Automatic Loop Closure Detection Using Multiple Cameras for 3D Indoor Localization

Nicholas Corso, John Kua, Jacky Chen, Avideh Zakhor
Video and Image Processing Lab, University of California, Berkeley
{ncorso, jkua, jackyc, avz}@eecs.berkeley.edu

Abstract

Automated 3D modeling of building interiors is useful in applications such as virtual reality and environment mapping. We have developed a human operated backpack data acquisition system equipped with a variety of sensors such as cameras, laser scanners, and orientation measurement sensors to generate 3D models of building interiors, including uneven surfaces and stairwells. An important intermediate step in any 3D modeling system, including ours, is accurate 6 degrees of freedom localization over time. In this paper, we propose two approaches to improve localization accuracy over existing methods. First, we develop an adaptive localization algorithm which takes advantage of the environment's floor planarity whenever possible. Secondly, we show that by including all the loop closures resulting from two cameras facing away from each other, it is possible to significantly reduce localization error in scenarios where parts of the acquisition path is retraced. We experimentally characterize the performance gains due to both schemes.

1. Introduction

In recent years, three-dimensional modeling has attracted much interest due to its wide range of applications such as virtual reality, disaster management, virtual heritage conservation, and mapping of potentially hazardous sites. Manual construction of these models is labor intensive and time consuming; as such, methods for automated 3D site modeling have garnered much interest.

An important component of any 3D modeling system is localization of the data acquisition system over time and space. Localization has been studied by the robotics and computer vision communities in the context of the simultaneous localization and mapping problem (SLAM). Recently much work has been done toward solving the SLAM problem with six degrees of freedom (DOF) [2, 14, 3], i.e. position and orientation. SLAM approaches with laser scanners typically rely on scan matching algorithms such as Iterative Closest Point (ICP) [12] to align scans from two poses in order to recover the transformation. In addition, recent advances in visual odometry algorithms have led to camera-based SLAM approaches [14, 7].

Localization in indoor environments is particularly challenging since GPS is unavailable inside buildings. In addition, 3D modeling of complex environments such as stairwells precludes the use of wheeled acquisition systems. To overcome this, a human operated backpack system equipped with a number of laser scanners, cameras, and an orientation measurement system has been developed to both localize the system and to construct geometry and texture for 3D indoor modeling [6, 10]. In [6], a set of localization algorithms for recovering all six DOF over time has been proposed and characterized over a 60 meter loop on a 30 meter hallway using manually detected loop closure (LC) events. In doing so, it has empirically been found that localization error is significantly reduced in situations where (a) the floor is planar, and (b) localization algorithms are designed to take advantage of the planarity. While such a localization algorithm is inapplicable to scenarios with stairwells or uneven surfaces, it can be applied to portions of the data acquisition path in which the planarity assumption does hold true. Thus, the challenge lies in classifying the acquisition path into planar and non-planar segments and to apply the appropriate localization algorithm to each portion. In this paper we develop such an adaptive localization algorithm and show that it can improve localization error in complex mixed environments made of both planar and non-planar floors.

Due to various process errors and sensor biases, localization based on any combination of scan matching, visual odometry, and wheel odometry can result in significant drifts in navigation estimates over time. Often this error becomes apparent when the acquisition system visits a landmark or traverses a loop. In the case of revisiting a previous location, the estimated trajectory from a localization algorithm may not form a perfect loop. This type of inconsistency can be remedied by detecting LC events and solving

an optimization problem to reduce the error [14, 13, 8, 9].

In general, finding LC events is a non-trivial task; accumulated localization error causes naïve detection schemes to miss them due to large errors in position estimates. Image data can be used to detect LCs independent of the current position estimate. Recently, a two step algorithm has been developed for automatic image based LC detection from a single camera for an indoor modeling system [10]; the first step, which is based on FAB-MAP [13], results in a rank ordered list of candidate image pairs. The list of image pairs is processed in the second step using keypoint matching [11] to filter out the erroneous candidates.

For image based LC detection using one camera, both the position and orientation of the camera, and hence the acquisition system, need to be similar during a revisit with a given location. However, with two side-looking cameras pointing 180° away from each other, it is conceivable to detect LC events *across* two cameras provided during the revisit the system is about 180° away from its initial yaw orientation. This condition is satisfied in situations where a system traverses up and down a hallway or a stairwell. Therefore in practice, it is possible to detect a large number of LCs by matching images from one side-looking camera while traversing up the hallway or stairwell, to images from the opposing side-looking camera while traveling in the opposite direction: the larger the amount of overlap in the retraced path, the larger the number of such LCs. In this paper, we apply the automatic LC detection algorithm of [10] to images from two opposite facing side cameras on a human operated backpack system in order to detect such LCs. In doing so, we show that the increased number of LCs in these scenarios results in significant reduction in 6 DOF localization error.

