

Once the correspondences are established, the misalignment criterion is expressed as a function of the sought-after parameters given the utilized model (rigorous, quasi-rigorous, etc.). Least squares adjustment (LSA) is then performed to estimate the modeled parameters.

In [62], the authors propose a method for LiDAR strip adjustment. Their approach is able to correct biases on the LiDAR–GNSS/INS calibration parameters as well as additional systematic errors such as laser-beam encoder offsets or scale factors and biases in the GNSS/INS observations. Their method is based on the ICP methodology, where discrepancies between robustly selected point-to-plane correspondences from overlapping LiDAR strips are minimized via of LSA.

The problem of optimal selection of LiDAR observations for strip adjustment was recently addressed in [63]. The authors rely on modeled measurement uncertainties of georeferenced LiDAR points in order to achieve a minimal LSA problem size.

Note that the authors of [61] claim that the vertical lever-arm component cannot be estimated by only observing discrepancies between strip-to-strip correspondences, because an error in the vertical lever-arm parameter produces the same effect regardless of the flying direction or flying height. The vertical lever-arm offset can only be estimated if at least one vertical ground control point (GCP) is used. This limitation is again highlighted in [11]. Consequently, the authors decided to bypass the problem by manually measuring and then fixing the vertical lever-arm component of the laser scanner during the optimization. Moreover, the authors of [12] state that, *"depending on the sensors assembly, flight configuration, and terrain geometry, some of these parameters* [mentioned in [12, Table 1]] *may be completely correlated and therefore not estimable."* However, they give no further information about the parameters that are likely not estimable.

2.3. GNSS/INS–Camera Alignment

In this section, we review studies that tackle the problem of estimating calibration parameters between a GNSS/INS and an optical imaging sensor mounted on a mobile system. This problem has mainly been studied in two research fields: robotics and photogrammetry. Both fields address the problem differently. In robotics, problems such as simultaneous localization and mapping (SLAM) are of interest and fast or real-time calibration is often preferred [64, 65]. This has led to extensive development of online filtering methods [66, 67], and more recently keyframe-based nonlinear optimization methods [68, 69]. High precision is preferred over fast computation in the photogrammetry and remote sensing field. A high performance optimization framework such as bundle adjustment [24] has been used to estimate the calibration parameters. Those methods can essentially be divided in two variants: two-step and single-step methods. In the two-step procedure, the calibration parameters are estimated by comparing the measured GNSS/INS trajectory with the camera motion that is estimated using aerotriangulation and bundle adjustment [70, 71]. This method originates from the photogrammetry community and is closely related to the robotic hand-eye calibration method mentioned in section 2.1.3, where the "hand" is replaced by a GNSS/INS positioning sensor. In single-step methods, the calibration parameters are expressed as unknown in the mathematical model of the bundle adjustment procedure. Solving the bundle adjustment problem will then directly yield their estimated values [72, 73]. However, in both methods, images usually must be captured over a calibration site with uniformly distributed markers or ground control points in order to achieve precise data alignment.

In [74], the authors studied flight configuration requirements for reliable estimation of GNSS/INS–camera calibration parameters. This study showed that images taken from two opposite flight lines with 100% overlap allow good estimation of the roll and pitch parameters as well as the planimetric lever-arm components. Having images from parallel flight lines with the least possible overlap allows a more reliable estimation of the heading parameter. Moreover, the authors of [11] claim that having images taken from flight lines flown at different altitudes enables decorrelation of the planimetric lever-arm components from the roll and pitch parameters. However, for the same reasons mentioned in section 2.2, the authors in [74] also state that the vertical lever-arm component cannot be estimated by observing discrepancies between conjugate image feature points unless vertical control information is available, such as a GCP.

In [75], a GNSS/INS–camera boresight angle calibration method is proposed using a LiDAR digital elevation model (DEM). Manually selected tie points from overlapping images are triangulated in the 3D object space and then refined using the LiDAR elevation data. The boresight angles are then estimated by minimizing the distance in the image space between the backprojected refined 3D points and the manually selected tie points. However, exact image–LiDAR conjugate points often do not exist, and refined elevations are computed by LiDAR DEM interpolation. That makes the method highly dependent on the accuracy and density of the LiDAR point cloud. Except for [75], these methods only address the GNSS/INS–camera calibration problem without using LiDAR data. Therefore, the data consistency cannot be guaranteed in the case of a joint LiDAR–GNSS/INS–camera calibration. The method in [75] relies on prior DEM and the authors do not mention the calibration problem between a LiDAR and a GNSS/INS sensor.

2.4. Joint LiDAR–GNSS/INS–Camera Alignment

In this section, studies addressing the joint calibration of a multi-sensor composed at least of a camera, a LIDAR and a GNSS/INS sensor are reviewed.

The authors of [76] propose a motion-based (see section 2.1.3) calibration method for a multi-sensor array composed of a camera, a 3D LiDAR and an GNSS/INS for automotive mobile platforms. Each sensor trajectory is independently estimated using sensor-specific techniques and then aligned by successively computing the rotational and translational offsets between all sensors of the array. The LiDAR trajectory is estimated by using the ICP algorithm between successive scans, while the camera poses are computed using a standard visual odometry approach. The missing camera motion scale is estimated by incorporating additional sensor information from the array. In this method, all calibration parameters are estimated by considering only pairwise transformations between the sensors. As stated in [76], the result leads to a non-consistent solution. The authors performed an additional optimization step to find a consistent transformation. However, in this step the authors only optimized the rotation part of all estimated transformations because the camera transforms contained scale ambiguity.

In [11], the authors propose a calibration method for a multi-sensor system composed of multiple cameras, multiple LiDARs and a GNSS/INS sensor and achieved consistent calibration of all those sensors. The method is based on a modified bundle adjustment model, where image point scale factors are not eliminated but treated as unknowns. This allows the pairing of conjugate 3D image points and LiDAR derived linear and planar features extracted from the acquired data. The main limitation of this approach is that the method relies on the presence of high level features in the scene, which reduces the flexibility of the approach. Consequently, the extraction of high level features requires manual intervention, which means the method cannot be fully automatic (see section 2.1.4).

The authors of [12] present a hybrid orientation of LiDAR point clouds and aerial images designed to solve LiDAR strip adjustment and aerial triangulation in the same optimization framework. The method matches LiDAR strips and a photogrammetric reconstructed point cloud by optimizing several parameters, such as the absolute LiDAR and image data orientations, as well as the interior and mounting parameters (boresight angles and lever-arm) of the laser scanner and camera. Inspired by the ICP methodology, the method performs a 3D-3D alignment (see section 2.1.2) by iteratively minimizing discrepancies between defined sensor observation point-to-point and point-to-plane correspondences. The established correspondences are: strip-to-strip, control point-to-strip, image tie point-to-image tie points, image tie point-to-control point and image tie point-to-strip. The authors claim that selecting the appropriate image tie point-to-strip correspondences is not clear-cut as both sensors capture different features of the scene and exact correspondences might not be possible. This is mainly true when dealing with natural scenes where flat areas are absent.

2.5. Point Cloud Colorization

Once the LiDAR–GNSS/INS–camera alignment is properly performed, a 3D georeferenced LiDAR point cloud and the position and orientation of each image can be reconstructed in the same reference frame. Each 3D LiDAR point can then be projected in captured images. A pixel color can be associated with each LiDAR point falling in an image. For a set of M images, a 3D LiDAR point has m pixel color candidates, with $m \leq M$. When multiple overlapping images are available ($m \geq 2$), a colorization strategy has to be developed to assign a single color to each LiDAR point. Moreover, in order to avoid erroneous colorization, determining whether a LiDAR point is visible or occluded from a specific view point is crucial. We first review methods concerning the problem of point visibility in a point cloud. Then we present different proposed colorization strategies.

The Z-buffer method [77] is commonly used to detect occluded areas by utilizing a DSM derived visibility map, namely the buffer. The authors of [78] introduce two new angle-based methodologies for occlusion detection of a LiDAR derived DSM. Those are based on checking the off-nadir angle to the line of sight connecting the perspective center of the imaging sensor and the DSM cells. In [79], an alternative method is proposed that uses height gradients of a LiDAR derived DSM for occlusion detection. The method analyses the surface height gradient at certain sampled directions, guiding the identification of occluded regions in the aerial images. In [80, 81], the authors propose an efficient technique, i.e. Hidden Point Removal (HPR), which does not require any surface reconstruction or normal estimation. The method first transforms the points to a new domain and then constructs the convex hull in that domain. Points that lie on the convex hull of the transformed set of points are the images of the visible points.

