2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)
September 24–28, 2017, Vancouver, BC, Canada

Stable Laser Interest Point Selection for Place Recognition in a Forest

Matthew Giamou, Yaroslav Babich, Golnaz Habibi and Jonathan P. How

Abstract— **Place recognition is an essential part of robot localization and mapping problems. Using lower data-rate sensors like 2D scanning laser rangefinders enables the robots to use less memory and computation in building maps. However, place recognition by a vehicle with 6-DOF dynamics like a quadrotor in unstructured, 3D environments like forests is challenging, especially with a sensor that only measures a planar slice of the environment. This paper extends the 2D geometry-based place recognition system of [1] to a challenging forest envirnoment with a novel procedure for selecting stable and salient 2D laser interest points using Dirichlet process clustering (DP-means). This method is tested on both synthetic and real data from a forest trail and compared with [1]. The result reveals the importance of salient interest point selection in allowing accurate and fast place recognition. Our approach also ensures a low bandwidth representation of visited areas, making it suitable for real-time, multi-agent SLAM applications.**

I. INTRODUCTION

Autonomous mobile robots possess great potential for assisting humans in challenging scenarios like wilderness search and rescue. Agile quadrotors equipped with onboard sensing and autonomy can navigate and map challenging GPS-denied environments such as building interiors and dense forests. Using a fleet of multiple autonomous vehicles in a large environment can improve system performance, but demands robust localization and mapping to correct for drift and merge maps.

Place recognition is a key component of any SLAM system and has a rich literature for single robot systems equipped with stereo or monocular cameras [2]–[5]. However, these approaches typically require access to multiple images which would be costly to communicate over a limited bandwidth communication network. Recently, a number of techniques inspired by vision-based place recognition have been applied to 2D lidar data to facilitate real-time loop closure on large datasets spanning 100s of meters [1], [6], [7]. The low data rate of 2D lidar measurements when compared with camera images and visual features makes these systems promising for multi-agent SLAM and exploration systems [8], [9]. However, these approaches have only been applied to datasets that are either indoors, outside of buildings, or in fairly simple outdoor environments (such as the Victoria Park dataset [10]). The case of wilderness search and rescue, which can occur in heavily forested regions with varying topography and irregularly shaped and positioned trees, poses an inherent difficulty for place recognition techniques that perform well in more structured environments. Motivated by the problem of rapidly exploring and mapping an expansive and dense forest, this paper presents an approach to multi-agent place recognition that is: 1- functional in a variety of natural and man-made environments, 2- lighting and viewpoint invariant, 3- operational over a low-bandwidth network, and 4- suitable for real-time applications. The system presented in this paper is an extension of the approach in [1]. Their work is extended by adding a novel interest point stabilization procedure for 2D laser data using DP-means clustering [11] that allows geometric loop closing techniques to work in difficult forest environments. A new 281 m forest trail dataset collected at Middlesex Fells Reservation exhibits the usefulness of our technique.

Fig. 1: Image of Middlesex Fells Reservation, a forested park north of Boston. Varying topography, tree angles, tree size, occlusions and strong perceptual aliasing make place recognition and subsequent loop closure challenging for mobile robots. Additionally, trails are usually traversed in one of two opposing directions, limiting the effectiveness of perspective cameras for place recognition.

The rest of the paper is organized as follows: related work is summarized in Section II. The baseline place recognition system of [1] is explained in Section III. Section IV describes stable interest point selection via RANSAC and DP-means clustering, constituting our main contribution. Experimental setup and results are discussed in Section V, and the paper is concluded in Section VI.

II. RELATED WORK

Place recognition is a large and active research problem. For vision based systems, bag-of-words (BoW) approaches have had notable success. In particular, the FAB-MAP 2.0 algorithm presented in [2] achieves excellent precision and

Aerospace Controls Laboratory, Massachusetts Institute of Technology
{mgiamou, ybabich, golnaz, jhow}@mit.edu

recall for very large datasets. FAB-MAP is extended to incorporate inter-feature geometry in [3], improving performance. The popular DBoW2 algorithm presented in [5] uses binary features to achieve state of the art speed and accuracy. However, most visual BoW approaches use geometric verification steps that would require the transmission of large amounts of image feature data. Additionally, they are not typically not viewpoint invariant unless omnidirectional cameras are used. The high bandwidth requirements are addressed with a distributed vocabulary in [12], but this does not reduce the information exchange by the extent required for networks with transmission rates on the order of 10 kb/s. In [4], Milford demonstrates visual route recognition with heavily compressed and downsampled images but requires either a omnidirectional camera or long overlapping trajectories in the same orientation for a perspective camera equipped robot, both of which can be prohibitive for some robots and missions.

Perceptual aliasing makes place recognition particularly challenging in forest environments. Bosse and Zlot [13], [14] developed a keypoint voting nearest neighbor algorithm to find loop closure in forest settings. Although their result is promising and finds loop closures in sublinear time, they have used 3D laser data to extract 3D regional point descriptors which requires a 3D laser and high-dimensional keypoint representations. This makes their approach prohibitive for light aerial vehicles with limited communication. The introduction of fast laser interest region transform (FLIRT) features in [15], inspired by scale and rotation invariant features in images, introduced a number of 2D laser-based loop closure techniques. A BoW approach using FLIRT features is developed in [6] and extended in [16]. These algorithms use the viewpoint-invariance of features and their ordering in a 2D lidar's angular scan to generate loop closure candidates with histograms of word frequencies. This approach, however, suffers from the limited expressiveness of 2D laser data in comparison with rich vocabularies built from higher dimensional vision data. This is especially true for forest environments, where individual trees have similar and nondescript shapes. Similarly to [3], the geometry of keypoint locations is directly used to build Geometric LAndmark RElations (GLARE) signatures in [7]. These GLARE descriptors are translation invariant and show superior performance over laser-based BoW techniques on publicly available datasets [1].