The outline of the paper is as follows. The architecture and conventions of our backpack system is described in Section 2. In Section 3, the adaptive algorithm for mixed paths with both planar and non-planar floors is discussed and evaluated. In Section 4, we describe image based LCs for the two side-looking cameras, and characterize its localization error. Section 5 contains the localization performance by combining the approaches in Sections 3 and 4. The conclusions are in Section 6.

2. Architecture and Conventions

We mount four 2D laser range scanners, two cameras, an orientation sensor, and an IMU onto a backpack system, which is carried by a human operator. Figure 1 shows the CAD model of such a system. The yaw scanner is a 40Hz Hokuyo UTM-30LX 2D laser scanner with a 30-meter range and a 270° field of view. Both the pitch and the side-looking vertical geometry scanners are 10Hz Hokuyo URG-04LX 2D laser scanners with a 4-meter range and a 240° field of view. These scanners are mounted orthogonally to

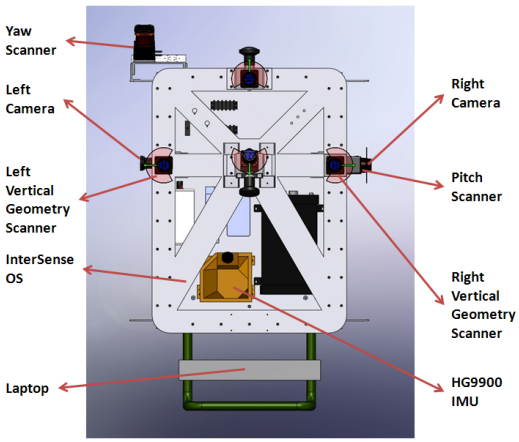


Figure 1. CAD model of the backpack system.

one another. The two cameras are Point Grey Grasshopper GRAS-14S5C units equipped with 5mm lenses, resulting in a 82°x67° field of view. The IMU, a Honeywell HG9900, is a strap-down navigation-grade sensor which combines three ring laser gyros with bias stability of less than 0.003°/hour and three precision accelerometers with bias of less than 0.245mm/sec^2. The HG9900 provides highly accurate measurements of all six DOF at 200Hz and thus serves as our ground truth. The orientation sensor (OS), an InterSense InertiaCube3, provides orientation parameters at a rate of 180Hz. As seen later, only the yaw and pitch scanners, and the InterSense OS are used to localize the backpack in all the localization algorithms discussed in this paper. In particular, the position of the system is estimated at a rate of 10Hz, the same rate as the pitch scanner. The left and right cameras are used to detect LC events while the side-looking vertical geometry scanners are only used to construct geometry.

Throughout this paper we assume a right-handed coordinate system. With the backpack system worn upright the x axis is forward, the y axis is leftward, and the z axis is upward. As shown in Figure 1, the yaw scanner scans the x-y plane, the pitch scanner scans the x-z plane, and the vertical geometry scanner scans the y-z plane. Thus, the yaw scanner can resolve yaw rotations about the z axis

3. Adaptive Localization

In this section, we begin by reviewing two of the 6 DOF localization algorithms originally developed in [6]. These algorithms combine scan matches from orthogonal scanners and OS data to recover the 6 DOF transformation from the pose at time t_1 to the pose at time t_2. Integrating the recovered transformations, an estimated trajectory for the backpack can be obtained. Due to process errors and sensor biases, the estimated transformation is somewhat erroneous as the error grows large over long trajectories. Once

LC events are known, they are enforced using a nonlinear optimization technique, the Tree-based netwORk Optimizer (TORO), to reduce localization error [6, 9]. In Section 3.1, we review the 2×ICP+OS method [6] for transformation recovery without a priori knowledge about the scene environment. This algorithm is applicable to paths with planar or non-planar floors. In Section 3.2, we review the localization algorithm 1×ICP+OS+Planar [6], which is based on the floor planarity assumption. In Section 3.3, we introduce a new algorithm which adaptively switches between the two localization algorithms by automatically segmenting the traversed path into planar and non-planar segments. Such an algorithm is useful in mixed environments with both planar and non-planar floors such as staircases. Specifically, it enjoys the low localization error of 1×ICP+OS+Planar in planar floor regions, at the same time as being able to handle non-planar floor regions the same way as 2×ICP+OS does. Thus, it can be thought of as a hybrid between 2×ICP+OS and 1×ICP+OS+Planar. Results for this adaptive localization algorithm are presented in Section 3.4.