In [82], the authors developed an efficient colorization strategy that colorize a LiDAR point cloud from a video stream. For each LiDAR point, a robust average color value of every pixel candidate is sequentially computed and updated once a color candidate is recorded. In this way, all color candidates for each LiDAR point do not have to be stored, thus avoiding an intractable problem. In addition, HPR is used to select visible LiDAR points from each image point of view. One colorization strategy also involves casting the colorization problem from multiple views to a single view colorization scheme. An option is to use the closest image to the considered LiDAR point, as accomplished in [83]. This method has the advantage of minimizing the impact of angular errors in the calibration or in the GNSS/INS measurements. Another approach is to use the image that has the closest view direction to the estimated point normal, as mentioned in [79]. A global colorization method for large-scale point clouds has also been developed in [84]. This method is based on an optimization framework that aims to assign the best color values to each LiDAR point according to a defined criterion. This approach first defines a graph-structure to the un-ordered set of 3D points. Secondly, an energy, composed of a data term and a smoothing term, is minimized to visually achieve pleasing point cloud colorization. The main drawback of this method is its computational cost. Moreover, as this technique does not entirely handle occlusions, artifacts can appear in the colorized point cloud.

3. Proposed Calibration

In this section, we propose a novel method that performs the LiDAR–GNSS/INS–camera calibration by successively determining T_s^g and T_g^c (see Fig. 1). The T_s^g and T_g^c estimations aim at respectively solving the LiDAR–GNSS/INS and GNSS/INS–camera data alignment. The steps involved in the proposed calibration method are depicted in Fig. 2. First, T_s^g is computed by a robust method inspired by a state-of-the-art approach [62]. Second, an estimation of T_g^c is performed based on the previous estimation of T_s^g, thus ensuring the data consistency of the entire multi-sensor system.

This work involves four coordinate systems: the reference mapping frame m, the GNSS/INS frame g whose time-dependent position and orientation are known in frame m, the laser scanner frame s and the camera frame c.

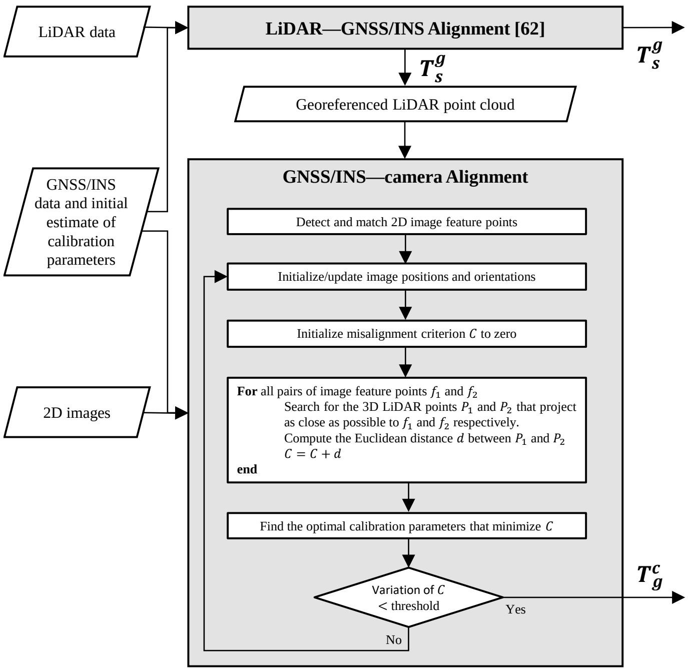


Figure 2: Flowchart of the proposed LiDAR–GNSS/INS–camera calibration method.

In the proposed method, estimations of both T_s^g and T_g^c transformations involve a prior coarse initialization based on the mounting configuration of the multi-sensor system. Unlike other dependence-based methods described in section 2.1.5, our approach is robust enough that no precise initialization is required as it is shown in section 6.4.

The LiDAR–GNSS/INS and the GNSS/INS–camera alignment described in the following sections are performed using data acquired by a UAV-borne multi-sensor system during a flight campaign. Details on the multi-sensor system used and the flight campaign carried out are given in section 4.

3.1. LiDAR–GNSS/INS Alignment

Boresight angles and the lever-arm between s and g are retrieved according to the methodology presented in [62]. Note that by using the method described in [62] we only estimate the boresight angles and the lever-arm offsets. Neither trajectory correction nor scanner intrinsic parameters are computed.

During the least squares adjustment, we fixed the vertical component of the lever-arm to a manually measured value in order to overcome the limitation presented in section 2.2.

3.2. GNSS/INS–Camera Alignment

3.2.1. Material

The estimated transformation T_s^g allows georeferencing in the global mapping frame m, a 3D point cloud from synchronized LiDAR and GNSS/INS measurements. Each image shot is time-tagged by the GNSS/INS and thus its precise position and orientation can be retrieved by interpolation. However, the georeferenced position and orientation of the camera frame c with respect to the m mapping frame are only approximately known due to the prior coarse initialization of T_g^c. Input data for our method consists of georeferenced point clouds and images derived from different flight lines collected by the UAV-borne multisensor system during a flight campaign (see section 4).

3.2.2. Basic Idea

Let us consider a pair of image points (f_1 and f_2) extracted respectively from two overlapping images (I_1 and I_2), and the two rays passing from the optical center of the camera through those image feature points (see Fig. 3). Let Q_1 denote the 3D LiDAR point that projects onto f_1 with Q_2 denoting the 3D LiDAR point that projects onto f_2. If f_1 is the conjugate point of f_2, then Q_1 and Q_2 should be identical. Therefore, the distance between Q_1 and Q_1 is a marker of the GNSS/INS–camera misalignment. However, Q_1 and Q_2 are most likely not measured. Nevertheless, an estimation of Q_1 and Q_2, denoted $\hat{Q}_1$ and $\hat{Q}_2$, has to be determined, e.g. by reconstructing a surface from measured LiDAR points. Usually this reconstruction is done by using interpolation methods. However, such a surface reconstruction, is computationally demanding and complex, often requiring additional information, such as normal estimations. In our method, we propose to simplify the estimation of $\hat{Q}_1$ and $\hat{Q}_2$ by using nearest neighbor interpolation. Let P_1 and P_2 be the 3D LiDAR points whose respective projections p_1 and p_2 in the images I_1 and I_2 are closest to f_1 and f_2. If no occlusion occurs, as in Fig. 3, then P_1 and P_2 are the nearest neighbor interpolation of Q_1 and Q_2, (i.e. $\hat{Q}_1$ and $\hat{Q}_2$). With this approach, any pair of conjugate image points can be associated with a pair of 3D LiDAR points. The above mentioned notations are illustrated in Fig. 3.

In case of a perfect alignment between the LiDAR and camera data, P_1 and P_2 must be identical (see Fig. 3d). In case of an imperfect LiDAR-camera alignment, P_1 and P_2 are most likely different (see Fig. 3c). The method proposed in this paper consists of expressing the distance between P_1 and P_2 as a function of the calibration parameters. Therefore, any pair of conjugate image points extracted from the image data may be considered to be able to estimate the calibration parameters by minimizing the sum of all squared distances between all pairs of associated 3D LiDAR points.

Our method, unlike that of [75], makes direct use of the LiDAR point cloud and does not rely on prior DSM to match 2D image feature points and 3D LiDAR points. We select 3D LiDAR points whose image projection is closest to the image feature points by performing a nearest neighbor search around each conjugate image feature point in the image space. Moreover, we propose a simple metric to evaluate the GNSS/INS–camera data misalignment by computing a sum of squared Euclidean distances between 3D LiDAR points. LiDAR point normals do not need to be estimated to compute a tie point-to-strip distance metric, as in [12], which is computationally more expensive than a point-to-point distance.

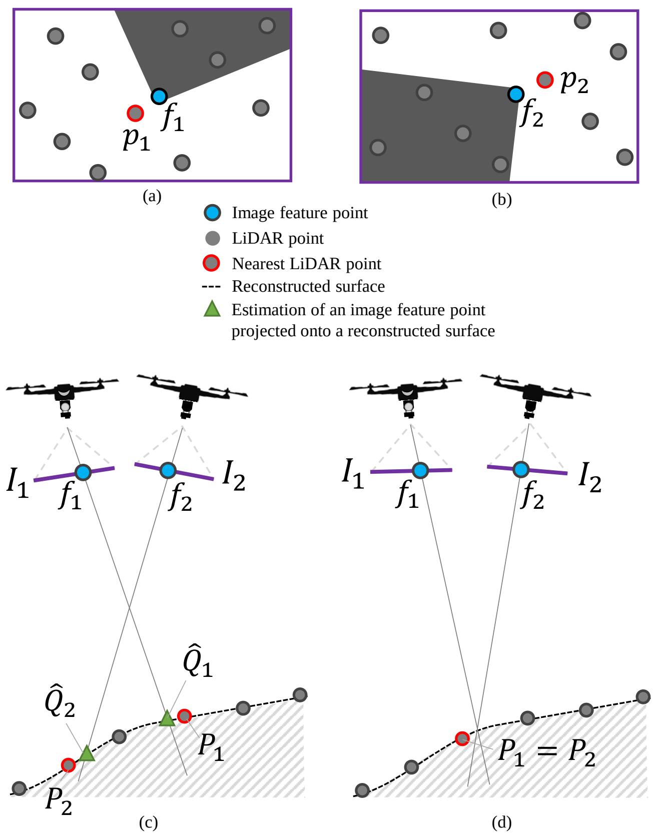


Figure 3: Basic idea of the GNSS/INS–camera alignment method. Figures (a) and (b) depict two images partially displaying the same scene taken at different positions, orientations and times. f_1 and f_2 is a pair of conjugate image feature points. p_1 is the projection of P_1 in the left image, while p_2 is the projection of P_2 in the right image. Figure (c) illustrates the case of imperfect GNSS/INS–camera alignment: P_1 and P_2 are different 3D LiDAR points. $\hat{Q}_1$ and $\hat{Q}_2$ are estimated projections of f_1 and f_2 onto the reconstructed surface. Figure (d) shows the perfect alignment case: P_1 and P_2 are the same 3D LiDAR point.