The work by Dong et al [9] presents a real-time multi-agent SLAM solution using FLIRT descriptors in approximate nearest neighbor-based histogram voting. Although this approach is less accurate than methods using inter-feature geometry like GLARE [1], they improve performance using robust outlier rejection in their back-end SLAM system. Their method required network bandwidth that could support each agent broadcasting at 25 kB/s, limiting operational range and reliability. Geometric descriptors like GLARE can mitigate this requirement by transmitting only the coordinates of a sparse subset of interest points. To this end, we extend the work in [1] by adding a stable interest point extraction procedure in our place recognition pipeline (see Figure 2). This enables us to choose stable interest points by using DP-means clustering. The results show improvement in speed and performance on a forest data set collected at Middlesex Fells Reservation.

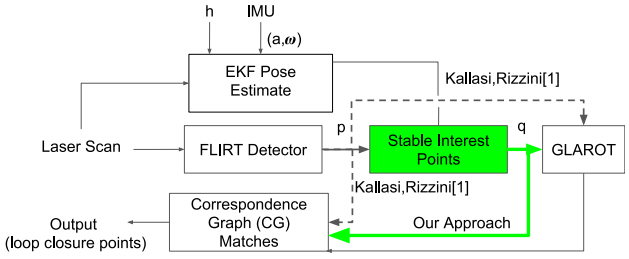

Fig. 2: Block diagram of place recognition. Our main contribution is to add stable interest point extraction module (green block). Only stable interest features are chosen from FLIRT output and are fed to GLAROT module (green line). Our work is an extension of [1], where the FLIRT output is directly fed to GLAROT module (dashed lines). The EKF block is the extended Kalman filter-based localization of [18].

III. PLACE RECOGNITION

The problem of place recognition is posed as accurately associating a stream of 2D laser scans $\mathcal{S}^t$ of an environment with previous measurements of the same location stored in a database $\mathcal{D}$. This database can be on board the vehicle or accessed remotely via a wireless network.

Figure 2 depicts the place recognition pipeline. Algorithm 1 describes the approach, which is an extension of the system found in [1]. The input of the algorithm at time t is the altitude h^t from an altimeter, body frame acceleration and angular velocity, (a^t, ω^t), which are estimated by an IMU, and 2D laser scans $\mathcal{S}^t$. First, a set of interest points $\{p_i^t\}$ is extracted from each incoming laser scan $\mathcal{S}^t$ by using FLIRT (line 8). In [1] these points are immediately used to form a GLARE signature G^t which summarizes the interest point geometry of scan $\mathcal{S}^t$ and allow a fast search of $\mathcal{D}$ for 20 candidate scans $\{\mathcal{S}'^t\}$ (line 13). We adds a 'stable interest point (SIP)' block as our contribution (line 12). This procedure generates a stable and salient set of points $\{q_i^t\}$ from a window of recent scans $\{\mathcal{S}^k\}_{k=t-T}^{t}$ for some length of time T (1 second in this work). This set of points is then used in the remainder of the pipeline instead of $\{p_i^t\}$. The interest points from these candidate scans are then checked using a correspondence graph (CG) matching procedure [17] (line 15). A match $(\mathcal{S}^*)^t \in \{\mathcal{S}'^t\}$ is accepted as a recognized place if the number of matching interest points exceeds some threshold N_{th} (line 15-17).

A. Laser Interest Point Detection

Modern laser interest point algorithms include FLIRT [15] and, more recently, FALKO [19]. Both consist of a configurable detector and descriptor pairing. Inspired by scale invariant feature detection in computer vision, FLIRT

Algorithm 1 Place Recognition Algorithm

$t' \leftarrow 0$
while true **do**
  Input: a^t, ω^t, h^t, S^t
  $P^t \leftarrow \varnothing$
  $(S^*)^t \leftarrow \varnothing$
  $S' \leftarrow \varnothing$
  $X^t \leftarrow \text{EKF}(a^t, \omega^t, h^t, S^t)$
  $\{p_i^t\} \leftarrow \text{FLIRT}(S^t)$
  $P^t \leftarrow \{p_i^t\}$
  if $t' \geq T$ **then**
    $t' \leftarrow 0$
    $\{q_j^t\} \leftarrow \text{StableInterestPoints}(P^{t-T:t}, X^t)$
    $\{s'\} \leftarrow \text{GLAROT}(\{q_j^t\})$
    for s'^k in S' **do**
      if $\text{CG}(\{q_i^t\}, s'^k) \geq N_{th}$ **then**
        $(S^*)^t.add(s'^k)$
      end if
    end for
  end if
  $t' \leftarrow t' + 1$
  Output: $(S^*)^t$
end while

Algorithm 2 RANSAC-based Stability Test

Input: $M, \{\mathcal{S}\}_{i=t-M}^{t}$
$\{p^t\} \leftarrow \text{FLIRT}(\mathcal{S}^t)$
$i \leftarrow 1$
while $i < M$ and $\{p^t\}$ not empty **do**
  $\{p^t\} \leftarrow \text{RANSAC}(\{p^t\}, \mathcal{S}^{t-i})$
  $i \leftarrow i + 1$
end while
Output: $\{p^t\}$

