



A sensor fusion framework for wearable sensor suites

Georges Chahine, Maxime Vaidis, François Pomerleau, Cédric Pradalier

► To cite this version:

Georges Chahine, Maxime Vaidis, François Pomerleau, Cédric Pradalier. A sensor fusion framework for wearable sensor suites. SN Applied Sciences, inPress. hal-03201502

HAL Id: hal-03201502

<https://hal.science/hal-03201502>

Submitted on 18 Apr 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Mapping in unstructured natural environment

A sensor fusion framework for wearable sensor suites

Georges Chahine · Maxime Vaidis ·
François Pomerleau · Cédric Pradalier

Received: date / Accepted: date

Abstract We present a generalized mapping framework that can withstand the challenges incurred by working in unstructured outdoor environments, such as a snowy forest. The proposed method takes advantage of a sensor fusion scheme, where sensors such as cameras and lidars are used in order to reconstruct the surrounding natural environment. Although mapping techniques such as SLAM and ICP cannot themselves properly handle the complexity of natural scenes, they do have the potential to contribute to the global solution in a proposed sensor fusion scheme, based on a factor graph architecture. In this paper, we propose an innovative map registration scheme for visual maps, and show how it can improve the reconstruction quality after data fusion. We also analyze the behavior and sensitivity of factor graphs to uncertainties, by comparing the residual error with different parameter combinations such as variances, using an exhaustive grid search with ground truth comparison. Finally, we suggest an ICP-inferred loop closure, capable of compensating position and attitude drift. The experiments are carried out by recording in a snowy forest using a wearable sensor suite.

In the experiments, ground truth was acquired using a millimeter-accurate total station. The proposed framework is shown to be robust and likewise

G. Chahine
Georgia Institute of Technology, Atlanta, GA, USA
E-mail: gchahine@gatech.edu

M. Vaidis
Université Laval, Québec City, Québec, Canada
E-mail: maxime.vaidis@norlab.ulaval.ca

F. Pomerleau
Université Laval, Québec City, Québec, Canada
E-mail: francois.pomerleau@norlab.ulaval.ca

C. Pradalier
Georgia Institute of Technology, UMI 2958 GT-CNRS, GeorgiaTech Lorraine
E-mail: cedric.pradalier@georgiatech-metz.fr

capable of providing estimates that are otherwise unattainable using classic techniques, such as visual SLAM and ICP for lasers. Finally, a visible improvement in the map reconstruction quality is shown, and the proposed framework achieves a translation error of 0.36 meters.

Keywords Mapping · Natural Environment · Sensor Fusion · Wearable Robots

1 Introduction

Autonomous robotic systems require a representation of their environment, such as a map, which is used to achieve specific objectives. For example, these maps are used to develop autonomous movement strategies (e.g., path planning, obstacle avoidance) or to perform specific tasks such as forest inventory [19]. In some cases, the map is the purpose of the mission. For example, data obtained after processing a map can be used to extract features such as tree diameters [23].

Several mapping solutions exist, yet they all depend on the environment in which they are evaluated, and on the desired task which might also be affected by environmental conditions. [4,29]. Based on that, the successful execution of a given mapping technique will have to take into account the environmental challenges that might cause it to fail. In this study, we will focus on the three-dimensional reconstruction of unstructured environments such as a forest, during challenging meteorological conditions. Such environments are a challenge for mapping algorithms such as Simultaneous Localization And Mapping (SLAM), but also for pose estimation of the robotic platform because of the roughness of the terrain. Traditionally, several types of sensors such as cameras and lidars can be used to obtain the pose of a robot.

Cameras are of a particular interest for applications involving robots, as they are considered sensors that can be used for localization and mapping *i.e.*, SLAM. Monocular cameras are particularly versatile, lightweight and comparatively cheaper than lidars, making them suitable especially for wearable sensor suites. Visual SLAM techniques allow the generation of pose and map estimates from a sequence of moving images. The booming of learning-based techniques meant that cameras can also significantly contribute to the decision-making process in artificial intelligence applications.

When extrinsics are available, cameras have the ability to complement lidars, given that they often include color or semantic data, that can be projected onto the lidar point cloud. Unfortunately in some environments, extracting robust features from camera images can be challenging, such as when working with featureless, snow covered landscapes. As a result, it becomes difficult to use a method purely based on visual odometry or vision-based SLAM.

As a result, convenience implies the use of other sensors such as lidars that perform better under visually challenging constraints. Unlike cameras, lidars provide high-precision point clouds and are capable of producing dense maps

within their range. Using the Iterative Closest Point (ICP) algorithm [2], lidar scans are incrementally matched to build a three dimensional map.

The ICP algorithm iteratively compares two point clouds in order to find the rigid transformation that minimizes the alignment error between them. Although this method ensures local consistency, a global drift appears over time [20]. In case the trajectory does not loop, the drift cannot be corrected using a loop closure, hence it is necessary to use other techniques to refine the pose of the robotic platform. These drifts can be also attenuated by SLAM methods that use optimization graphs [5], or by constraining the position of the data acquisition system using Global Position System (GPS) priors.

Some research has used deep learning with range sensors in order to decrease the processing time during map registration such as [25]. These methods have the advantage of not adhering to the rigid motion assumption to compute the pose of the system. Other point cloud registration techniques include the Normal Distribution Transform (NDT) algorithm [3], although favorable results require a lidar with at least 32 beams [16]. The affordability and increasing popularity of 3D lidars have driven an interest in 3D point cloud descriptors such as 3DSmoothNet [11] and in correspondence alignment techniques such as TEASER++ [28].

The state-of-the art in SLAM has so far failed to address the challenges encountered in unstructured natural environments [6]. Further, visual SLAM techniques lack both scale and the density seen in maps generated using 3D lidars. In a previous work, we surveyed the performance of monocular visual SLAM algorithms in natural environments [6], and found that Direct Sparse Odometry (DSO) [10] has a performance edge when it comes to natural environment applications.

We call data fusion any scheme or approach that takes advantage of multiple inputs from sensors such as GPS, Inertial Measurement Unit (IMU), and possibly some kind of odometry including, wheel, visual and ICP-based. The purpose of sensor fusion is to obtain a refined estimate of a quantity. In this paper, it is the 6 Degrees of Freedom (DoF) pose, which is otherwise unattainable using an individual sensor [13]. For the purpose of fusing pose estimates, factor graphs are capable of integrating different estimates such as odometry and GPS readings *i.e.*, geometric constraints [17]. The trajectory to be estimated is discretized into several points of reference along it, which we call nodes. These nodes represent the robot 6 DoF pose and they are interconnected using geometric constraints, as well as priors that reflect absolute knowledge on the pose of a robotic platform. What follows is a smoothing and optimization step to refine the pose estimate at each node [9]. Factor graphs are therefore powerful tools for sensor fusion since constraints and priors can be inferred from multiple sources.

The goal of this paper is to evaluate and adapt available techniques such as factor graphs, ICP and scale-lacking monocular visual odometry for the purpose of creating visual representations of natural environments, using visual and range sensors such as cameras and lidars, respectively. We show how we can recover the scale in the context of a factor graph, improve the reconstructed

visual map after sensor fusion and use visual odometry to further refine pose estimates.