3.2.3. Implementation

Let $\varphi = [\theta_x, \theta_y, \theta_z, t_x, t_y, t_z]$ denote the calibration parameters to be estimated, with θ_x, θ_y, θ_z being the boresight angles and t_x, t_y, t_z the lever-arm components associated with the GNSS/INS–camera transformation T_g^c.

As a preprocessing step, we first extract image feature points from all the images considered for this calibration process. Then, we define a set of N pairs of image feature points by matching feature points from pairs of overlapping images. This can be achieved by using any feature point detector and descriptor. A comparative analysis of different image feature detectors and descriptors can be found in [85].

Let us consider a pair of image feature points (f_1, f_2) and its associated pair of nearest 3D LiDAR points (P_1, P_2). In order to find P_1, we first project every 3D LiDAR point on the image where f_1 is located. We then perform a nearest neighbor search in the image space around f_1 using a KD-tree [86]. P_1 is the point whose projection p_1 is the nearest to f_1 and is located in the image. P_2 is found similarly (see Fig.3a-b). In this way, we avoid performing an expensive nearest neighbor search in the 3D object space to find the nearest LiDAR point to the ray passing through f_1. Although performing a nearest neighbor search in the 2D image space or in the 3D object space is not equivalent, this is a reasonable approximation since images are usually located at high distances above the point cloud.

Let $d(P_1, P_2)$ be the squared Euclidean distance between P_1 and P_2. Expressing $d(P_1, P_2)$ as a function of φ gives:

$$d(P_1, P_2) = \\ \|T_g^m(t_2)T_c^g(\varphi)P_2^c(t_2) - T_g^m(t_1)T_c^g(\varphi)P_1^c(t_1)\|^2, \quad (1)$$

where $P_i^c(t_i)$ ($i \in \{1, 2\}$) is the coordinate of the 3D LiDAR point P_i in c the camera frame at time t_i, $T_g^m(t_i)$ ($i \in \{1, 2\}$), the transformation between the g and m frames at time t_i and $T_c^g(\varphi)$ the transformation between c and g that only depends on the φ calibration parameters.

Let us consider N pairs of conjugate image features points and d_k the squared distance between the two 3D LiDAR points associated with the k-th pair of conjugate image feature points. We determine a GNSS/INS–camera alignment criterion C which we compute as follows:

$$C(\varphi) = \sum_{k}^{N} d_k \quad (2)$$

As this criterion is a non-linear function of φ, we use a non-linear least square algorithm to find $\varphi^\star$, the global minimum of C. In addition, a Huber loss function [87] is used to minimize the impact of outliers on this calibration. Note that at each iteration i in the non-linear least square algorithm, the current value φ^i is used to update the nearest 3D LiDAR points of each pair of image feature points. The optimization stops when the variation of C between two iterations is below a specific threshold.

Note that during the optimization, we set the vertical component t_z of the lever-arm to a manually measured value in order to overcome the limitation presented in section 2.3.

4. Study Site and Datasets

This section briefly addresses the acquired data used in this paper for the proposed LiDAR–GNSS/INS–camera calibration (section 3) and its evaluation (section 5).

4.1. Actual Data

Two flight campaigns (F_1 and F_2) were conducted over a flight area next to the city of Montpellier in the South of France. Data acquired during F_1 is used to perform the LiDAR–GNSS/INS–camera calibration, while data acquired during F_2 is used to evaluate the performed calibration. During both campaigns, camera, LiDAR and GNSS/INS data were simultaneously captured by the same multi-sensor system.

The multi-sensor system is mounted on an DJI-M600 multicopter. It is composed of a YellowScan Surveyor LiDAR system and a rigidly mounted SONY UMC-R10C camera.

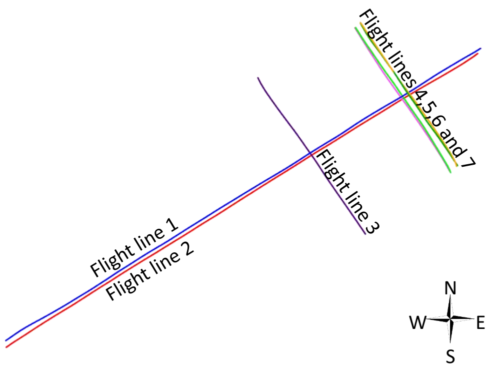
The YellowScan Surveyor is composed of a Velodyne VLP-16 laser scanner and an APX-15 GNSS/INS. The manufacturer claims that the Velodyne VLP-16 has a typical range precision of 3 cm. The APX-15 has 0.025° post-processed roll and pitch precision and 0.08° heading. The positioning accuracy given is 2 cm horizontal and 5 cm vertical. Therefore, the manufacturer YellowScan guarantees 5 cm absolute accuracy for a georeferenced LiDAR point cloud resulting from the Surveyor measurements taken below 50 m altitude.

The camera CMOS sensor size is 23.2×15.4 mm with an array dimension of 5456×3632 pixels (~ 20 megapixels). The camera is mounted with a 9 mm lens, and offers a 113° horizontal field of view.

LiDAR and photogrammetric targets were spread on horizontal and tilted surfaces at different elevations to be captured during the F_2 flight campaign (see Fig. 4). LiDAR targets are high reflective planar surfaces, and photogrammetric targets are planar objects marked with a checkerboard pattern. LiDAR and photogrammetric targets were surveyed using a combination of total station and differential GNSS measurements, resulting in 53 LiDAR checkpoints and 120 image checkpoints. Both checkpoints have an absolute accuracy of 1 cm or less. LiDAR checkpoints were measured at the center of the LiDAR targets or on the ground, where it is known to be approximately flat. Image checkpoints were measured at each corner of the photogrammetric targets. Examples of LiDAR and image checkpoints are illustrated in Fig. 5.

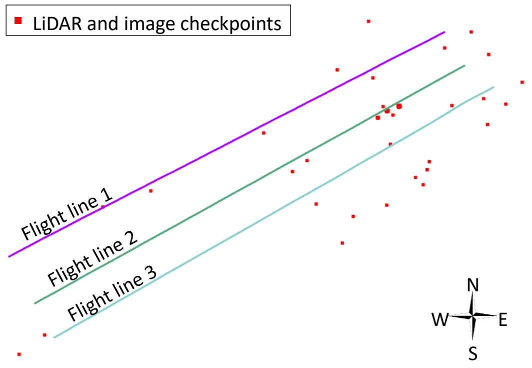
From the F_1 flight campaign, we extracted an input dataset consisting of camera, LiDAR and GNSS/INS data from parallel and perpendicular flight lines. Those flight lines fulfill the flight configuration recommended in [60, 61] for reliable LiDAR–GNSS/INS alignment and the flight configuration recommended in [74] for reliable estimation of the GNSS/INS–camera calibration parameters. The flight lines used as input data are illustrated in Fig. 4a. This includes 20 images and 7 LiDAR strips consisting of an overall point cloud of approximately 10 M points.

From the F_2 flight campaign, we extracted a test dataset consisting of camera, LiDAR and GNSS/INS data from the 3 parallel flight lines depicted in Fig. 4b. This includes an overall LiDAR point cloud of approximately 15 M LiDAR points and 100 images.



Flight lines	Altitude	Direction
1	25 m	NE—SW
2	45 m	SW—NE
3	25 m	NW—SE
4	25 m	SE—NW
5	25 m	NW—SE
6	45 m	SE—NW
7	45 m	NW—SE

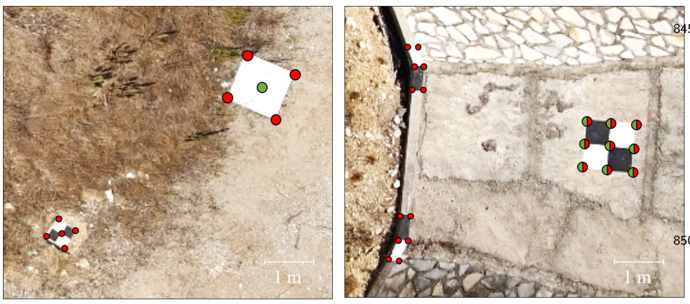
(a)



Flight lines	Altitude	Direction
1	45 m	SW—NE
2	65 m	SW—NE
3	45 m	NE—SW

(b)

Figure 4: Figure (a) illustrates the flight lines of the input dataset extracted from F_1. The input dataset includes 3 flight lines at 45 m altitude and 4 flight lines at 25 m altitude. Figure (b) depicts the flight lines of the test data extracted from F_2. The test dataset includes 2 flight lines at 45 m altitude and one flight line at 65 m altitude. The red dots represent the positions of all image and LiDAR checkpoints.