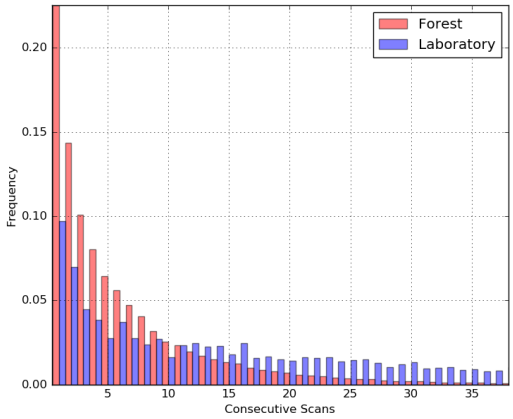

Fig. 3: Distribution of stability of FLIRT interest points on laser scans from Middlesex Fells reservation. The x-axis is the average number of consecutive scans a feature remains in an inlier set computed with RANSAC. The indoor data contains corners and other features that are highly structured and more uniform in the vertical direction.

detects interest points in either the range, local normal, or curvature of a scan at multiple geometric scales. The FLIRT descriptors are grids capturing the shape local to their corresponding interest point. FALKO detects interest points that can be fitted by intersecting lines describing corners. FALKO is designed to be more orientation and scan density invariant than FLIRT and outperforms FLIRT on many datasets in [1]. However, since a forest environment's most salient features are trees of varying sizes, using FLIRT range interest points outperformed FALKO. This is due to FLIRT range detection's ability to easily capture the sharp range discontinuities caused by trees, which do not closely resemble corners.

Even with its multi-scale detection's ability to pick out salient features in a forest, FLIRT features are not very stable in our Middlesex Fells Reservation dataset, pictured in Figure 1. To measure stability, we performed a consecutive RANSAC matching procedure on FLIRT interest points $\{p_i^t\}$ from scan $\mathcal{S}^t$ against the previous M scans. The RANSAC stabilization performed by Algorithm 2 selects a set of inlier points denoted $\{p_i^t\}_1$ that best agree with a 2D transformation between scan $\mathcal{S}^t$ and $\mathcal{S}^{t-1}$. This procedure is then repeated between $\{p_i^t\}_1$ and points $\{p_i^{t-2}\}$ from $\mathcal{S}^{t-2}$ until the "surviving" points $\{p_i^t\}_M$ are obtained or there are no inliers remaining. The normalized histogram in Figure 3 visualizes the distribution of interest point lifetimes for data collected in the forest and data collected in an indoor flight laboratory. The mean lifetime for forest data interest points is 6.32 consecutive scans, whereas for indoor data the mean is 21.26 scans.

B. GLARE Signatures

Geometric landmark relations (GLARE) were introduced in [7] as signatures for capturing viewpoint-invariant geometric properties in a set of interest points. Given a set of interest points $\{p_i^t\}$ extracted from a scan $\mathcal{S}^t$, a GLARE signature is essentially a distribution over the pairwise euclidean distance $\rho_{i,j}^t$ and angular distance $\theta_{i,j}^+$ between pairs. Figure 4 depicts this process: for every pair of points (p_i^t, p_j^t), their distance and relative angle in the scan frame are computed as

$$\rho_{i,j}^t = \left\| p_i^t - p_j^t \right\|, \tag{1}$$
$$\theta_{i,j}^t = \text{atan2}(y_i^t - y_j^t, x_i^t - x_j^t), \tag{2}$$
$$\theta_{i,j}^+ = \max(\theta_{i,j}^t, \theta_{j,i}^t). \tag{3}$$

Once the pairs $(\theta_{i,j}^+, \rho_{i,j}^t)$ have been computed, they are assigned to bins in a histogram

$$(\theta_{i,j}^+, \rho_{i,j}^t) \in \text{bin}(n_\theta, n_\rho), \tag{4}$$

where n_θ and n_ρ are the integers corresponding to the quantization of $(\theta_{i,j}^+, \rho_{i,j}^t)$ in the range $[0, \rho_{max}]m \times [0.0, \pi]rad$. To account for noise in the interest point detection, a weight drawn from a Gaussian distribution is computed for cell $n = (n_\theta, n_\rho)$ and adjacent cells. Thus, each feature pair

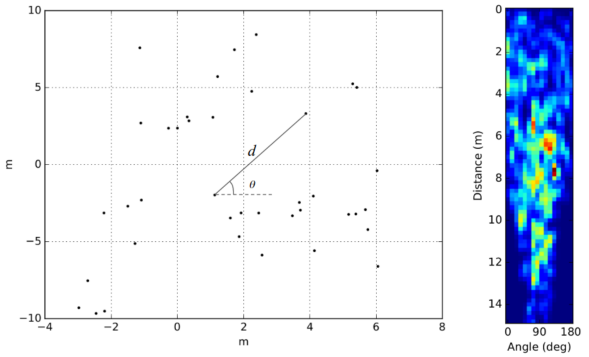

Fig. 4: GLARE signature (right) for a set of FLIRT interest points (left) extracted from a laser scan collected at Middlesex Fells Reservation. Hotter colors represent higher frequency of point pairs at the distance and angular offset indicated by a bin.