We highlight the sensitivity of factor graphs to uncertainties and determine unknown parameters such as uncertainties on visual odometry, by running an exhaustive grid search to determine the best parameter combination. We also explain how noisy visual map reconstruction can be even after pose optimization, and suggest as a solution a piecewise reconstruction and alignment scheme. The suggested framework includes a loop closure step, inferred using ICP in revisited geographic areas. In this paper, we also evaluate the suitability of the state-of-the-art in correspondence alignment for mapping in natural environment. We underline performance of ICP and its suitability for lidar scan matching on slowly moving robotic platforms such as wearable sensor suites. Our approach was tested by recording data in forest during sunny and snowy weather, with the sensor suite presented in [Figure 2](#). In [Section 4](#), the results were compared to a ground truth given by a theodolite in order to have an accurate assessment of the pose given by the factor graph.

2 Related work

The generation of 3D maps from a moving platform is an evolving field of study, due to the increasing availability and affordability of exteroceptive sensors such as cameras and lidars. Unfortunately, the majority of available techniques have either focused on indoor or structured outdoor scenes, with few of these methods such as [\[10\]](#) and [\[6\]](#) for visual methods along with [\[1\]](#) for laser-based methods having partial success in unstructured outdoor scenes, and in particular the reconstruction of trees. In the context of a natural environment, the state-of-the-art does not provide specialized solutions for 3D reconstruction, yet recent works such as [\[27\]](#) and [\[7\]](#) are starting to address some of the challenges associated with 3D reconstruction of natural organic structures. The task of trees crown reconstruction remains unsolved due to the variability of trees phenotype, with some research focused on CAD model replacement [\[8\]](#) or radius estimation from flying laser measurements [\[15\]](#). Additionally, works such as [\[27\]](#) and [\[26\]](#) show that it might be easier to model tree structure such as a trunk, branches and twigs when they are exposed and observable, such as during winter when the trees are bare. Another challenge is meshing the point clouds to create surfaces that are suitable for autonomous robot navigation such as path planning, with few libraries such as [\[21\]](#) and [\[24\]](#) providing useful meshing tools. The topological representation of a point cloud is, however, open-ended and in many cases specific to the application. Nevertheless, some libraries such as [\[24\]](#) provide hole detection and filling tools, and most point cloud processing libraries such as [\[21\]](#) and [\[24\]](#) provide some form of plane fitting for surface estimation, usually using RANSAC or marching cube techniques for surface fitting. HDL Graph SLAM [\[16\]](#) presents a graph-based sensor fusion approach to mapping, capable of integrating GPS, IMU and laser scanners. The use of the word “HDL” references to the authors’ experiments

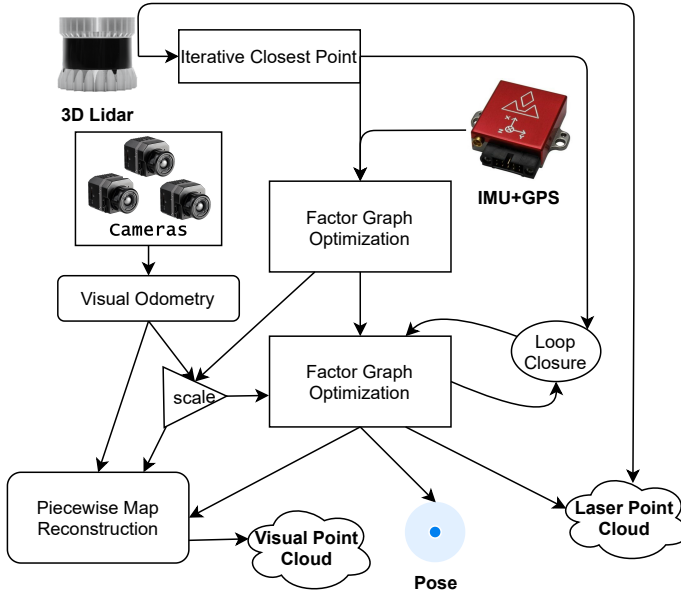


Fig. 1: Flow chart of the proposed framework, including the sensors used during experimentation.

with the velodyne HDL-32e laser. HDL SLAM does not support the integration of visual odometry and when testing the method, it has been noticed that there is no scheme for dense reconstruction in between graph nodes.

In a recent work [7], we presented a real-time 2D laser-supported visual tracking solution for natural environment, dependent on a camera feed and a sparse disparity image generated by projecting a 2D laser onto the corresponding image. However, laser-supported visual odometry assumes an overlap between the lidar field of view and that of the camera. In addition that, it does not support the integration of multiple cameras or inertial measurements.

3 Framework

The proposed framework, shown in Figure 1, is a sensor fusion scheme based on a factor graph and in particular the implementation of Dellaert *et al.* [9]. The data is assumed to be time synchronized, as detailed in the implementation presented in Section 4. For the purpose of this section, the system is assumed to have, at least, one monocular camera, one 3D laser, an IMU and a GPS receiver. An approximation of the camera intrinsics and the various extrinsics between sensors is assumed to be known as well. The 3D laser can be replaced with a 2D laser, as long as that the field of view of the laser intersects with that of at least 1 camera. Unfortunately, laser-based SLAM is unsuitable for vertically installed 2D lasers such as in Figure 2a, given the lack of overlap

in between measurements. In this paper, the use of ‘global frame’ refers to the Universal Transverse Mercator (UTM) frame and in particular the North-East-Down (NED) convention, representing x , y and z respectively.

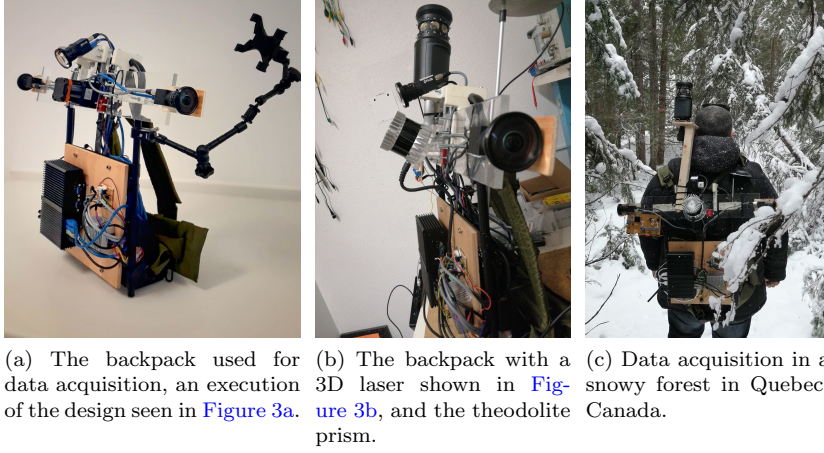


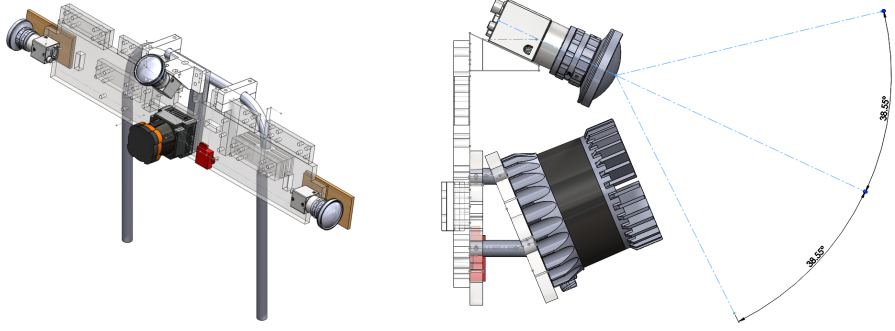
Fig. 2: Implementation of the design shown in Figure 3.