(a) (b)

Figure 5: LiDAR and image checkpoints. The green dots describe LiDAR checkpoints and the red dots represent image checkpoints. Dots with both red and green are used simultaneously as LiDAR and image checkpoints.

4.2. Simulated Data

The simulated data consists of 25 point clouds and 20 images. All data are positioned and oriented in a global reference frame. Simulated LiDAR point clouds are generated by sampling an artificially created planar DSM. The images are generated by simulating a camera with a similar focal length, sensor size and array dimension as the camera presented in section 4.1. Simulated images are positioned and oriented according the same flight configuration mentioned in section 4.1. In the proposed calibration method, since the image color information (pixel color) is only used to establish correspondences between image feature points, we directly simulate exact conjugate image feature points instead of creating image color information. We thus generate on the planar DSM, artificial markers (points) which we then project on each image. A marker that projects itself on two simulated images creates an artificial pair of image feature points.

5. Evaluation Methods

In this section, we assess our algorithm performance on both actual and simulated data. Using actual and ground truth data, we quantitatively assess the performance of both LiDAR–GNSS/INS and GNSS/INS–camera alignment. The LiDAR–camera alignment is also evaluated qualitatively by visually examining the point cloud colorization. Using simulated data, we study the sensitivity of the GNSS/INS–camera alignment method to the initial calibration parameters and to the point cloud density and noise.

The proposed LiDAR–GNSS/INS–camera calibration is performed using data from F_1 and is assessed using data from F_2. LiDAR checkpoints are used to assess the LiDAR–GNSS/INS alignment. Image checkpoints are used to evaluate the GNSS/INS–camera alignment (see section 4.1).

5.1. LiDAR–GNSS/INS Alignment Evaluation

In order to assess the LiDAR–GNSS/INS alignment, in the point cloud from the test data, we select LiDAR points that we assume to be locally close to every LiDAR checkpoint. We thus consider the points contained in the sphere of 20 cm radius which is centered at the considered LiDAR checkpoint. It should be recalled that LiDAR checkpoints are measured on locally planar surfaces. Let LiDAR checkpoints l_1 and Ω_1 denote the set of LiDAR points selected around l_1. In case of accurate LiDAR–GNSS/INS alignment, then all LiDAR points in Ω_1 are located on the real planar surface around l_1. Consequently, the distance $D(l_1, \Omega_1)$, which we define as the orthogonal distance between l_1 and the plane that best fits Ω_1, should be compatible with the precision of the acquisition system, i.e. with the laser scanner and GNSS/INS. However, in case of incorrect estimation of the LiDAR–GNSS/INS calibration parameters, Ω_1 is most likely located far from the true neighborhood of l_1. Therefore, by considering all 53 LiDAR GCPs measured in the flight area, we propose two metrics D_L and σ_L to evaluate the LiDAR–GNSS/INS alignment. While D_L assesses the absolute accuracy of the alignment, σ_L assesses its relative accuracy. We define D_L as the mean of all $D(l_j, \Omega_j)$ computed for each LiDAR checkpoint j. We define σ_L as the mean of all σ_j, where σ_j is the standard deviation of the distribution of the point set Ω_j around the best plane that fits Ω_j. Since LiDAR checkpoints were measured on horizontal and tilted surfaces at different elevations in the flight area, both metrics, D_L and σ_L, are likely symptomatic of the LiDAR–GNSS/INS data misalignment.

5.2. GNSS/INS–Camera Alignment Evaluation

We compute the back-projection error in the image space to assess the accuracy of the GNSS/INS–camera calibration. We select, in the F_2 flight campaign, 9 images taken at 45 m altitude that contains several photogrammetric targets. The corners of each photogrammetric target are detected in each image. Each corner can be associated with a 3D image checkpoint. The distance between the projection of each image checkpoint and the corner it is associated with is a marker of the accuracy of the GNSS/INS–camera alignment. Fig. 6 illustrates the detected corners (in blue) and the projected image GCPs in the image space (in red). The root mean square error (RMSE) is used to assess the alignment accuracy.

5.3. LiDAR–Camera Alignment Evaluation

We assess the LiDAR–camera alignment which was performed indirectly by achieving the GNSS/INS–camera and the LiDAR–GNSS/INS calibration. We perform a visual evaluation by colorizing the point cloud using LiDAR and the camera data from the test dataset. To this end, we use a state-of-the-art method to colorize the point cloud by choosing, for each LiDAR point, the pixel value from the spatially closest image, as in [79]. As an extension of our colorization method, we use of a Z-buffer method [77] to detect occluded areas.

5.4. Sensitivity Analysis to Initial Calibration Parameters

In order to study the influence of the initial calibration parameters in the proposed GNSS/INS–camera calibration method, we performed experiments on simulated and actual data.

We repeated 100 calibrations using simulated data starting with different initial calibration offsets. The initial boresight angles (θ_x, θ_y, θ_z) and planimetric lever-arm offsets (t_x, t_y) were randomly set to values respectively comprised in the intervals $[-10, 10]°$ and $[-1, 1]$ m centered at the parameters true values. The vertical lever-arm offset t_z was set at its true value during the optimization to avoid a biased estimation, as mentioned in section 2.3.

The experiment on actual data involved repeating the calibration by varying the approximate initial calibration values obtained by the sensor mounting configuration. For this experiment, we used initial boresight angles and planimetric lever-arm offsets, ranging respectively from $-10°$ to $10°$ and from -1 m to 1 m with a 1° and 10 cm step. Note that a reasonable approximate measurement of the sensor mounting configuration usually enables the operator to estimate the initial parameters with an accuracy of up to a few degrees and centimeters.

5.5. Sensitivity Analysis to Point Cloud Density and Noise using Simulated Data

In this section, we study the sensitivity of the GNSS/INS–camera alignment method to point cloud density and noise. Simulated data are used to perform a quantitative analysis in a controlled environment. We conducted 25 experiments, one for each of the 25 generated point clouds. Each experiment consists of performing a GNSS/INS–camera alignment (see section 3.2.3) using the simulated images and one point cloud.

Each of the 25 point clouds are generated from a horizontal planar DSM with 5 different noise level and 5 different density characteristics. The 5 densities are 100, 10, 1, 0.1, and 0.01 pt/m^2. Noise is modeled by a centered random error with a normal distribution which is applied to the position of each LiDAR point. Five different noise levels are investigated that have different standard deviations: 0, 0.01, 0.1, 0.5, and 2 m. Previous experiments have shown that point cloud densities which are too low prevent joint estimation of both boresight angles and lever-arm offsets. By using densities below 1 pt/m^2, the optimized planimetric lever-arm components do not converge but instead oscillate around their optimal values. Therefore, we divide the 25 point cloud set into two subsets: S_1 and S_2. S_1 consists of the 15 point clouds having densities greater or equal to 1 pt/m^2 which are used to jointly estimate both boresight angles and lever-arm offsets. S_2 includes the remaining 10 point clouds which are only used to estimate the boresight angles.

Using the point clouds of the S_1 subset, we simultaneously optimize the boresight angles and the lever-arm. The boresight angles (θ_x, θ_y, θ_z) and the planimetric lever-arm offsets (t_x, t_y) are initialized far from their true values, while the vertical lever-arm offset t_z is set at its true value during the optimization to avoid a biased estimation (see section 2.3). Note that the initial calibration parameter values do not influence the resulting optimal parameters as long as nearest LiDAR points to the conjugate image feature points can be found (see section 3). We only optimize the boresight angles using the point clouds of the S_2 subset. The boresight angles (θ_x, θ_y, θ_z) are initialized far from their true values, while the lever-arm offsets (t_x, t_y, t_z) are set at their true values.

Each experiment is run until convergence (see section 3). In order to compare the results between experiments, we compute two error values: the rotational $\Delta\theta$ and the translation Δt differences between the resulting rigid-body transformation and the ground truth transformation. $\Delta\theta$ and Δt are computed as follows:

$$\Delta\theta = \arccos\left(\frac{tr(\Delta R) - 1}{2}\right), \quad (3)$$

and

$$\Delta t = \|\Delta T\|, \quad (4)$$

where ΔR is the difference rotation matrix between the two 3D rotations given, respectively, by the estimated boresight angles and their true values and ΔT is the difference translation vector between the two 3D translations given, respectively, by the estimated lever-arm and its true value. Moreover, to ensure that the results are representative, we repeat each experiment 75 times by bootstrapping each input point cloud, thus the error is computed as the mean μ of the 75 bootstrapped values.

6. Evaluation Results

6.1. LiDAR–GNSS/INS Alignment Evaluation

D_L is equal to 2.45 cm and σ_L to 2.81 cm according to the experiment described in section 5.1. This result indicates that both D_L and σ_L, which are related to the error in the LiDAR–GNSS/INS alignment, are indiscernible from the noise caused by the 5 cm accuracy of the YellowScan Surveyor provided by the manufacturer. This indicates a proper LiDAR–GNSS/INS alignment with respect to the utilized sensor accuracy.