produces a histogram matrix $H_{i,j} \in \mathbb{R}^{N_\rho \times N_\theta}$ with values

$$H_{i,j}(m) = \mathcal{N}(m - n, \Sigma_H), \tag{5}$$

where $\Sigma \in \mathbb{R}^{2\times 2}$ is a tunable covariance matrix describing the noise. The glare descriptor G^t is defined as

$$G^t = \eta \sum_{i,j} H_{i,j}, \tag{6}$$

where η is a normalizing factor. As noted in [1], GLARE features are invariant to rigid rotations of the constituent interest points up to circular shifting of columns of G^t. Thus, their GLAROT (GLAre ROTation-invariant) procedure is used to compare two GLARE features. This defines the distance between two signatures G^s and G^t as the *shifted* L_1 *norm*, SL_1:

$$SL_1(G^s, G^t) = \min_{0\le k<N_\theta} \sum_{i=0}^{N_\rho-1} \sum_{j=0}^{N_\theta-1} |G^t_{i,j} - G^s_{i,(j+k) \bmod N_\theta}|). \tag{7}$$

This procedure involves $N_\rho \times N_\theta^2$ pairs of fast array lookups, subtractions, and absolute value operations per query. This allows SL_1 to be efficiently computed over the GLARE features of the database of previously encountered scans $\mathcal{D}$ and the GLARE feature of a query scan $\mathcal{S}$, selecting the N previously measured scans that have the least GLAROT distance with the query as loop closure candidates. In experiments, the GLAROT search procedure never exceeded 4 ms of runtime with N_ρ=120 and N_θ=12 over a maximum of 200 queries.

C. Correspondence Graph Matching

Once a set of N scans has been retrieved by the GLAROT procedure, a correspondence graph (CG) matching procedure is used to find the 2D geometric transformation producing the largest inlier set. The method in [17] displayed superior performance over a Generalized Hough Transform-based approach and a RANSAC matcher on a number of publicly available datasets in [1].

Given two sets of points $\{p_{i_s}\}$ and $\{p_{i_t}\}$, each vertex v_{i_s,i_t} of the CG represents a hypothetical correspondence between points p_{i_s} and p_{i_t}. Any given pair of vertices v_{i_s,i_t} and v_{j_s,j_t} is connected if

$$||p_{i_s} - p_{j_s}| - |p_{i_t} - p_{j_t}|| < \epsilon \tag{8}$$

for a tolerance ϵ. Once formed, the largest possible correspondence that represents a ϵ-tolerant rigid 2D transformation from a subset of $\{p_{i_s}\}$ to a subset of $\{p_{i_t}\}$ is given by the maximum clique in the CG. To find the maximum clique, the fast *maximum clique dynamic* algorithm presented in [20] is used. Forming the CG and computing the maximum clique for the N candidates retrieved by GLAROT constitutes the most computationally expensive stage of the pipeline.

IV. STABLE INTEREST POINT SELECTION

Datasets collected in highly structured environments such as the Intel or MIT-CSAIL datasets contain features that can be reliably identified in a single laser scan. By contrast, our Middlesex Fells dataset does not produce interest points that are stable enough through consecutive scans. The reasons for this instability include: 1) trees, branches and elements of bushes whose thickness is on the order of laser rangefinder noise and angular resolution, 2) varying topography and objects (trees, bushes, hillsides) with a lot of 3D structure, as opposed to the highly "2.5D" nature of structured environments, 3) platform altitude/attitude that is less stable than data collected on a wheeled vehicle, and 4) evolving occlusions caused by trees at a multitude of ranges from the laser. These effects lead to poor performance in the GLAROT and CG-matching procedures when interest points from single scans are used. To remedy this situation, we propose two methods by which only a stable and salient subset of interest points are chosen.

Our experimental data differs from previous datasets amenable to FLIRT and GLARE in that it was collected on a hand-carried sensor rig subject as opposed to a wheeled vehicle. Since, like a quadrotor in flight, the vehicle is free to roll, pitch, and change its altitude, consecutive scans from the laser rangefinder do not necessarily lie within the same plane. In order to utilize the 2D place recognition techniques, we make the assumption that trees are roughly perpendicular to the xy-plane of the global odometry frame. Coordinates of raw laser measurements are projected from the laser's frame of reference into a frame local to the laser but rotated such that its z-axis is perpendicular to the global odometry frame's z-axis and its x-axis is aligned with the laser frame's x-axis's projection onto the global xy-plane. This transformation is made possible with the EKF-provided odometry measurements. Once transformed into this frame, the points are projected onto to the xy-plane of the frame and FLIRT extraction proceeds on these 2D points.

A. RANSAC-Stabilized FLIRT Interest Points

The first method by which stable interest points are selected is the RANSAC procedure described in section III and Algorithm 2. This reduces the points used to a subset that is more reliably detected across consecutive scans. One problem with this approach is that stable or salient features

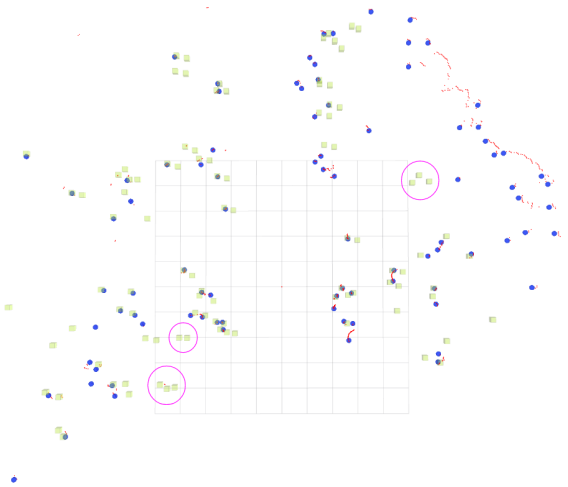

Fig. 5: Interest points taken from a sample scan in the forest. The small red points are raw laser measurements, large blue circles are FLIRT interest points, and the light green boxes are the 100 most stable DP-means cluster centers. The magenta circles identify regions that are not interest points of the current scan due to occlusions or some other source of instability. Using the M largest cluster centers in a 1s window of time as interest points keeps the descriptiveness of these points.