3.1 Data synchronisation

Different sensors exhibit different latencies that cannot be precisely measured or compensated. Further, cameras suffer additional delays such as image integration time (shutter length), and are susceptible to delays during data transmission, due to high bandwidth requirements in high frequency transmission. Consequently, asynchronous data collection does not guarantee data association and temporal synchronization, as highlighted in [12] and [14]. To solve the latter problem, the suggested framework takes advantage of a synchronized physical trigger.

3.2 Factor graph

We define a factor graph node distancing parameter d , that constrains the number of nodes in a factor graph. For instance, between two consecutive nodes, the robotic platform would have traveled at least d meters along the trajectory. The distancing parameter cannot be inferred from visual odometry due to lack of scale. Instead, other solutions such as IMU-aided GPS, laser-based ICP (for 3D lasers) or laser-aided visual odometry (for 2D lasers) can be used to determine the node distancing parameter. In this paper, we construct



(a) The backpack CAD model, showing a 2D Hokuyo laser (center), an Inertial Navigation System (INS) here in red, and three cameras: two on the sides and a center camera.

(b) Backpack CAD model, showing the Ouster-16 lidar model (bottom) on a custom-made mount to avoid interference with the field of view of the camera (top).

Fig. 3: All fixtures were 3D printed or laser-cut according to a pre-designed CAD model, so that to minimize geometric projection error in between components. The instruments are mounted on a rigid acrylic plate, the latter affixed onto the metal frame of a backpack.

our nodes by downsampling the input GPS-IMU data, using a distance-based downsampling scheme, *i.e.*, discarding all but the first measurement, each time the system moves by d meters. The latter scheme therefore accounts for occasional movement pauses while traversing rough terrain despite non-interrupted data recording, such as in human-carried sensor suites.

Let d_{gps} represent the sparsity of GPS prior constraints, according to which the GPS measurements are distributed over the factor graph nodes. In other words, we constrain our pose with GPS measurements every d_{gps} nodes *i.e.*, $d_{gps} * d$ meters. We now construct a factor graph with:

- A GPS prior every d_{gps} nodes.
- An attitude prior inferred from the INS, which describes the attitude of the IMU with respect to the global frame (UTM-NED) with bias compensation using the GPS direction of motion propagation. In the experimental section, bias compensation is computed internally inside the INS sensor.
- Multiple odometry readings, either from laser-based ICP and/or laser-supported visual odometry. A robotic platform with wheel odometry can also be used as an alternative. In this paper, we constrain our system using ICP odometry and scaled visual odometry.

3.3 Scale recovery and interpolation

The lack of scale in visual odometry measurements prevents the insertion of scaleless constraints into the factor graph. Instead, we need to scale these

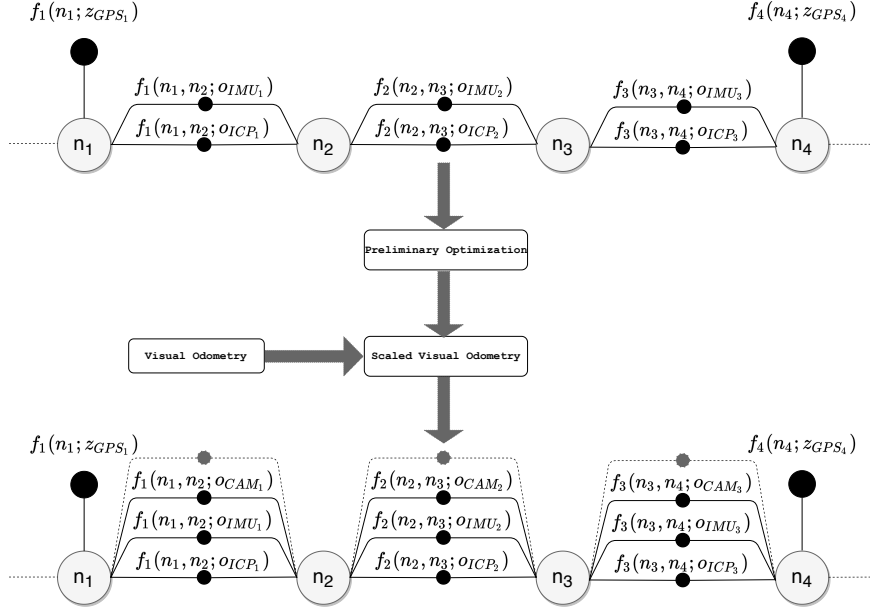


Fig. 4: Factor graph flow chart.

measurements either using GPS or a less noisy estimate from a preliminary factor graph optimization, given the availability of ICP odometry. Using the factor graph defined above, a preliminary factor graph optimization using GPS and attitude priors, as well as ICP odometry is performed and the resulting pose is used to scale visual maps, as shown in Figure 4.

The scale for a visual odometry constraint at a given factor graph node n is estimated by simply accumulating the traveled distance along the trajectory from visual odometry, over the traveled distance inferred from the optimized pose *i.e.*, $s_n = D_{vo}/D_{opt}$, where s_n is the scale for a visual SLAM constraint, with D_{vo} and D_{opt} representing the cumulative traveled distance over the last 100 meters before reaching the current position at node n . D_{vo} and D_{opt} are inferred from visual odometry and the preliminary factor graph optimization, respectively.

For each camera, we then interconnect nodes with scaled odometry constraints, followed by another optimization run to refine the pose and consequently the estimated scale. Additionally, three-dimensional reconstruction of the maps requires the recovery of the 6 DoF pose at the original rate. This is also important for cameras due to the keyframe-based nature of many visual algorithms, and therefore there is a need to interpolate the optimized 6-DoF pose. Using timestamp queries with the optimized pose from factor graph, we linearly interpolate the translation component of the pose. As for the rotation part, we use quaternions Spherical Linear Interpolation (SLERP) [22].

3.4 Point cloud registration

Point cloud registration is dependent on the context in which the data was captured and the sensor used. For instance, vertically installed 2D lasers cannot be registered using ICP, due to lack of overlap in between scans. In this subsection, we exploit the various solutions to point cloud registration, for both 3D lasers and cameras.

3.4.1 Merging of visual maps

Maps captured by most Visual SLAM algorithms work on the basis of keyframes, which do not necessarily overlap, hence ruling out ICP as a solution to keyframe registration. Unlike lasers, cameras can see far away objects that might show up in a keyframe even after the sensor suite has long left the vicinity of the object. Without breaking keyframe integrity, this results in a noisy, coarse reconstruction when attempting to reconstruct the map by simply transforming the keyframes using the optimized pose from factor graph, or any other pose but the original pose from visual odometry. We therefore suggest a hybrid approach for point cloud registration consisting of local reconstruction of the map using visual odometry, with ICP smoothing for the piecewise alignment of map fragments.

For a set of visual SLAM keyframes S_i captured along a portion of the trajectory, we transform any point P_{ij} in the sensor frame to the global frame using $P_{zj} = T_{ij}T_{ext}P_{ij}$, where:

T_{ij} represents the forward transformation from the base frame to the global frame, T_{ext} is the extrinsic transformation, also known as the sensor to base transformation, P_{zj} is the projected point in the global frame, and j represents the keyframe number, with $j = 1$ referring to the first keyframe.

Let \mathcal{M}_c and \mathcal{M}_l be 2 preliminary point clouds used to construct a refined point cloud \mathcal{M}_f . All three point clouds, shown in [Figure 5](#), represent the same scene but are differently constructed.