3.1. Overview of 2×ICP+OS Localization

Given the input laser scans and OS data at t_1 and t_2, the 2×ICP+OS algorithm provides an estimate of the linear 6 DOF transformation from the pose at t_1 to the pose at t_2 by running ICP twice, once on the yaw scanner and once on the pitch scanner. As suggested in [6], the 6 DOF localization problem can be approximately decoupled into a series of 2D scan matching problems. Only the yaw and pitch scanners are needed to recover the translation between successive poses. Specifically, we use Censi's PLICP algorithm for scan matching [5] on the yaw scanner data to obtain t_x, t_y, and the change in yaw, $\Delta\psi$. Since accurate covariance measurements are needed for TORO optimization, we employ Censi's method [4] to estimate the covariances Σ_{t_x}, Σ_{t_y}, and $\Sigma_{\Delta\psi}$. Similarly, by applying scan matching to the pitch scanner data, the estimate of t_z, as well as its covariance measure, Σ_{t_z}, are obtained. The InterSense OS provides absolute orientation estimates, pitch and roll, which allows for construction of the incremental orientation from time t_1 to t_2. The covariance measures for the OS's orientation parameters are taken from the product specifications.

3.2. Overview of 1×ICP+OS+Planar Localization

With a priori knowledge that the backpack is moving through an environment with a planar floor, the 1×ICP+OS+Planar algorithm allows for a much more accurate estimate of the transformation between successive poses. Similar to the 2×ICP+OS algorithm, performing PLICP and Censi's method on the yaw scanner data allows for recovery of t_x, t_y, and $\Delta\psi$ as well as their covariance measures. In addition, the InterSense OS provides pitch and roll. However, as shown in Figure 2, if a line can be fit to the scan samples on the floor, absolute z can be estimated at every time instant. Successive estimates for z allow for the construction of t_z, and the covariance can be estimated using the method described in [6]. This estimate of z has been empirically shown to be more accurate than the one obtained via 2×ICP+OS in regions where the planarity assumption holds. This is because whereas in 2×ICP+OS the change in z, i.e. t_z, is estimated via scan matching from one instant in time to another, in 1×ICP+OS+Planar, absolute z values are estimated from scratch at each time instant, and hence do not get a chance to drift over time.

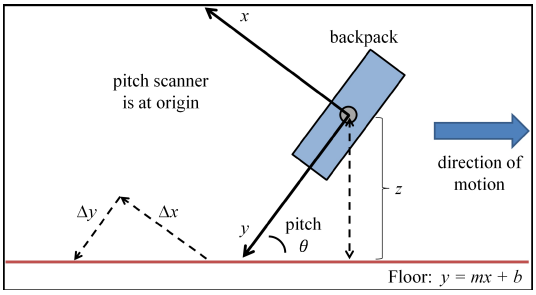


Figure 2. Overview of the floor planarity assumption; the axis are shown in the coordinate system of the pitch laser scanner; it is assumed that the backpack is worn upright by a human operator.

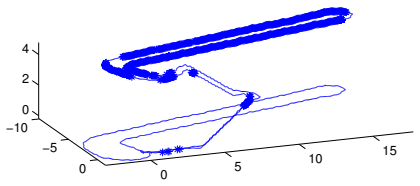
3.3. Adaptive Localization Method

The main drawback of the 1×ICP+OS+Planar algorithm is that a priori knowledge of a planar floor is required for the entire acquisition path. The planarity assumption breaks down in more complex, mixed environments. In order to adaptively switch between 1×ICP+OS+Planar and 2×ICP+OS, the data must be segmented into planar and non-planar regions.

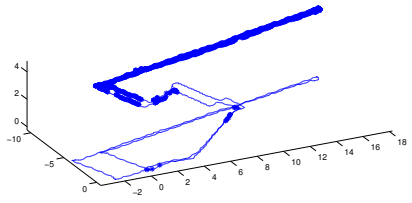
To identify planar portions of the acquisition path, we need to locate range data that coincides with the floor. Assuming that the system is worn approximately upright, laser data points from the pitch scanner that are from the floor generally fall within a specific angular range. Given the entire 240° field of view of the scene, we search for all data points that originate from the 30° slice of the scan corresponding to the downward direction. Next, we fit a line to the resulting data points using least squares regression, and compare the data points to the fitted data to compute the correlation coefficient via:

$$\rho = \frac{\sum_i (x_i - \mu_x)(\hat{x}_i - \mu_{\hat{x}})}{\sqrt{\sum_i (x_i - \mu_x)^2 \sum_i (\hat{x}_i - \mu_{\hat{x}})^2}} \qquad (1)$$

where x and $\hat{x}$ are the original and fitted data points and μ_x and $\mu_{\hat{x}}$ are their respective means. For locations where the floor is planar, ρ should be almost unity. To allow for sensor noise, a lower bound of 0.99 is set for ρ to check for