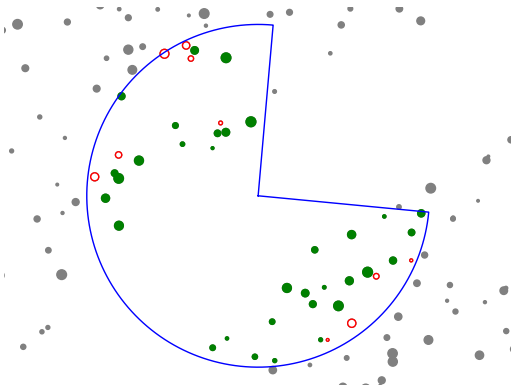

Fig. 6: A scan of the simulated forest. Solid green circles are detected trees, hollow red circles are occluded trees, and grey circles are trees outside of the scanner's range. The trees are distributed via a 2D Poisson point process. Tree radii are drawn from the uniform distribution over 0.05 m to 0.2 m. Radial Brownian motion is used to simulate the vehicle moving closer to trees at certain points in its trajectory, increasing the occluded regions' areas.

that are temporarily not detected due to interruption from rolling/pitching of the sensor rig, brief occlusions, or noisy range measurements may be detrimentally omitted.

B. DP-Means Clustering

Since RANSAC-based stabilization requires an interest point to remain in view for multiple consecutive scans, it can unwittingly throw out interest points that show up frequently in the same neighborhood but with some interruptions. To remedy this, we propose clustering interest points in order to select the M clusters containing the most interest points extracted over a 1 second window of scans. Selecting only the clusters with the most points rejects erroneous outliers (e.g. FLIRT interest points caused by sensor noise or unstable 3D elements of the environment). The effects of clustering are depicted in Figure 5, where the circled clusters represent interest points that have gone out of view in the last second.

The DP-means clustering algorithm is preferred over traditional k-means for our problem because instead of requiring the number of clusters k as input, it requires the cluster penalty parameter λ which corresponds to the maximum radius of a cluster. Since the number of salient interest points in a particular sequence of scans is unknown *a priori*, it is hard to choose an intuitive value for k. On the other hand, a forest typically has a range of tree radii that is easy to deduce and whose upper limit is a good value for λ. We use the open source C++ implementation of the dynamic means algorithm [21], which is a batch-sequential extension of DP-means.

V. RESULTS

We tested our system on both synthetic 2D data and on real data collected by a sensor rig carried by hand through a forest. The synthetic experiments establish the viability of geometric place recognition techniques in an environment with randomly distributed point features. All experiments were performed on a 3.4GHz Intel Core i7-4670K processor.

A. Simulation Experiment

A Python simulation was created in order to test the effectiveness of GLARE and CG matching for finding loop closures in a forest environment. A 2D Poisson point process is used in the simulation to generate a uniform random distribution of points representing tree positions [22]. Each point is also given a uniformly sampled radius, as displayed in Figure 6. A simulated robot travels through a cleared circular path with 5.0 m width and 32.5 m radius, while undergoing bounded radial Brownian motion. After completing one circle, it turns around and returns along the same path, forming many candidate loop closures. During this time, the robot collects 100 range scans in a 270° radial field of view to match the hardware experiments in section V-B, with a synthetic normal standard deviation of 0.18 m (close to the diameter of large trees commonly found in the dataset) applied to the range measurements. A sample simulated scan can be seen in Figure 6.

GLAROT signatures for each scan are created and compared to find 10 candidates for CG matching. Finally, CG matching is performed, and the CG threshold clique sizes are varied to give the precision-recall curves in Figure 7. The high precision (ratio of correct loop closures to the total number of posited loop closures) that can be achieved indicates that local tree geometry is sufficiently descriptive to perform localization and mapping. The imperfect recall (fraction of loop closures found) is a result of laser range noise overcoming the CG matching algorithm; running the simulation with no error results in perfect recall. Additionally, some recall is lost due to occlusion, but can be partially

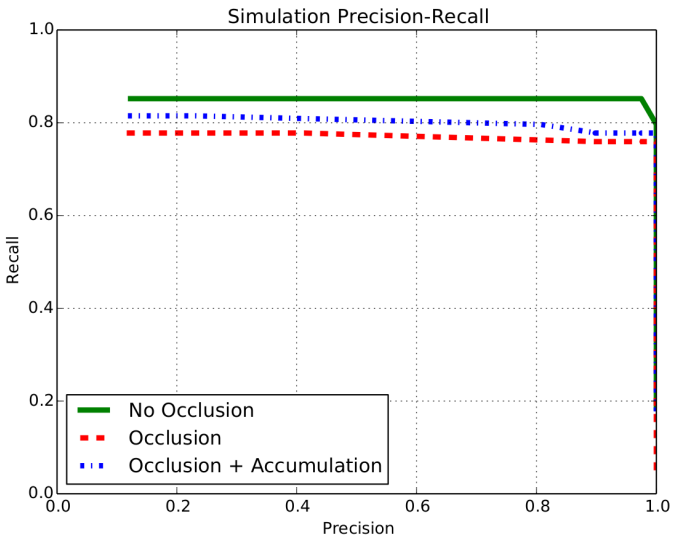

Fig. 7: Precision-recall curves for three scenarios. In the no occlusion scenario, any tree within the scanner range is detected. Occlusion accounts for trees blocking vision of other trees behind. Accumulation uses data from several consecutive laser scans, seeing some previously occluded trees.