\mathcal{M}_c is a coarse point cloud constructed by simply projecting visual SLAM keyframes using the corresponding optimized pose from factor graph. In other words, \mathcal{M}_c is constructed by inferring T_{ij} from the factor graph optimized pose.

\mathcal{M}_l is a hybrid map constructed using pose increments from sensor original data *i.e.*, visual SLAM. The latter pose increments build on sparse factor graph poses, serving as anchors distributed along the optimized trajectory. What follows is the derivation needed to reconstruct a point cloud \mathcal{M}_l .

At the beginning of every set of keyframes, we set $T_{i,j=1} = T_{o_k}$, where T_{o_k} represents the optimized pose from factor graph at the corresponding node k . This has an anchoring effect and prevents the map from significantly drifting away from the associated pose. We now derive the forward transformation that locally projects the points to the global frame, *i.e.*, T_{ij} , dropping the subscript i for brevity with $j \neq 1$:

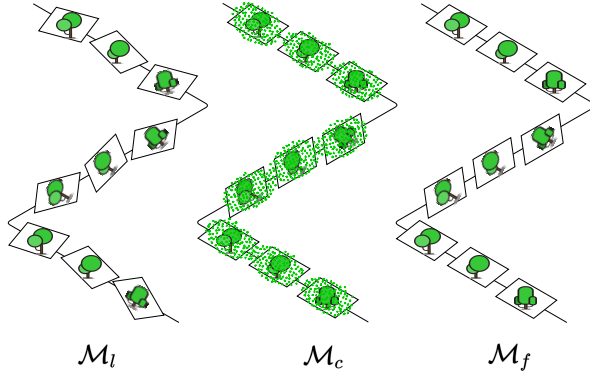


Fig. 5: Schematic showing the 2 intermediate maps \mathcal{M}_c , \mathcal{M}_l and the final map \mathcal{M}_f . Snapshots of \mathcal{M}_c and \mathcal{M}_f are shown in Figure 11.

$$T_j = T_{j-1} T_{vo_{j-1}}^{-1} T_{vo_j} \quad (1)$$

where T_{vo} represents the sensor to world transformation, from visual odometry.

Setting $T_{i1} = T_{o_k}$ on the first keyframe of every set implies a misalignment of map fragments, due to the resetting effect upon which Equation 1 builds up subsequent pose increments. To solve this, we correct the point cloud \mathcal{M}_l to obtain a point cloud \mathcal{M}_f by aligning, using ICP, the points in a given set S_i with the corresponding points in the coarse point cloud \mathcal{M}_c .

3.4.2 Merging of range maps

Range scans obtained from 3D lasers overlap and therefore are suitable for ICP-based solutions. In this paper, a hybrid approach for map registration identical to the method presented in Section 3.4.1 is used by replacing visual odometry by an ICP mapper that takes advantage of an interpolated pose prior from factor graph. In other words, we register \mathcal{M}_l using ICP in-between factor graph nodes, and \mathcal{M}_c using the interpolated pose from factor graph.

3.5 Exhaustive parameter search

An exhaustive search for parameters is a brute force technique consisting of trying all possible values of a parameter, while looking for desired changes in a system dependent on the said parameter. Feedback on the search process is dependent on the application and the system at hand. Exhaustive search techniques are especially useful when *a priori* knowledge on the value of a parameter is poor or unavailable.

The proposed parameter search scheme is an exhaustive search of factor graph uncertainties as well as other parameters that might affect the outcome.

For instance, we noticed that varying d_{gps} has a considerable effect on the outcome, and therefore d_{gps} was included in the parameter search scheme along with the uncertainties. The proposed exhaustive search implies trying all parameter combinations in a constrained range of values, followed by a factor graph optimization step. We then compute the error feedback and log all the parameters with the associated error value.

Setting the covariance matrices for a factor graph can be challenging: unlike sensors that come with a datasheet, laser-based ICP or visual odometry do not have a defined uncertainty. Although the parameters can be manually fixed, a better estimation can be achieved through an exhaustive parameter search scheme, which runs through a constrained range of values to determine the covariances associated with visual odometry, ICP-generated pose, GPS and IMU measurements as well as the factor graph parameter d_{gps} .

Subsequently, parameter search is performed by logging different parameter combinations against alignment error with the reference (*i.e.*, ground truth) trajectory. Unfortunately, an exhaustive parameter search is a computationally expensive task, and in this context assumes the availability of ground truth data. Further, uncertainties for different measurements are dependent on the sensor as well as the environment around it. For instance, laser measurements depends on the surface material upon which laser beams reflect, as well as climatic conditions. For the purpose of the exhaustive parameter search, the bulk of the uncertainty is assumed to be interoceptive to the sensor or method used to generate the measurements. The latter assumption hypothetically allows the generalization of the parameters for a sensory platform, disregarding exteroceptive stimuli. Further discussions in [Section 4](#) show that the parameters we find for a given survey remain valid for other surveys that might not include the ground truth required to run the exhaustive parameter search.

3.6 Loop closure

Position and attitude drift is expected with most incremental sensor systems. For an outdoor system such as the one presented in this paper, both visual and laser-based measurements accumulate errors, the latter constrained by IMU-aided GPS priors, in the context of a factor graph. In this section, we suggest a two-step loop closing scheme; the first step consisting of a correspondence match generation function followed by a laser-based, ICP-inferred geometric constraint.

3.6.1 Correspondence generation

The underlying assumption for correspondence generation is that translational drift, constrained by GPS priors, is well below the laser range threshold, currently between 30 and 50 meters for commercially available 3D lasers. Given a factor graph with n nodes, with the first node being N_1 and the last node being N_n . We generate the $n \times n$ correspondence map \mathcal{C} , such that $\mathcal{C}_{(i,j)} = 1$ if

the laser-generated maps associated with neighbor nodes N_i and N_j overlap, and $\mathcal{C}_{(i,j)} = 0$ everywhere else. To check for corresponding nodes, the fifty nearest neighbors for a node N are tested against two distance constraints. The first constraint is a proximity parameter, typically between five and fifteen meters that guarantees map overlap. The second constraint is an elapsed distance over the trajectory constraint typically between fifteen and thirty meters. Every node N_i is associated with a distance D_i , which measures the elapsed distance over the trajectory from N_1, N_2, \dots to N_i . The second constraint guarantees that the nearest neighbor lies in a revisited space, after the sensor suite has initially left the area.

3.6.2 Geometric constraint

We propose an ICP-based geometric constraint, using the correspondence map found in [Section 3.6.1](#). Owing to the constrained nature of the field of view and namely the opening angle of 3D lasers, the point clouds associated with two corresponding nodes *i.e.*, just two laser scans, are not guaranteed to significantly overlap. As a solution to that, using the optimized pose from the factor graph, we locally reconstruct the laser map observed from consecutive factor graph nodes $N_1 \dots N_k$, using interpolated factor graph poses $P_1 \dots P_k$ to create sub-maps. A matching sub-map is subsequently found using the correspondence map \mathcal{C} . A series of nodes and their associated maps form a sub-map if and only if every node in the series associate with at least a single correspondence in \mathcal{C} , and that no significant gaps exist in between consecutive scans to guarantee map continuity. Further, we define the map size limit such using the length of the trajectory $(P_k - P_1) < Th_{lc}$, where Th_{lc} is the map length limit over the trajectory typically between ten and forty meters, over which a new match pair is initialized. Finally, for every match pair, we infer the geometric transformation using ICP and add the inferred constraint to the factor graph.