(a)



(b)

Figure 3. Segmentation results for (a) dataset 1; (b) dataset 2; the boldface markers indicate regions where planar floors are detected.

planarity. In some degenerate cases, such as in buildings with highly reflective floors, few floor data points are collected. In these situations, ρ may not always serve as an accurate means of segmentation; thus we additionally require the number of points being fit to the floor line to be 99% of the number of data points measured over the 30° window, e.g. 85 points in the case of the Hokuyo URG-04LX.

The segmentation algorithm is tested on two separate datasets. Both consist of two long hallways connected by a stairwell. Unlike the upper floor, the lower floor has an extremely reflective floor with no returned laser points. The segmentation results are shown in Figure 3. The boldface regions correspond to planar floor detections. As seen, our proposed segmentation algorithm correctly identifies planar regions on the top floor. Even though the lower floor is planar, it is not detected as planar since it is reflective and has no returns.

The output of the segmentation algorithm is a disjoint, binary partitioning of the transformations between successive poses. Specifically, T_p denotes all transformations that should be computed under the planar assumption and T_{np} denotes those without. The segmentation results are combined with the localization procedures of Sections 3.1 and 3.2 to generate an adaptive localization algorithm. At each time interval, the segmentation algorithm is run and the transition from the pose at t_{i-1} to the pose t_i is classified as planar, T_p, or non-planar, T_{np}. If the transition is a member of T_p, then the transformation is computed according to the 1×ICP+OS+Planar algorithm; otherwise, it is computed via 2×ICP+OS.

3.4. Adaptive Localization Results

The adaptive localization algorithm is characterized on three separate data sets, referred to as datasets 1, 2, and 3. The first two datasets consist of two long hallways connected by a stairwell. Dataset 1 is comprised of two roughly 20-meter hallways connected by a stairwell roughly 4.5-meters in height. Similarly, dataset 2 consists of two 20-meter hallways connected by a 4.5-meter stairwell. Since the complex environment does not allow for the planarity assumption for the entire path, the adaptive localization algorithm is compared to 2×ICP+OS for datasets 1 and 2. On the other hand, for dataset 3, which is entirely planar, we compare the adaptive algorithm to both 2×ICP+OS and 1×ICP+OS+Planar. A manually detected set of LCs are used for all results presented in this section. Comparison is made to the ground truth data collected by the Honeywell HG9900 IMU. Global errors are computed in a world reference frame such that x is east, y is north, and z is upwards.

It is important to clearly distinguish between what we call incremental and global errors. Incremental errors refer to the error in any of 6 DOF parameters from one time instant to the next. Global error is computed by (a) successively applying incremental transformations from our proposed localization algorithms for all times to reconstruct the entire 6 DOF localization path, including position and orientation, and (b) comparing their values with those obtained via the ground truth. Thus, global errors result from accumulated incremental errors. As such, the magnitude of global error for each localization parameter is for the most part decoupled from that of its corresponding incremental error. In particular, various components of incremental localization errors can either cancel each other out to result in lower global errors, or they can interact with each other in such a way so as to magnify global errors.

Figures 4(a) and 4(b) show the global RMS and peak position errors for datasets 1 and 2. As seen, the adaptive algorithm substantially lowers both the peak and RMS localization error for the z-axis. Specifically, for datasets 1 and 2, the reduction in RMS z-error is roughly 20% and 40%, respectively. The translation along other axes as well as global and incremental rotations remain largly unchanged. This is a simple consequence of the fact that both 2×ICP+OS and 1×ICP+OS+Planar estimate all other parameters in the same manner.

Figure 4(c) shows the global RMS and peak errors for dataset 3. Dataset 3 consists of a single T-shaped hallway with a non-glossy floor. The performance of the adaptive algorithm on dataset 3 is almost identical to that of 1×ICP+OS+Planar and roughly 40% better than 2×ICP+OS for z. This is to be expected because the adaptive algorithm classifies almost the entire dataset as planar, 91% in this case. In fact, the average position error for this dataset is 1.2% lower for the adaptive algorithm than for 1×ICP+OS+Planar which assumes planarity throughout.