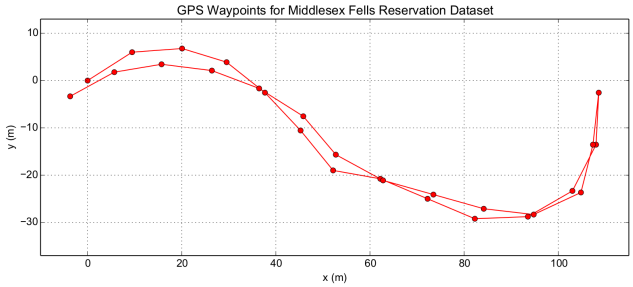

Fig. 8: Sparse GPS localization for Middlesex Fells dataset. The path has a loop with two legs overlapping on the same trail, providing many opportunities for place recognition between. The roughly 180° angular offset between headings on the outbound and inbound leg make the loop closure detections more difficult given that the laser rangefinder's field of view is 270°.

regained by using several consecutive scans to see trees that were previously blocked from vision. This accumulation effect is provided by DP-means as it returns cluster centers from interest points over the previous T scans.

B. Hardware Experiment

For the hardware experiment, a dataset was collected by a sensor rig carried through a forest trail in Middlesex Fells Reservation, pictured in 1. The data was collected in a loop that is essentially an outbound traversal of a path followed by a return that traces the same path with opposite heading. The path was recorded with coarse GPS measurements displayed in Figure 8. The sensor rig consisted of a horizontally mounted Hokuyo UTM-30LX laser rangefinder, a Pixhawk unit providing inertial measurements, a downward-facing LidarLite for altitude measurements, and an Intel NUC computer to collect the data in the form of Robot Operating System (ROS) messages. The Hokuyo produced laser measurements at a rate of 40 Hz over an angular field of view of 270° with 0.25° angular resolution. The inertial measurements and LidarLite measurements were processed at 100 Hz. The laser was manually stabilized as best as possible but still experienced significant pitch and roll due to natural oscillations in the carrier's gait, occasionally exacerbated by rough terrain and changing topography. Future work will utilize quadrotor flight in order to characterize the significance of vehicle dynamics on performance.

To detect ground-truth loop closures, we estimated the rig's trajectory using the extended Kalman filter (EKF) described in [18] on 2D laser scan matching, LidarLite measurements h, and IMU measurements of acceleration a and angular velocity ω. The collected data was then used offline to simulated loop closure using the methods described in sections III and IV. The trajectory was "played back" in sequence, allowing the algorithm to accumulate a database $\mathcal{D}$ of scans with which to compare incoming scans. The precision recall-curves in figure 9 are generated by considering all correspondence graph matches with a number of matching pairs of points above a variable threshold N_{th} as positive matches. Candidates for correspondence graph matching are gathered with the 20 smallest GLAROT signature distances. Place recognition is attempted once per second, and multiple false positives from a single recognition attempt (i.e. erroneous matches that have more than N_{th} points) all contribute to the precision computation. Since the instability of features means that FLIRT features chosen from a single scan are sensitive to the set of features used, 4 runs with randomly offset starting points in the first second of data were used and averaged to create the precision-recall curves of Figure 9. The starting time also slightly affected the DP-means approach at high values of N_{th} (i.e. high precision). This is because DP-means cluster locations are sensitive to the order in which data is processed, and a random seed is used to shuffle data input order in our implementation [21]. Selecting parameters or augmenting the algorithm to reduce this instability is an important avenue for future work that can be tested with larger datasets.

In Figure 9, we observe that choosing only FLIRT points that survive at least 1 round of RANSAC does not have a strong effect on the resultant precision and recall. However, table I indicates that this culling of superfluous points leads to a mean runtime over 6 times faster. A speed-up is also exhibited with using the M=70 largest DP-means cluster centers instead of raw FLIRT interest points, with significant precision-recall performance improvements. Figure 11 visualizes the increased number of valid correspondences that are discovered by the DP-means approach at N_{th}=12 over using the unfiltered FLIRT features. The faster runtime and higher recall values at precisions greater that 0.5 of DP-means feature stabilization could enable single or multi-agent systems with robust SLAM backends to perform in real time as in [9]. Additionally, sharing only 70 points over a period of 1-2 seconds uses far less bandwidth than that system.