4 Experiments

The dataset used in this paper is composed of a forest environment, in the area surrounding Montmorency Forest in Quebec, Canada, shown in [Figure 6](#). A secondary survey was also captured in an open field next to Université Laval. Data were captured using the sensor suite fitted on the backpack shown in [Figure 2a](#). As shown in [Figure 3](#), the data acquisition system has been designed to include three 2.0 megapixels RGB cameras from FLIR, a 16-beam Ouster-16 lidar from Ouster, and the VN-200 INS from Vectornav, the latter consisting of an IMU and a GPS. The system is physically synchronized using a laser-generated trigger, that can be downsampled using an on-board frequency divider circuit. Finally, the data is recorded on the on-board computer. For the remainder of this paper, the term ‘forest survey’ refers to the dataset captured in the snowy forest shown in [Figure 6b](#).



(a) Ground truth recording using the theodolite total station. (b) Snapshot from the center, backward-facing camera on the backpack sensor suite.

Fig. 6: Data acquisition in forestry environment

The versatility of a backpack allows survey recording in narrow or difficult to navigate regions, as shown in Figure 2c. Further, designing sensor fixtures and having them rigidly affixed to the backpack frame, means that sensor extrinsics can be calculated by using measurements inferred from the design model. For cameras, this includes taking into account the location of the image plane inside the camera hull, the latter provided by the camera manufacturer datasheet. Further, we used a Trimble S7 theodolite to obtain a translational ground truth. By affixing a theodolite prism onto the backpack as shown in Figure 2b and Figure 6a, we were able to have an accuracy to the millimetre of its trajectory.

The exhaustive parameter search was performed by defining four lists L_1 , L_2 , L_3 , L_4 that contain, respectively, a range of uncertainties for GPS, visual odometry, and ICP measurements as well as the sparsity parameter for GPS constraints d_{gps} . The lists of uncertainties are designed to be inclusive of all logical potential values:

- The list of GPS uncertainties L_1 ranges between 1 and 19, discretized with steps of 1.5 meters, for a total of 13 possible parameters.
- The list of visual odometry uncertainties L_2 ranges between 1 and 16, discretized with steps of 1.5 meters, for a total of 11 possible parameters.
- The list of sparsity constraints L_3 ranges between 1 and 30, discretized with steps of 1 node, for a total of 30 possible parameters.
- The list of ICP uncertainties L_4 ranges between 1 and 25, discretized with steps of 1.5 meters, for a total of 17 possible parameters.

In practice, with 4 parameters we have a $O(\prod_{i=1}^4 n_i)$ complexity, where n_i is the number of elements in a list L_i . The exhaustive search for the four parameters took around 60 hours, using an off-the-shelf Intel Core i7 computer with 16 GB of RAM. What follows is an exhaustive parameter search as presented

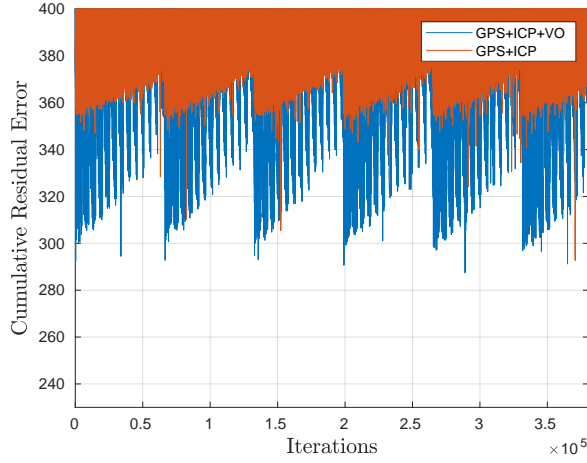


Fig. 7: Cumulative trajectory alignment error, in function of an expanded set of parameter combinations, with and without visual odometry.

in Algorithm 1.

Algorithm 1: Exhaustive parameter search snippet

```

logger=[];                                     // empty list
for every GPS uncertainty parameter p1 in list L1 do
    for every visual odometry uncertainty parameter p2 in list L2 do
        for every GPS sparsity constraint parameter p3 in list L3 do
            for every ICP uncertainty parameter p4 in list L4 do
                Get optimized pose from factor graph
                Calculate trajectory alignment error
                state = [error, p1, p2, p3, p4]
                append(logger, state)

```

5 Results

5.1 Quantitative analysis

Figure 7 shows the error variations in a partial parameter search. Still in the same figure, it is shown that the inclusion of visual odometry improves the overall pose estimates, in almost all possible parameter combinations. Moreover, the periodicity in error oscillations suggests that certain parameters tend to be more dominant than others with respect to their contribution to the error term, therefore imposing their own periodicity on that of the error. Based on that and given the mixture of parameters at hand, we generate graphs to highlight the effect of individual parameter variability, as shown in Figure 8.

The sensitivity of factor graphs to the estimated uncertainties, here expressed in variances, is highlighted in Figure 8, proving the existence of a narrow range of variances, at which the factor graph optimization performs at its best. Further, certain variances such as those associated with GPS measurements show a more forgiving range compared to the parameter associated with visual odometry or the parameter d_{gps} , the latter two parameters seemingly having a considerable effect on the optimization process. Still in the same figure, the standard deviation of the error can influence parameter selection. For example, Figure 8b shows a considerable fluctuation of the error for laser-based ICP odometry, in the range between 1 and 5 meters. Therefore, it is convenient to choose the values between 5 and 10 meters, as they suggest an improved stability and therefore a higher repeatability. In contrast to the latter, Figure 8c suggests that uncertainties attached to GPS measurements do influence the system performance yet less significantly than the other three parameters in the same figure. This is due to the small range in which the error varies for variance values ranging between 1 and 20 meters, reinforced by a largely consistent standard deviation all throughout the range of values.

Comparing the optimized pose from factor graph with the ground truth, we generate the Root Mean Square (RMS) error for the translational component. The alignment RMS error, inferred from the residual RMS error associated with the ground truth trajectory alignment, was added to the evaluation since it penalizes for trajectory deviations and rotation inaccuracies. Given that even small deviations from ground truth, as shown in Figure 9a are carried through the trajectory and globally affect the alignment quality, the associated error is expected to be significantly higher than its translational counterpart. Nevertheless, a perfect alignment is desirable and therefore it is a suitable metric in comparing different methodologies.

Table 1: Accuracy metrics for the forest dataset and a secondary survey.