6.2. GNSS/INS–Camera Alignment Evaluation

The RMSE is equal to 2.55 pixels according to the experiment described in section 5.2. At 45 m altitude, the approximate ground sampling distance (GSD) of the test images is 2.1 cm/pixel. The checkpoints have a maximum possible measurement error of 1 cm in the mapping frame, so this maximum error projected in the test images is about 0.5 pixel. In addition, the corners are detected with a maximum possible error of 1 pixel. Therefore, the RMSE is 2.55 ± 1.5 pixels. This error is therefore compatible with the YellowScan Surveyor LiDAR measurement accuracy of 5 cm, which is equal to 2.38 pixels in the image space at 45 m altitude.

6.3. LiDAR–Camera Alignment Evaluation

According to the experiment described in section 5.3, the resulting point cloud colorization is illustrated in Fig. 7. The image and LiDAR checkpoint locations are displayed in red. We observe on the checkerboard targets that the colors seem to be correctly assigned the each LiDAR point. Correct color assignment can also be observed on the 3D structures, e.g. on the white truck in frame E in Fig. 7. No miscolorization is visible, e.g. white color on the ground or the ground color on the truck. These results indicate that our proposed alignment approach could possibly also generate an accurate colored representation of the LiDAR point cloud without explicitly estimating the LiDAR–camera calibration parameters.

6.4. Sensitivity Analysis to Initial Calibration Parameters

All calibrations performed on simulated data converged to a solution close to the true calibration parameters. The RMSE between final and true values for θ_x, θ_y, θ_z, t_x, and t_y are given on the left side of Table 1. Since at 45 m altitude and at nadir a 1e−3° angular error in the calibration parameters produces a misalignment error in the object space smaller than 1 mm, these RMSE are considered small enough to indicate the robustness of our proposed calibration to parameter initialization.

In the experiment on actual data, the standard deviation of the final calibration parameters θ_x, θ_y, θ_z, t_x, and t_y are given on the right side of Table 1. Even though the true calibration values are not known, these results show that the dispersion around the estimated values is low and independent of the parameter initialization, even when actual data is used. The differences in magnitude with the standard deviation obtained using simulated data is due to the fact that the actual LiDAR point cloud has a point density approximately 200 times greater than the simulated one. Further results on the influence of the point cloud density are given in section 6.5.

6.5. Sensitivity Analysis to the Point Cloud Density and Noise using Simulated Data

The results are summarized in Table 2. A double vertical line separates the experiments conducted using the S_1 and S_2 point cloud subsets. On the left, experiments using S_1 are presented, while the right side shows the experiment conducted using S_2.

By looking at the experiments performed on S_1 on the left side of Table 2 (from 100 to 1 pt/m^2), we notice that the calibration error increases as the noise level increases. The alignment error also slightly increases as the point cloud density decreases, but the influence of the density level is lower than the influence of the noise level. This highlights the robustness of the algorithm to density variation. Moreover, error values are reasonably small for noise levels up to 0.1 m, with a respective maximal rotational and translational error of 0.0106° and 0.0131 m

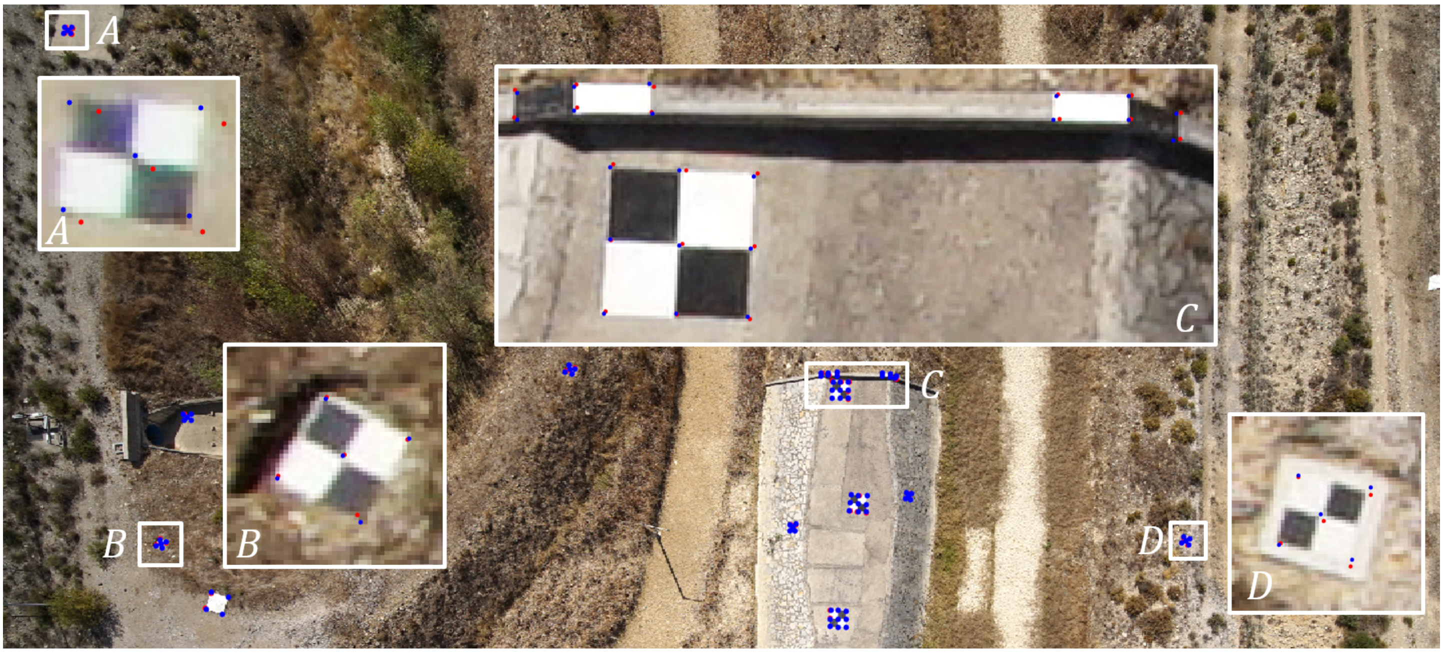


Figure 6: Evaluation of the GNSS/INS–camera alignment. The blue dots depict the detected corners. The red dots represent the projection of the image checkpoints on the image. The root mean square distance between corresponding blue and red dots is computed to assess the GNSS/INS–camera calibration accuracy.

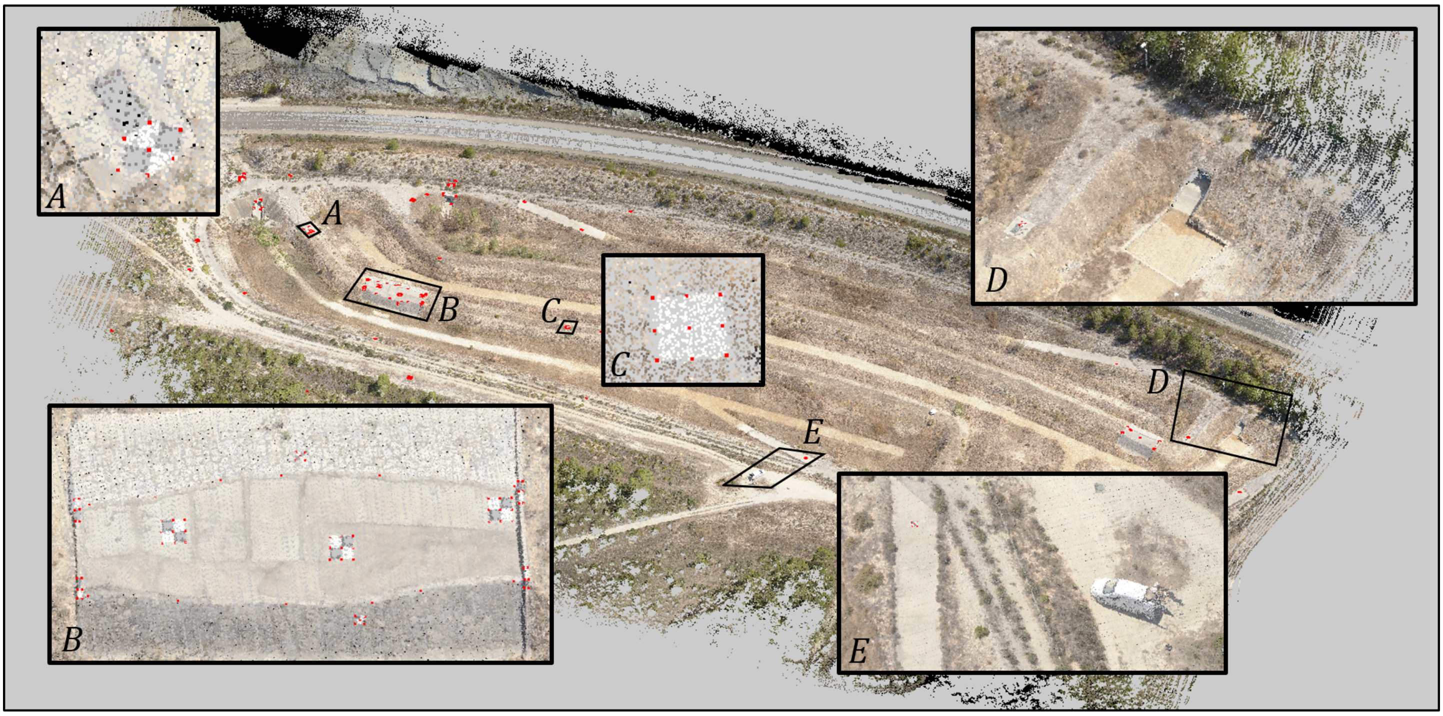


Figure 7: Visual evaluation of LiDAR–camera alignment by colorizing the LiDAR point cloud using multiple images. LiDAR and image checkpoints are shown in red.

for a density of 1 pt/m^2. Generally, both rotational error and translational error increase with increasing noise level.