TABLE I: Runtime and Number of Points Used

Algorithm	Runtime: μ (Max.)	No. of Points μ (Max.)
Baseline [1]	2.43 (7.55)	75.61 (119)
RANSAC 2	0.35 (1.53)	44.58 (88)
70 DP-means Clusters	1.13 (1.55)	70 (70)

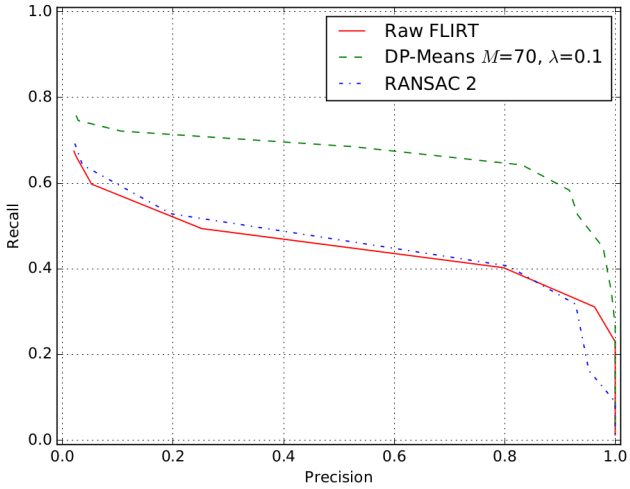

Fig. 9: Precision-recall curve for Middlesex Fells path 1. Using un-clustered FLIRT features from a single scan (red curve) leads to much worse performance than choosing the N largest clusters from a DP-means clustering of 1 second's worth of scans.

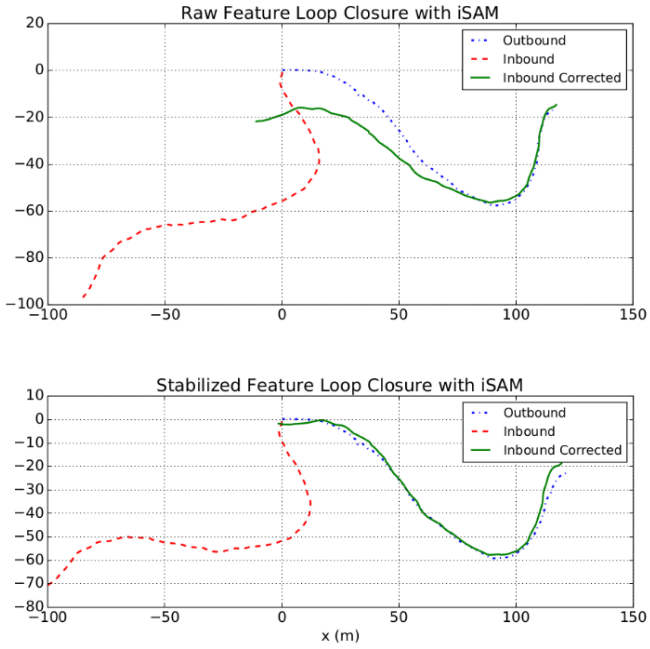

Fig. 10: Multi-agent pose SLAM results using iSAM on EKF odometry and detected loop closures for N_{th}=18, a threshold level with no false positives for either unstabilized or stabilized features. The loop closures found with unstabilized (top) and stabilized features (bottom) were used to produce a two agent pose graph with an unknown relative transform between starting reference frames. At this very conservative threshold, the stabilized method is able to find a greater number of correct loop closures, greatly improving the alignment of outbound and inbound trajectories.

VI. CONCLUSION

This paper presents an extension of the work in [1] for use in a challenging forest environment on a vehicle with 3D dynamics. Recent geometrical place recognition techniques are shown in simulation and on real world data to be useful for place recognition in a forest, provided a stable and salient subset of laser interest points are selected to describe a place. The DP-means clustering algorithm is exhibited as a fast and intuitive to tune way to select a small but descriptive set of points to represent a place on a forest trail.

Our results indicate that 2D laser-based place recognition techniques are able to identify loop closures with reasonable precision and recall in an environment with lots of 3D structure. This is a potentially useful strategy for UAVs with limited sensors, computing power, storage, or communication with a base station or other vehicles in a multi-agent SLAM system. Specifically, we show that a place in a forest can be summarized by 70 or fewer 2D points and compared against a database at almost 1 Hz. This is an order of magnitude reduction on the number of points in a single scan of the Hokuyo UTM-30LX used (1080 points), and far less than image features, which are commonly on the order of 100s of kB per query [12]. Future work can leverage this approach on real time multi-agent systems in large scale mapping and localization missions.

ACKNOWLEDGMENTS

This work was supported by the NASA Convergent Aeronautics Solutions project Design Environment for Novel Vertical Lift Vehicles (DELIVER). Thanks to Loc Tran, Kyel Ok, and Pon Jirachuphun for their invaluable help collecting data.

REFERENCES

[1] F. Kallasi and D. L. Rizzini, "Efficient loop closure based on FALKO lidar features for online robot localization and mapping," in *Intelligent Robots and Systems (IROS), 2016 IEEE/RSJ International Conference on*. IEEE, 2016, pp. 1206–1213.

[2] M. Cummins and P. Newman, "Appearance-only slam at large scale with FAB-MAP 2.0," *The International Journal of Robotics Research*, vol. 30, no. 9, pp. 1100–1123, 2011.

[3] R. Paul and P. Newman, "FAB-MAP 3d: Topological mapping with spatial and visual appearance," in *Robotics and Automation (ICRA), 2010 IEEE International Conference on*. IEEE, 2010, pp. 2649–2656.

[4] M. Milford, "Visual route recognition with a handful of bits," *Proc. 2012 Robotics: Science and Systems VIII*, pp. 297–304, 2012.

[5] D. Gálvez-López and J. D. Tardos, "Bags of binary words for fast place recognition in image sequences," *IEEE Transactions on Robotics*, vol. 28, no. 5, pp. 1188–1197, 2012.