Survey	Translational RMS	Alignment RMS
Main (Forest)	0.36 m	2.4 m
Secondary (Open field)	0.45 m	8.29 m

Table 1 shows the RMS performance for the forest survey, and an additional survey that has been added for the purpose of evaluating the suitability of previously found parameters in other surveys. Still in the same table, a secondary survey run showed the usability of the previously found parameters on another survey, yet with a slight increase in the translational RMS and a 4-fold increase in the alignment RMS. Note that, to the best of the authors' knowledge, no other parameter combination has achieved better results. This is most likely due to the nature of the dataset and the constrained field of view of the laser, as it happens for the secondary survey, the laser mostly sees the ground therefore the degradation of ICP-based estimators. In a future work,

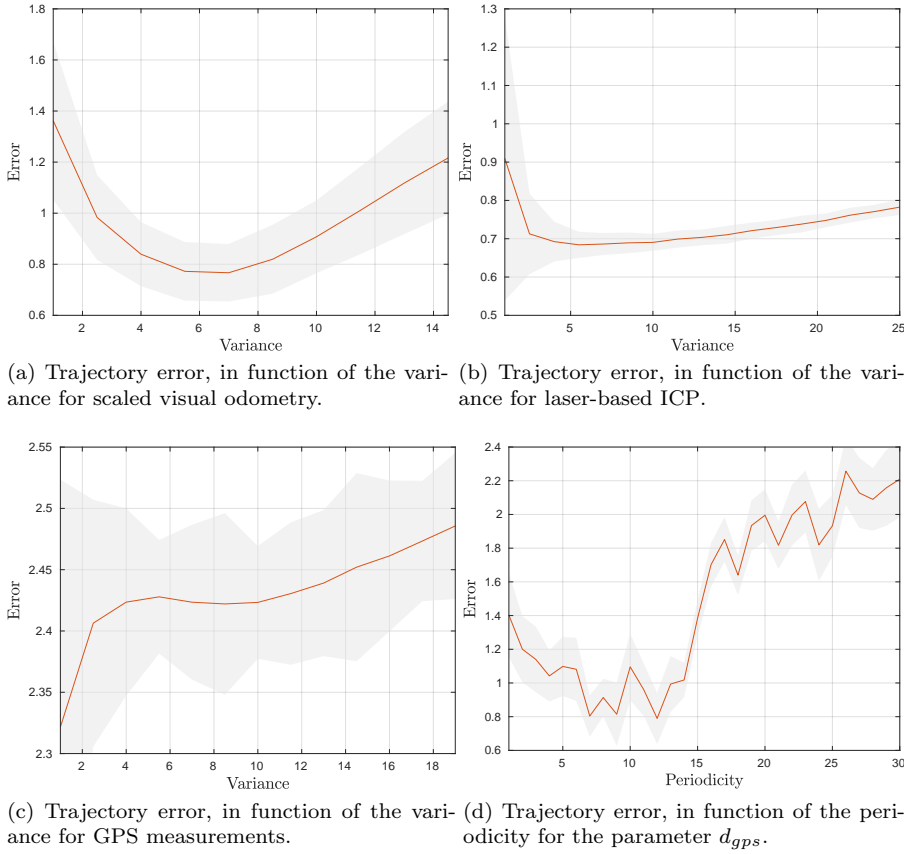


Fig. 8: Exhaustive parameter search, to evaluate the effect of certain parameters against performance metrics obtained from ground truth. The error values shown above, are the lowest error values (lowest 1%), obtained by fixing a given parameter and looking at error values against all other parameter combinations. The shaded area represents the standard deviation of the error.

the cylindrical 3D laser will be replaced by a hemispherical laser with a more suitable field of view.

5.2 Qualitative analysis

The increasing availability and affordability of 3D lasers reverberate in the demand for map registration techniques. In that context and given the dependability of the suggested system on laser-inferred odometry, we study the performance of ICP in natural environment, one of the most used registration algorithms for laser scan registration. The initial approach was to evaluate the

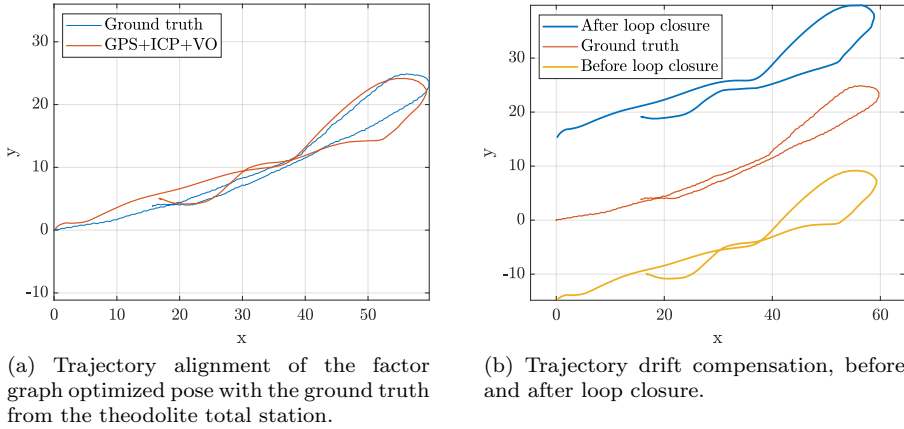


Fig. 9: Comparison of trajectories with the ground truth.

usability, in natural environment, of the state-of-the-art in ICP registration techniques such as TEASER++ [28]. The latter was evaluated by randomly sampling nearest neighbors for correspondence matching, but was found to be largely unreliable. The experiment was subsequently modified to take advantage of 3DSmoothNet [11] for correspondence matching, yet with no discernible performance improvement.

The performance of classical ICP *i.e.*, with nearest neighbor correspondences was found to be easily affected by the rate at which laser scans are matched. To further elaborate on the latter hypothesis, we set the lidar operating at its maximum frequency (20 Hz) and then down sample the data, resulting in the plot shown in Figure 10 and the metrics shown in Table 2. Still in the same table, it is shown that increasing the data rate increases the performance of ICP in comparison with the ground truth, yet to a certain limit after which the performance is severely degraded. Presumably, the performance degradation at the maximum frequency of 20 Hz is due to the relatively low speed (< 1 m/s) of the person carrying the sensor suite, making small pose differentials estimated by ICP in between consecutive scans to be overshadowed by the noise surrounding the estimation.

Table 2: Accuracy metrics for ICP at different data rates.

Frequency	Alignment RMS	Frequency	Alignment RMS
0.5 Hz	33.57 m	5 Hz	4.43 m
1 Hz	21.60 m	10 Hz	3.88 m
2 Hz	11.37 m	20 Hz	32.91 m

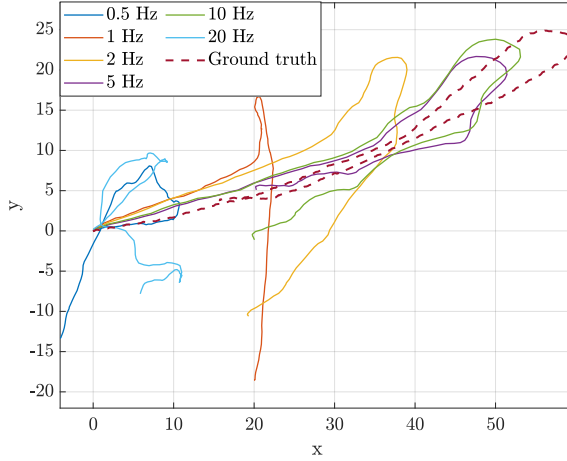


Fig. 10: Comparison between laser-based ICP and ground truth trajectories, at different data rates.