The results on the right side of Table 2 (from 0.1 to 0.01 pt/m^2) show a high error increase with decreasing point density. Moreover, the errors at 0.1 pt/m^2 are mainly larger for higher noise levels, but for a density of 0.01 pt/m^2, while the noise level no longer seems to have a significant role in the resulting error.

Table 1: Dispersion of the estimated calibration parameters according to different initial calibration offsets. SD, MAE and MAD stand for standard deviation, mean absolute error, and mean absolute deviation respectively.

	Simulated data						Actual data			
	true value	**RMSE**	**SD**	MAE	min	max	**SD**	MAD	min	max
θ_x [°]	0	**3.1e−3**	**1.7e−4**	3.1e−3	2.9e−3	3.4e−3	**4.0e−5**	1.9e−5	−1.7e−4	3.4e−5
θ_y [°]	0	**1.6e−4**	**1.1e−4**	1.4e−4	−3.2e−4	1.3e−4	**1.9e−5**	1.4e−5	−2.8e−5	5.0e−5
θ_z [°]	0	**9.2e−4**	**2.1e−4**	9.0e−4	−1.4e−3	−5.6e−4	**2.5e−5**	1.5e−5	−2.9e−5	7.9e−5
t_x [m]	0	**1.9e−3**	**1.5e−4**	1.9e−3	1.5e−3	2.1e−3	**4.4e−5**	1.8e−5	−9.5e−6	1.9e−4
t_y [m]	0	**3.8e−3**	**2.2e−4**	3.8e−3	−4.1e−3	−3.4e−3	**1.1e−5**	4.5e−6	−4.8e−5	2.4e−6

Table 2: Calibration parameter errors according to the point cloud density (D) and noise (N).

	D [pt/m^2]	100	10	1	0.1	0.01
N [m]		μ	μ	μ	μ	μ
0	$\Delta\theta$ [°]	0.0017	0.0016	0.0056	0.0237	0.1750
	Δt [m]	0.0017	0.0028	0.0078	-	-
0.01	$\Delta\theta$ [°]	0.0021	0.0029	0.0065	0.0227	0.1914
	Δt [m]	0.0022	0.0032	0.0064	-	-
0.1	$\Delta\theta$ [°]	0.0068	0.0075	0.0106	0.0258	0.1800
	Δt [m]	0.0106	0.0125	0.0131	-	-
0.5	$\Delta\theta$ [°]	0.0173	0.0185	0.0244	0.0350	0.1800
	Δt [m]	0.0287	0.0294	0.0378	-	-
2	$\Delta\theta$ [°]	0.0368	0.0396	0.0415	0.0420	0.1943
	Δt [m]	0.0664	0.0662	0.0576	-	-

7. Discussion

The origins of persistent camera and LiDAR data misalignment errors after system calibration are usually hard to identify, as the errors can be caused by a combination of several factors, e.g. camera and LiDAR time-synchronization, the GNSS/INS measurement quality, errors in the LiDAR–GNSS/INS calibration, altitude at which the calibration data are acquired, camera intrinsic parameters, occlusions, and image feature point detection. The contribution of individual error sources is hard to determine based only on ground truth. Therefore, a thorough analysis using simulated data could help to understand this contribution on the final alignment result. We discuss the mentioned error sources in this section.

First, the GNSS/INS–camera alignment method markedly depends on the presence and the quality of the image feature points. The overflown area must be sufficiently textured so that evenly distributed image feature points can be automatically detected in images. Consequently, areas with homogeneous colors should be avoided or at least should not prevail in the data. Otherwise the algorithm is prone to failure. By our approach, a few wrong image feature correspondences do not impact the alignment thanks to the use of the Huber loss function in the optimization process. However, the method will likely fail if the number of wrong matches exceed our estimator's breakdown point. However, a high number of correspondences, well-distributed in the images, will positively impact the alignment quality. Therefore, the feature detector and descriptor has to be carefully selected. The GSD also plays a significant role in the alignment quality, since the image feature location is less precise as the GSD increases. Consequently, lower image resolution and high flying altitudes can negatively impact the final alignment quality.

Second, point cloud areas which may be occluded in the real world but are visible from a specific image point of view can be a limitation in the proposed GNSS/INS–camera alignment method. This phenomenon is likely to arise with the presence of high vertical objects like buildings or trees. For every image feature, the computation of nearest LiDAR points is performed in the image space, so occlusions are prone to cause erroneous selection of non-visible LiDAR points. Although the Huber loss function reduces the impact of false selected nearest LiDAR points in the minimization process, to much occlusion would cause the algorithm to fail, as the computed cost would no longer be representative of the GNSS/INS–camera data misalignment. In this case, visible LiDAR points from each image view point should be determined before selecting the nearest LiDAR points.

Third, the effect of the used camera model and the estimation of its parameters on the GNSS/INS–camera alignment accuracy should be investigated. Indeed, the mapping-to-image frame projection function involves distortion modeling. Regarding the photogrammetric target A in Fig. 6, we note that the data alignment is worse than on targets located in the middle of the image. This is certainly due to the fact that the camera distortion model or its estimation has some weaknesses in this part of the image.

Fourth, the flight pattern is crucial for the algorithm to work properly. In [74], the authors mention having difficulty in decorrelating the translational and rotational calibration parameters using different flight line altitudes. However, no indication is given about optimal altitudes or altitude variations that should be considered. We noticed that the algorithm converged better at lower altitudes. First because, as previously mentioned, the quality of detected image feature will improve as the GSD decreases. Moreover, the LiDAR point density will also increase along with the number of image feature points. As we have shown with simulated data, the algorithm performs better with denser input data. Consequently, if less data is available, e.g. with low point cloud densities, the algorithm convergence is no longer ensured. However, due to the nature of the rotational parameters, small angular calibration errors obtained with data collected at low altitude will highly impact the data alignment at higher altitudes. Therefore, calibration should be performed close to the operating altitudes if possible. In addition, the reader might have noticed that in the results on simulated data described in section 6.4, the RMSE for θ_x (3.1e−3°) is about 20 times larger than the RMSE for θ_y (1.6e−4°). This problem is due the non-symmetric flight pattern recommended in [74] with respect to the roll axis caused by flight line 3 in F_1 (see Fig. 4a). This problem can, however, be solved by adding to the flight pattern in Fig. 4a an extra flight line at 25 m altitude in SE–NW direction such that it is symmetrical to flight line 3 with respect to the orientation given by the flight lines 4–7. By carrying out this experiment with this new flight configuration, we obtain a 1.3e−4° RMSE for θ_x, which is within the same order of magnitude as the RMSE of the other angles.

Fifth, the estimation of the GNSS/INS–camera calibration parameters is purposely based on the quality of the georeferenced point cloud in order to preserve the data consistency. This has the advantage of offsetting small errors in the LiDAR–GNSS/INS calibration estimation, while ensuring the consistency of the LiDAR–camera alignment. However, larger errors in the LiDAR–GNSS/INS alignment can potentially bias the whole multi-sensor system calibration, therefore resulting in an incorrect point cloud colorization.

Finally, the main limitation of our multi-sensor calibration method is its strong dependency on having sufficiently accurate position and orientation information at the time of every camera and LiDAR measurement. This means that accurate GNSS/INS measurements must be available and precise and robust time-synchronization is required between GNSS/INS measurements and camera and LiDAR acquisitions. Indeed, both LiDAR–GNSS/INS and GNSS/INS–camera alignment methods rely on GNSS/INS observations and high errors in the multi-sensor position and orientation will directly impact the resulting alignment quality.

8. Conclusion

In this paper we presented a new alignment method for a multi-sensor composed of a camera, LiDAR and GNSS/INS. The method is fully automatic, does not require any calibration markers, LiDAR intensity data or precise initialization of the calibration parameters. Moreover, our approach is not based on exact image–LiDAR conjugate features so the method is also suited for natural environments. Our method successively executes the LiDAR–GNSS/INS and the GNSS/INS–camera alignment in order to preserve the data consistency. We show in an experiment using actual data that this approach is suitable for performing accurate LiDAR–camera alignment. Quantitative metrics are applied to evaluate LiDAR–GNSS/INS and GNSS/INS–camera alignments using checkpoints. The results indicate that the obtained calibration accuracy is compatible with the 5 cm accuracy of the georeferenced LiDAR point cloud given by the manufacturer. This result is promising as the proposed system calibration relies on a georeferenced LiDAR point cloud and thus on its accuracy. Simulations on synthetic data show the robustness of our method to initial calibration parameters, low LiDAR point cloud density and noise levels. Moreover, we have shown that the accuracy of the colorized point cloud is appropriate for the intended application when system calibration is performed close to the operating altitudes. There is still room for improving the accuracy of the system calibration (see section 6.2) due to the different limitations discussed in section 7. If the contribution of the vertical lever-arm error in the total error budget becomes significant, then at least one GCP is required.