[6] G. D. Tipaldi, L. Spinello, and W. Burgard, "Geometrical FLIRT phrases for large scale place recognition in 2d range data," in *Robotics and Automation (ICRA), 2013 IEEE International Conference on*. IEEE, 2013, pp. 2693–2698.

[7] M. Himstedt, J. Frost, S. Hellbach, H.-J. Böhme, and E. Maehle, "Large scale place recognition in 2d lidar scans using geometrical landmark relations," in *Intelligent Robots and Systems (IROS 2014), 2014 IEEE/RSJ International Conference on*. IEEE, 2014, pp. 5030–5035.

[8] E. Olson, J. Strom, R. Goeddel, R. Morton, P. Ranganathan, and A. Richardson, "Exploration and mapping with autonomous robot teams," *Communications of the ACM*, vol. 56, no. 3, pp. 62–70, 2013.

[9] J. Dong, E. Nelson, V. Indelman, N. Michael, and F. Dellaert, "Distributed real-time cooperative localization and mapping using an uncertainty-aware expectation maximization approach," in *2015 IEEE*

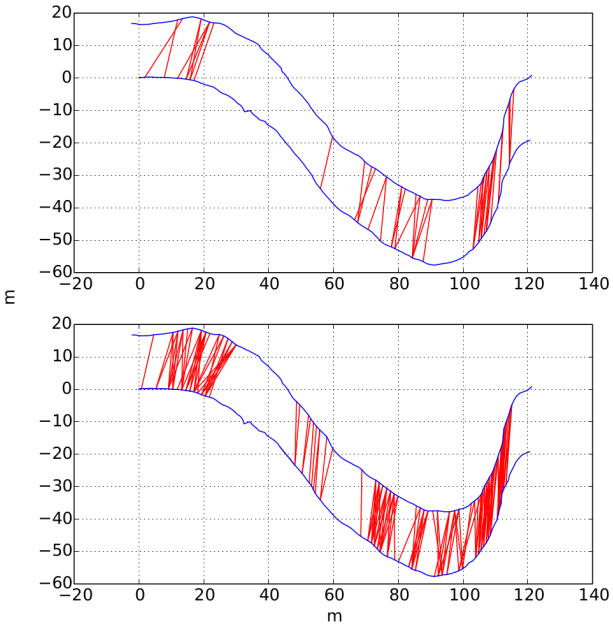

Fig. 11: Correspondences (red lines) detected by raw FLIRT features (top figure) and DP-means clustered FLIRT features (bottom figure) for N_{th}=12. The outbound (bottom blue curve) and inbound (top blue curve) path in each sub-figure has been separated to make correspondences more visible.

International Conference on Robotics and Automation (ICRA). IEEE, 2015, pp. 5807–5814.

[10] J. Guivant, J. Nieto, and E. Nebot, "Victoria park dataset," 2012.

[11] B. Kulis and M. I. Jordan, "Revisiting k-means: New algorithms via bayesian nonparametrics," *arXiv preprint arXiv:1111.0352*, 2011.

[12] T. Cieslewski and D. Scaramuzza, "Efficient decentralized visual place recognition using a distributed inverted index," *IEEE Robotics and Automation Letters*, vol. 2, no. 2, pp. 640–647, 2017.

[13] M. Bosse and R. Zlot, "Place recognition using regional point descriptors for 3d mapping," in *Field and Service Robotics*. Springer, 2010, pp. 195–204.

[14] ——, "Place recognition using keypoint voting in large 3d lidar datasets," in *Robotics and Automation (ICRA), 2013 IEEE International Conference on*. IEEE, 2013, pp. 2677–2684.

[15] G. D. Tipaldi and K. O. Arras, "FLIRT-interest regions for 2d range data," in *Robotics and Automation (ICRA), 2010 IEEE International Conference on*. IEEE, 2010, pp. 3616–3622.

[16] J. Deray, J. Sola, and J. Andrade-Cetto, "Word ordering and document adjacency for large loop closure detection in 2d laser maps," *IEEE Robotics and Automation Letters*, 2017.

[17] T. Bailey, E. M. Nebot, J. Rosenblatt, and H. F. Durrant-Whyte, "Data association for mobile robot navigation: A graph theoretic approach," in *Robotics and Automation, 2000. Proceedings. ICRA'00. IEEE International Conference on*, vol. 3. IEEE, 2000, pp. 2512–2517.

[18] M. Achtelik, A. Bachrach, R. He, S. Prentice, and N. Roy, "Stereo vision and laser odometry for autonomous helicopters in gps-denied indoor environments," in *SPIE Defense, security, and sensing*. International Society for Optics and Photonics, 2009, pp. 733 219–733 219.

[19] F. Kallasi, D. L. Rizzini, and S. Caselli, "Fast keypoint features from laser scanner for robot localization and mapping," *IEEE Robotics and Automation Letters*, vol. 1, no. 1, pp. 176–183, 2016.

[20] J. Konc and D. Janezic, "An improved branch and bound algorithm for the maximum clique problem," *proteins*, vol. 4, no. 5, 2007.

[21] T. Campbell, M. Liu, B. Kulis, J. P. How, and L. Carin, "Dynamic clustering via asymptotics of the dependent dirichlet process mixture," in *Advances in Neural Information Processing Systems 26*, 2013.

[22] E. Tomppo, *Models and methods for analysing spatial patterns of trees*. Metsäntutkimuslaitos, 1986.