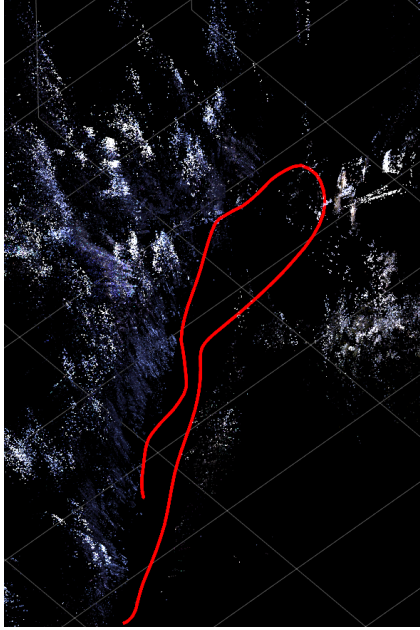
The integration of GPS measurements into factor graph architectures is not a trivial task, due to non-Gaussian noise contamination of position estimates [18]. GPS priors, however, do have the ability to globally constrain factor graphs and therefore eliminate accumulating drift from odometry. A GPS prior could have been implemented either:

- by imposing a loose, frequent constraint *i.e.*, a prior on every node with a relatively high variance (> 15 meters)
- by imposing a tight, sparse constraint *i.e.*, a prior every d_{gps} nodes with a small variance (< 7 meters)

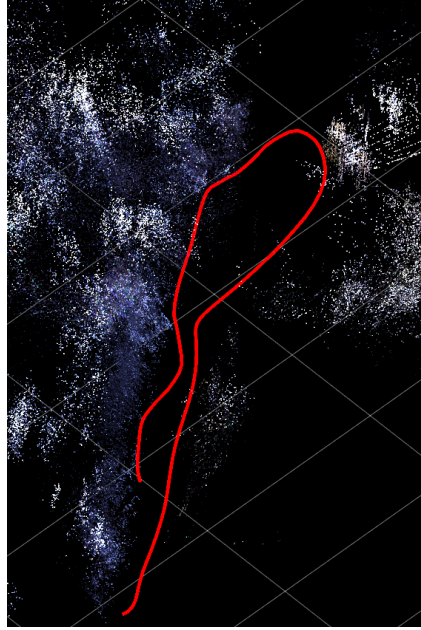
Figure 9 shows the trajectory alignment, before and after loop closure. Notice in Figure 9b how loop closure eliminates accumulating drift in the overlapping part, although it has little effect on the overall outcome in non-overlapping parts.

Since point cloud perception is not trivial, and since surface reconstruction improves human perception such as when displaying the maps in virtual reality, we used the Las Vegas reconstruction toolkit [24] in order to mesh the point cloud shown in Figure 12b. Since all maps are represented in the same global frame, the colors in the latter figure have been inferred by averaging the red, green and blue channels of the 10 nearest neighbor points in the corresponding visual map, the latter generated by the center camera on the sensor suite (Figure 6b). The lighting discrepancy is caused by varying lighting conditions, and the available shade at the moment when the data was recorded.

Although human perception of sparse point clouds can be difficult and especially in unstructured scenes, the effect of the visual reconstruction scheme suggested in Section 3.4.1) is nevertheless discernible in Figure 11. Comparing Figure 11a and Figure 11b, notice how noisy global reconstruction is (Fig-



(a) 3D reconstruction using the optimized pose with local reconstruction and map alignment, *i.e.*, \mathcal{M}_f



(b) Coarse map using the optimized pose without local reconstruction and map alignment, *i.e.*, \mathcal{M}_c .

Fig. 11: Comparison of the effect of local reconstruction with map alignment as discussed in [Section 3.4.1](#). The above figure shows an isometric view of a scene consisting of trees and a house on the top right corner.

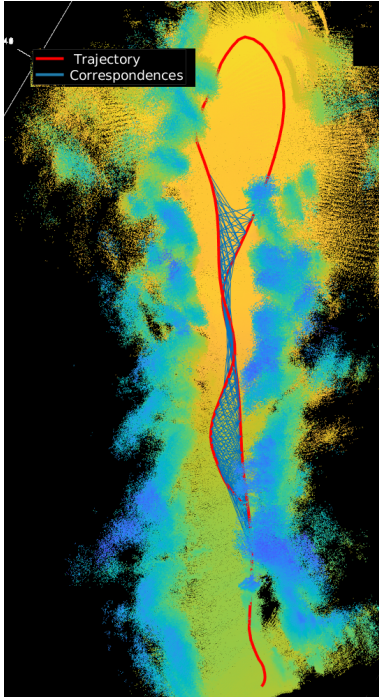
[ure 11b](#)), compared to the left-side image that uses local piece wise reconstruction with global alignment.

Unlike monocular cameras, laser-generated point clouds are dense and therefore more desirable and aesthetic. [Figure 12a](#) shows the top view of a laser-generated dense map with loop closure correspondences generated as previously discussed in [Section 3.6.1](#). Still in the same figure, notice how corresponding point couples fit well within the lidar range *i.e.*, the ensuing point cloud from local map reconstruction is guaranteed to overlap in the vicinity of the correspondences, as previously discussed in [Section 3.6.1](#).

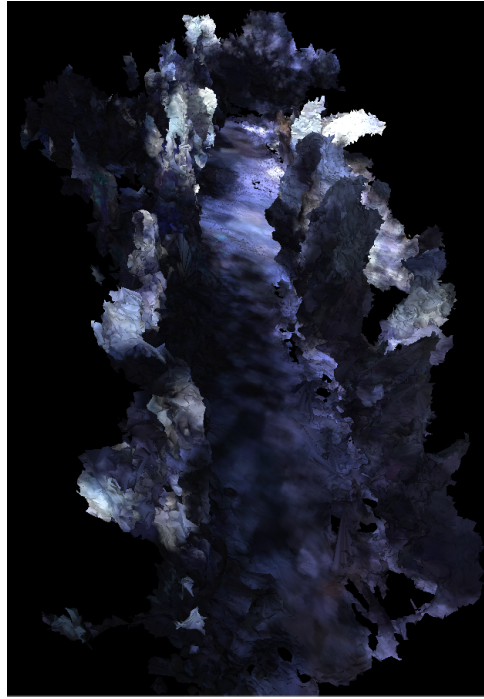
A video showing the colored point cloud being manipulated is available [here](#).

6 Conclusion

We have presented a mapping framework in an extremely challenging unstructured environment, capable of integrating monocular cameras, lidars and



(a) Laser point cloud reconstructed using the interpolated pose from factor graph, showing a trail between 2 columns of fir trees. The blue lines are drawn using the correspondence map (Section 3.6.1).



(b) Laser-generated point cloud, colored by averaging the color information of nearest neighbors in the sparse point clouds, produced from visual odometry, such as the ones shown in Figure 11. The point cloud was also meshed using the Las Vegas reconstruction toolkit. (video)

Fig. 12: Laser-generated point cloud, showing the same scene as a colorless point cloud and a colored mesh representation.

inertial measurements with loop closure. It has been demonstrated that the presented map integration scheme is robust and results in human-readable virtual depictions of natural environment elements, using a wearable sensor suite. The importance of this achievement is of great value for scientists looking to detect changes in natural environment. The advantages of the proposed method lies in its robustness, since the state-of-the-art in visual mapping such as DSO, and the state-of-the-art in laser mapping such as ICP, cannot survive the full length of the survey without catastrophic failure. Another advantage is the flexibility of adding as many exteroceptive sensors as needed, creating a multi-layered map. For example, a near-infrared mapping layer can be added using a suitable camera, a technique that is often used by environmentalists to characterize plant type and growth. In dense forests, by filtering laser points to include only the points in a cone above the sensor suite, our method can

be used to generate special types of maps, that reflect the canopy density and sky openness. The latter can be used to estimate light availability for certain plant species. Using a factor graph architecture, it is also possible to add a temporal layer, to compare maps taken at two different points in time. This approach will be addressed in a future work.

In this work, It has been shown that techniques such as ICP are of considerable utility, and can be adapted to challenging natural environment applications, for the purpose of local map alignment as well as loop closure. It has been also shown that factor graphs, when correctly manipulated, are suitable for refining pose estimates from different observations as well as to scale unitless estimates, generated using monocular techniques.