The advantage of our method is its flexibility. Our approach allows us to conduct any acquisition campaign with a quick system calibration process at its operating altitude.

We currently use a classical approach for colorizing the LiDAR point cloud, by assigning a color to each LiDAR point using a single pixel. If the calibration is accurate enough, several candidate pixels can be selected for each LiDAR point. This information could be used to increase the accuracy of the assignment of each pixel to each LiDAR point or to help equalize the overall color on the point cloud.

References

[1] A. Jaakkola, J. Hyyppä, A. Kukko, X. Yu, H. Kaartinen, M. Lehtomäki, Y. Lin, A low-cost multi-sensoral mobile mapping system and its feasibility for tree measurements, ISPRS journal of Photogrammetry and Remote Sensing 65 (6) (2010) 514–522.

[2] Q. Li, X. Cheng, Comparison of different feature sets for tls point cloud classification, Sensors 18 (12) (2018) 4206.

[3] Z. Taylor, J. Nieto, D. Johnson, Automatic calibration of multi-modal sensor systems using a gradient orientation measure, in: IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), IEEE, 2013, pp. 1293–1300.

[4] G. Pandey, J. R. McBride, S. Savarese, R. M. Eustice, Automatic extrinsic calibration of vision and lidar by maximizing mutual information, Journal of Field Robotics 32 (5) (2015) 696–722.

[5] E. G. Parmehr, C. S. Fraser, C. Zhang, J. Leach, Automatic registration of optical imagery with 3d lidar data using statistical similarity, ISPRS Journal of Photogrammetry and Remote Sensing 88 (2014) 28–40.

[6] F. Lv, K. Ren, Automatic registration of airborne lidar point cloud data and optical imagery depth map based on line and points features, Infrared Physics & Technology 71 (2015) 457–463.

[7] B. Yang, C. Chen, Automatic registration of uav-borne sequent images and lidar data, ISPRS Journal of Photogrammetry and Remote Sensing 101 (2015) 262–274.

[8] B. Khaleghi, A. Khamis, F. O. Karray, S. N. Razavi, Multisensor data fusion: A review of the state-of-the-art, Information fusion 14 (1) (2013) 28–44.

[9] J. Zhang, X. Lin, Advances in fusion of optical imagery and lidar point cloud applied to photogrammetry and remote sensing, International Journal of Image and Data Fusion 8 (1) (2017) 1–31.

[10] J. Li, B. Yang, C. Chen, A. Habib, 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 (2019) 123–145.

[11] R. Ravi, Y.-J. Lin, M. Elbahnasawy, T. Shamseldin, A. Habib, Simultaneous system calibration of a multi-lidar multicamera mobile mapping platform, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 11 (5) (2018) 1694–1714.

[12] P. Glira, N. Pfeifer, G. Mandlburger, Hybrid orientation of airborne lidar point clouds and aerial images, ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences (2019) 567–574.

[13] S. Wasielewski, O. Strauss, Calibration of a multi-sensor system laser rangefinder/camera, in: IEEE Intelligent Vehicles Symposium (IV), IEEE, 1995, pp. 472–477.

[14] Q. Zhang, R. Pless, Extrinsic calibration of a camera and laser range finder (improves camera calibration), in: IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Vol. 3, IEEE, 2004, pp. 2301–2306.

[15] R. Unnikrishnan, M. Hebert, Fast extrinsic calibration of a laser rangefinder to a camera, Robotics (2005).

[16] A. Kassir, T. Peynot, Reliable automatic camera-laser calibration, in: Australasian Conference on Robotics & Automation, ARAA, 2010.

[17] P. Núñez, P. Drews Jr, R. P. Rocha, J. Dias, Data fusion calibration for a 3d laser range finder and a camera using inertial data., in: European Conference on Mobile Robots (ECMR), 2009, pp. 31–36.

[18] A. Geiger, F. Moosmann, Ö. Car, B. Schuster, Automatic camera and range sensor calibration using a single shot, in: IEEE International Conference on Robotics and Automation (ICRA), IEEE, 2012, pp. 3936–3943.

[19] L. Zhou, Z. Li, M. Kaess, Automatic extrinsic calibration of a camera and a 3d lidar using line and plane correspondences, in: IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), IEEE, 2018, pp. 5562–5569.

[20] V. Fremont, P. Bonnifait, et al., Extrinsic calibration between a multi-layer lidar and a camera, in: IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems (MFI), IEEE, 2008, pp. 214–219.

[21] M. Pereira, D. Silva, V. Santos, P. Dias, Self calibration of multiple lidars and cameras on autonomous vehicles, Robotics and Autonomous Systems 83 (2016) 326–337.

[22] Z. Hu, Y. Li, N. Li, B. Zhao, Extrinsic calibration of 2-d laser rangefinder and camera from single shot based on minimal solution, IEEE Transactions on Instrumentation and Measurement 65 (4) (2016) 915–929.

[23] Y. Xie, R. Shao, P. Guli, B. Li, L. Wang, Infrastructure based calibration of a multi-camera and multi-lidar system using apriltags, in: IEEE Intelligent Vehicles Symposium (IV), IEEE, 2018, pp. 605–610.

[24] B. Triggs, P. F. McLauchlan, R. I. Hartley, A. W. Fitzgibbon, Bundle adjustment—a modern synthesis (1999) 298–372.

[25] Y. Chen, G. Medioni, Object modelling by registration of multiple range images, Image and Vision Computing 10 (3) (1992) 145–155.

[26] P. J. Besl, N. D. McKay, Method for registration of 3-d shapes, in: Sensor fusion IV: Control Paradigms and Data Structures, Vol. 1611, International Society for Optics and Photonics, 1992, pp. 586–606.

[27] S. Rusinkiewicz, M. Levoy, Efficient variants of the icp algorithm., in: International Conference on 3-D Digital Imaging and Modeling (3DIM), Vol. 1, 2001, pp. 145–152.

[28] W. Zhao, D. Nister, S. Hsu, et al., Alignment of continuous video onto 3d point clouds, IEEE Transactions on Pattern Analysis and Machine Intelligence (8) (2005) 1305–1306.

[29] R. Y. Tsai, R. K. Lenz, A new technique for fully autonomous and efficient 3d robotics hand/eye calibration, IEEE Transactions on Robotics and Automation 5 (3) (1989) 345–358.

[30] Y. C. Shiu, S. Ahmad, Calibration of wrist-mounted robotic sensors by solving homogeneous transform equations of the form ax= xb, IEEE Transactions on Robotics and Automation 5 (1) (1989) 16–29.

[31] C.-C. Wang, Extrinsic calibration of a vision sensor mounted on a robot, IEEE Transactions on Robotics and Automation 8 (2) (1992) 161–175.

[32] J. Schmidt, F. Vogt, H. Niemann, Calibration–free hand–eye calibration: a structure–from–motion approach, in: Joint Pattern Recognition Symposium, Springer, 2005, pp. 67–74.

[33] N. Andreff, R. Horaud, B. Espiau, Robot hand-eye calibration using structure-from-motion, The International Journal of Robotics Research 20 (3) (2001) 228–248.

[34] J. Heller, M. Havlena, A. Sugimoto, T. Pajdla, Structure-from-motion based hand-eye calibration using l• minimization, in: IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, 2011, pp. 3497–3503.

[35] S. Schneider, T. Luettel, H.-J. Wuensche, Odometry-based online extrinsic sensor calibration, in: IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), IEEE, 2013, pp. 1287–1292.

[36] R. Ishikawa, T. Oishi, K. Ikeuchi, Lidar and camera calibration using motions estimated by sensor fusion odometry, in: IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), IEEE, 2018, pp. 7342–7349.

[37] P. Rönnholm, Registration quality-towards integration of laser scanning and photogrammetry, EuroSDR, 2011.

[38] D. González-Aguilera, P. Rodríguez-Gonzálvez, D. Hernández-López, J. L. Lerma, A robust and hierarchical approach for the automatic co-registration of intensity and visible images, Optics & Laser Technology 44 (6) (2012) 1915–1923.

[39] M. Gong, S. Zhao, L. Jiao, D. Tian, S. Wang, A novel coarse-to-fine scheme for automatic image registration based on sift and mutual information, IEEE Transactions on Geoscience and Remote Sensing 52 (7) (2013) 4328–4338.

[40] H. Kim, C. D. Correa, N. Max, Automatic registration of lidar and optical imagery using depth map stereo, in: IEEE International Conference on Computational Photography (ICCP), IEEE, 2014, pp. 1–8.

[41] P. Rönnholm, M. Karjalainen, H. Kaartinen, K. Nurminen, J. Hyyppä, Relative orientation between a single frame image and lidar point cloud using linear features, Photogrammetric Journal of Finland 23 (2) (2013).

[42] T. Cui, S. Ji, J. Shan, J. Gong, K. Liu, Line-based registration of panoramic images and lidar point clouds for mobile mapping, Sensors 17 (1) (2017) 70.

[43] L. Wang, U. Neumann, A robust approach for automatic registration of aerial images with untextured aerial lidar data, in: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, 2009, pp. 2623–2630.

[44] M. A. Fischler, R. C. Bolles, Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography, Communications of the ACM 24 (6) (1981) 381–395.

[45] D. Scaramuzza, A. Harati, R. Siegwart, Extrinsic self calibration of a camera and a 3d laser range finder from natural scenes, in: IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), IEEE, 2007, pp. 4164–4169.

[46] N. Williams, K.-L. Low, C. Hantak, M. Pollefeys, A. Lastra, Automatic image alignment for 3d environment modeling, in: Brazilian Symposium on Computer Graphics and Image Processing (SIBGRAPI), IEEE, 2004, pp. 388–395.

[47] A. Mastin, J. Kepner, J. Fisher, Automatic registration of lidar and optical images of urban scenes, in: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, 2009, pp. 2639–2646.

[48] E. G. Parmehr, C. S. Fraser, C. Zhang, J. Leach, Automatic registration of optical imagery with 3d lidar data using statistical similarity, ISPRS Journal of Photogrammetry and Remote Sensing 88 (2014) 28–40.

[49] R. Wang, F. P. Ferrie, J. Macfarlane, Automatic registration of mobile lidar and spherical panoramas, in: IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), IEEE, 2012, pp. 33–40.

[50] Z. Taylor, J. Nieto, Automatic calibration of lidar and camera images using normalized mutual information, in: IEEE International Conference on Robotics and Automation (ICRA), 2013.

[51] M. Guislain, J. Digne, R. Chaine, G. Monnier, Fine scale image registration in large-scale urban lidar point sets, Computer Vision and Image Understanding 157 (2017) 90–102.

[52] J. Levinson, S. Thrun, Automatic online calibration of cameras and lasers., in: Robotics: Science and Systems (RSS), Vol. 2, 2013.

[53] K. W. Morin, Calibration of airborne laser scanners, University of Calgary (2002).

[54] H. Burman, Calibration and orientation of airborne image and laser scanner data using gps and ins, Ph.D. thesis, Institutionen för geodesi och fotogrammetri (2000).

[55] A. Habib, K. I. Bang, A. P. Kersting, J. Chow, Alternative methodologies for lidar system calibration, Remote Sensing 2 (3) (2010) 874–907.

[56] P. Friess, Toward a rigorous methodology for airborne laser mapping, Proceedings EuroCOW (2006) 25–27.

[57] J. Skaloud, D. Lichti, Rigorous approach to bore-sight self-calibration in airborne laser scanning, ISPRS journal of photogrammetry and remote sensing 61 (1) (2006) 47–59.

[58] M. Hebel, U. Stilla, Simultaneous calibration of als systems and alignment of multiview lidar scans of urban areas, IEEE Transactions on Geoscience and Remote Sensing 50 (6) (2011) 2364–2379.

[59] K. I. Bang, A. Habib, A. Kersting, Estimation of biases in lidar system calibration parameters using overlapping strips, Canadian Journal of Remote Sensing 36 (sup2) (2010) S335–S354.

[60] J. Skaloud, P. Schaer, Towards automated lidar boresight self-calibration, International Symposium on Mobile Mapping Technology (2007).

[61] A. P. Kersting, A. Habib, K.-I. Bang, J. Skaloud, Automated approach for rigorous light detection and ranging system calibration without preprocessing and strict terrain coverage requirements, Optical Engineering 51 (7) (2012) 076201.

[62] P. Glira, N. Pfeifer, G. Mandlburger, Rigorous strip adjustment of uav-based laserscanning data including time-dependent correction of trajectory errors, Photogrammetric Engineering & Remote Sensing 82 (12) (2016) 945–954.

[63] R. Keyetieu, N. Seube, Automatic data selection and boresight adjustment of lidar systems, Remote Sensing 11 (9) (2019) 1087.

[64] M. Li, H. Yu, X. Zheng, A. I. Mourikis, High-fidelity sensor modeling and self-calibration in vision-aided inertial navigation, in: 2014 IEEE International Conference on Robotics and Automation (ICRA), IEEE, 2014, pp. 409–416.

[65] K. Eckenhoff, P. Geneva, G. Huang, Closed-form preintegration methods for graph-based visual–inertial navigation, The International Journal of Robotics Research 38 (5) (2019) 563–586.

[66] E. S. Jones, S. Soatto, Visual-inertial navigation, mapping and localization: A scalable real-time causal approach, The International Journal of Robotics Research 30 (4) (2011) 407–430.

[67] T.-C. Dong-Si, A. I. Mourikis, Motion tracking with fixed-lag smoothing: Algorithm and consistency analysis, in: IEEE International Conference on Robotics and Automation (ICRA), IEEE, 2011, pp. 5655–5662.

[68] J. Nikolic, J. Rehder, M. Burri, P. Gohl, S. Leutenegger, P. T. Furgale, R. Siegwart, A synchronized visual-inertial sensor system with fpga pre-processing for accurate real-time slam, in: IEEE International Conference on Robotics and Automation (ICRA), IEEE, 2014, pp. 431–437.

[69] E. D. Nerurkar, K. J. Wu, S. I. Roumeliotis, C-klam: Constrained keyframe-based localization and mapping, in: IEEE International Conference on Robotics and Automation (ICRA), IEEE, 2014, pp. 3638–3643.

[70] M. Cramer, D. Stallmann, System calibration for direct georeferencing, International Archives of Photogrammetry Remote Sensing and Spatial Information Sciences 34 (3/A) (2002) 79–84.

[71] N. Yastikli, K. Jacobsen, Direct sensor orientation for large scale mapping—potential, problems, solutions, The Photogrammetric Record 20 (111) (2005) 274–284.

[72] L. Pinto, G. Forlani, A single step calibration procedure for imu/gps in aerial photogrammetry, International Archives of Photogrammetry and Remote Sensing 34 (B3) (2002) 210–213.

[73] J. Skaloud, P. Schaer, Towards a more rigorous boresight calibration, in: ISPRS International Workshop on Theory Technology and Realities of Inertial/GPS/Sensor Orientation, 2003.

[74] A. P. Kersting, A. Habib, K. I. Bang, Mounting parameters calibration of gps/ins-assisted photogrammetric systems, in: 2011 International Workshop on Multi-Platform/Multi-Sensor Remote Sensing and Mapping (M2RSM), IEEE, 2011, pp. 1–6.

[75] C. Siying, M. Hongchao, Z. Yinchao, Z. Liang, X. Jixian, C. He, Boresight calibration of airborne lidar system without ground control points, IEEE Geoscience and Remote Sensing Letters 9 (1) (2011) 85–89.

[76] Z. Taylor, J. Nieto, Motion-based calibration of multimodal sensor arrays, in: IEEE International Conference on Robotics and Automation (ICRA), IEEE, 2015, pp. 4843–4850.

[77] F. Amhar, J. Jansa, C. Ries, et al., The generation of true orthophotos using a 3d building model in conjunction with a conventional dtm, International Archives of Photogrammetry and Remote Sensing 32 (1998) 16–22.

[78] A. F. Habib, E.-M. Kim, C.-J. Kim, New methodologies for true orthophoto generation, Photogrammetric Engineering & Remote Sensing 73 (1) (2007) 25–36.

[79] H. C. de Oliveira, M. Galo, A. P. Dal Poz, Height-gradient-based method for occlusion detection in true orthophoto generation, IEEE Geoscience and Remote Sensing Letters 12 (11) (2015) 2222–2226.

[80] S. Katz, A. Tal, R. Basri, Direct visibility of point sets 26 (3) (2007) 24.

[81] S. Katz, A. Tal, On the visibility of point clouds, in: IEEE international conference on computer vision (ICCV), 2015, pp. 1350–1358.

[82] P. Vechersky, M. Cox, P. Borges, T. Lowe, Colourising point clouds using independent cameras, IEEE Robotics and Automation Letters 3 (4) (2018) 3575–3582.

[83] F. Zeng, R. Zhong, The algorithm to generate color point-cloud with the registration between panoramic image and laser point-cloud, in: IOP Conference Series: Earth and Environmental Science, Vol. 17, IOP Publishing, 2014, p. 012160.

[84] S. Cho, J. Yan, Y. Matsushita, H. Byun, Efficient colorization of large-scale point cloud using multi-pass z-ordering, in: International Conference on 3D Vision (3DV), Vol. 1, IEEE, 2014, pp. 689–696.

[85] S. A. K. Tareen, Z. Saleem, A comparative analysis of sift, surf, kaze, akaze, orb, and brisk, in: International Confer-

ence on Computing, Mathematics and Engineering Technologies (iCoMET), IEEE, 2018, pp. 1–10.

[86] J. L. Bentley, Multidimensional binary search trees used for associative searching, Communications of the ACM 18 (9) (1975) 509–517.

[87] P. J. Huber, Robust statistics, Springer, 2011.