The proposed method is however limited to the availability of extrinsic parameters, which in this paper, is extracted from the CAD model of the sensor suite. Other limitations include the ability to systematically detect the failure of visual SLAM algorithms, to prevent local discontinuities in the factor graph setup. Other practical limitations for data acquisition include battery life, which is severely reduced in cold weather.

This work however exposed a considerable lack of mapping techniques suitable for natural environment applications, as well as the need to develop descriptors for 3D point clouds, and highlights the dependence of factor graphs to the non-trivial choice of uncertainty. Future work will include semantic mapping as well as temporal alignment of maps, a great interest for biologists and environmentalists who seek to detect changes in the natural environment over a long period of time. Finally, we aim to improve topological understanding of natural environment point clouds by further exploration of meshing techniques and therefore improve human perception of natural environment elements such as trees.

Declarations

Funding

This work was funded in part by the Grand-Est Region, the Zone Ateliers Moselle, as well as the Rhine-Meuse Water Agency in France.

Conflict of Interest Statement

On behalf of all authors, the corresponding author states that there is no conflict of interest.

Availability of data and material

The dataset will be made public upon the maturity of a cross-seasonal dataset.

Code availability

Not publicly available.

Ethics approval

Not applicable.

Consent to participate

Not applicable.

Consent for publication

Not applicable.

References

1. Babin, P., Dandurand, P., Kubelka, V., Giguère, P., Pomerleau, F.: Large-scale 3D mapping of subarctic forests. In: *Proceedings of the Conference on Field and Service Robotics (FSR)*. Springer Tracts in Advanced Robotics (2019)
2. Besl, P.J., McKay, N.D.: A method for registration of 3-D shapes (1992)
3. Biber, P., Straßer, W.: The normal distributions transform: A new approach to laser scan matching. pp. 2743 – 2748 vol.3 (2003)
4. Bosse, M., Zlot, R., Flick, P.: Zebedee: Design of a spring-mounted 3-D range sensor with application to mobile mapping. *IEEE Transactions on Robotics* **28**(5), 1104–1119 (2012)
5. Cadena, C., Carlone, L., Carrillo, H., Latif, Y., Scaramuzza, D., Neira, J., Reid, I., Leonard, J.J.: Past, present, and future of simultaneous localization and mapping: Toward the robust-perception age. *IEEE Transactions on Robotics* **32**(6), 1309–1332 (2016)
6. Chahine, G., Pradalier, C.: Survey of monocular slam algorithms in natural environments. In: *2018 15th Conference on Computer and Robot Vision (CRV)*, pp. 345–352 (2018)
7. Chahine, G., Pradalier, C.: Laser-supported monocular visual tracking for natural environments. In: *2019 19th International Conference on Advanced Robotics (ICAR)*, pp. 801–806 (2019)
8. Chang, C.C., Chiang, H.: Three-dimensional image reconstructions of complex objects by an abrasive computed tomography apparatus. *The International Journal of Advanced Manufacturing Technology* **22**, 708–712 (2003)
9. Dellaert, F., Kaess, M.: Square root sam: Simultaneous localization and mapping via square root information smoothing. *The International Journal of Robotics Research* **25**(12), 1181–1203 (2006)
10. Engel, J., Koltun, V., Cremers, D.: Direct sparse odometry. *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2017)
11. Gojcic, Z., Zhou, C., Wegner, J.D., Andreas, W.: The perfect match: 3D point cloud matching with smoothed densities. In: *International conference on computer vision and pattern recognition (CVPR)* (2019)

12. Gravina, R., Alinia, P., Ghasemzadeh, H., Fortino, G.: Multi-sensor fusion in body sensor networks: State-of-the-art and research challenges. *Information Fusion* **35**, 68 – 80 (2017)
13. Hall, D.L., Llinas, J.: An introduction to multisensor data fusion. *Proceedings of the IEEE* **85**(1), 6–23 (1997)
14. Huck, T., Westenberger, A., Fritzsche, M., Schwarz, T., Dietmayer, K.: Precise time-stamping and temporal synchronization in multi-sensor fusion. In: 2011 IEEE Intelligent Vehicles Symposium (IV), pp. 242–247 (2011)
15. Kato, A., Moskal, L.M., Schiess, P., Swanson, M.E., Calhoun, D., Stuetzle, W.: Capturing tree crown formation through implicit surface reconstruction using airborne lidar data (2009)
16. Koide, K., Miura, J., Menegatti, E.: A portable three-dimensional lidar-based system for long-term and wide-area people behavior measurement. *International Journal of Advanced Robotic Systems* **16** (2019)
17. Kümmerle, R., Grisetti, G., Strasdat, H., Konolige, K., Burgard, W.: g2o: A general framework for graph optimization. In: 2011 IEEE International Conference on Robotics and Automation, pp. 3607–3613. IEEE (2011)
18. Liu, L., Amin, M.G.: Performance analysis of gps receivers in non-gaussian noise incorporating precorrelation filter and sampling rate. *IEEE Transactions on Signal Processing* **56**(3), 990–1004 (2008)
19. Pierzchała, M., Giguère, P., Astrup, R.: Mapping forests using an unmanned ground vehicle with 3D LiDAR and graph-SLAM. *Computers and Electronics in Agriculture* **145**(February), 217–225 (2018)
20. Pomerleau, F., Colas, F., Siegwart, R., Magnenat, S.: Comparing ICP variants on real-world data sets: Open-source library and experimental protocol. *Autonomous Robots* **34**(3), 133–148 (2013)
21. Pomerleau, F., Magnenat, S., Colas, F., Liu, M., Siegwart, R.: Tracking a depth camera: Parameter exploration for fast icp. In: 2011 IEEE/RSJ International Conference on Intelligent Robots and Systems, pp. 3824–3829 (2011)
22. Shoemake, K.: Animating rotation with quaternion curves. *SIGGRAPH Comput. Graph.* **19**(3), 245–254 (1985)
23. Tremblay, J.F., Béland, M., Pomerleau, F., Gagnon, R., Giguère, P.: Automatic 3D mapping for tree diameter measurements in inventory operations. In: *Proceedings of the Conference on Field and Service Robotics (FSR)*. Springer Tracts in Advanced Robotics (2019)
24. Wiemann, T., Nuechter, A., Hertzberg, J.: A toolkit for automatic generation of polygonal maps - las vegas reconstruction. In: *ROBOTIK 2012; 7th German Conference on Robotics*, pp. 1–6 (2012)
25. Wu, W., Wang, Z., Li, Z., Liu, W., Fuxin, L.: Pointpwc-net: A coarse-to-fine network for supervised and self-supervised scene flow estimation on 3D point clouds. *arXiv preprint arXiv:1911.12408* (2019)
26. Wu, X., Pradalier, C.: Illumination robust monocular direct visual odometry for outdoor environment mapping. In: 2019 International Conference on Robotics and Automation (ICRA), pp. 2392–2398 (2019)
27. Wu, X., Vela, P., Pradalier, C.: Robust monocular edge visual odometry through coarse-to-fine data association (2019)
28. Yang, H., Shi, J., Carlone, L.: Teaser: Fast and certifiable point cloud registration. *ArXiv abs/2001.07715* (2020)
29. Zhang, J., Singh, S.: LOAM: Lidar odometry and mapping in real-time. *IEEE Transactions on Robotics* **32**(7), 141–148 (2015)