The average position errors for 2×ICP+OS and the adaptive algorithm for all three datasets are shown in Table 1. As seen, the adaptive algorithm outperforms 2×ICP+OS for all

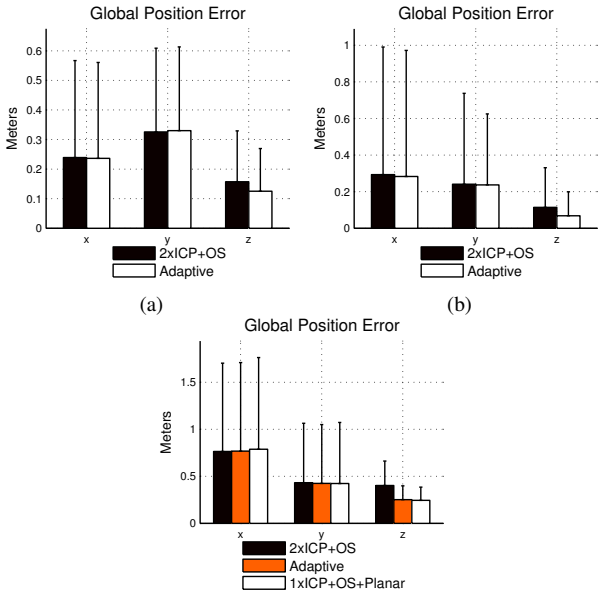


Figure 4. Global RMS position error for 2×ICP+OS and adaptive; markers above each bar denote peak error; (a) dataset 1; (b) dataset 2; (c) dataset 3.

Dataset	Average Position Error		% Change
	2×ICP+OS	Adaptive	
1	0.401 m	0.396 m	-1.2%
2	0.343 m	0.321 m	-6.4%
3	0.903 m	0.847 m	-6.2%

Table 1. Average position error for each dataset for 2×ICP+OS and adaptive localization methods.

three datasets. To conclude, the proposed adaptive algorithm has been shown to be more accurate than 2×ICP+OS in mixed environments with planar and non-planar floors, at the same time as being as accurate as 1×ICP+OS+Planar in planar floor environments.

4. Loop Closures from Multiple Cameras

Rather than using images from only a single camera to look for LCs, in this section we propose to use images from a pair of diametrically opposed side-looking cameras, where one points to the operator's left and the other to the operator's right. In doing so, many more LC points are detected when the operator retraces the path in opposite directions. Specifically, in our 3 datasets this increases the number of LCs by three-fold or better, allowing the localized path to be better constrained in the regions around each LC. In this localization process, a graph is constructed where each node represents a pose at a moment in time, and the edges connecting the nodes are the incremental transformations and the covariance matrices of those transformations.

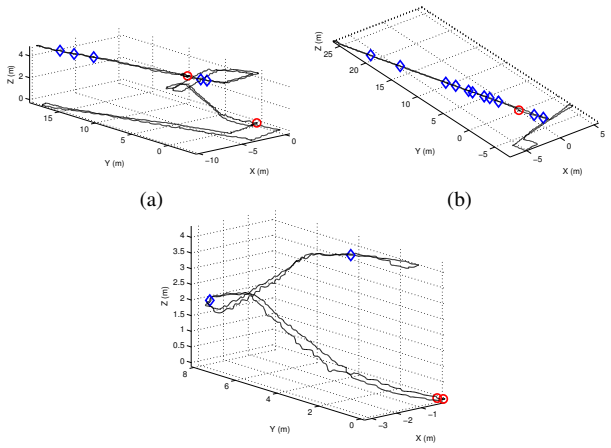


Figure 5. The LCs automatically detected from camera images for (a) dataset 2 with 7 closures, (b) dataset 3 with 12 closures, and (c) dataset 4 with 4 closures; closures detected by a single (double) camera are shown with red circles (blue diamonds).

When an LC is identified, an extra edge is added to connect different parts of the graph together.

Figure 5 shows LCs for 3 datasets detected by a single camera in red circles, and the additional LCs detected by two diametrically opposed side-looking cameras in blue diamonds. As seen in Figure 5(a), not many additional LCs are detected on the lower hallway for dataset 2; this is due to the presence of repeating patterns on the wall from a long set of lockers, making it difficult to distinguish unique images.

Once an LC is detected, the incremental transformation and covariance is computed and inserted as an edge to the pose graph. This is done by scan matching the laser scans associated with the LC in a process similar to the way incremental transformations are estimated for the algorithms described in Section 3. However, as the scan events for LCs are from different points in time and do not necessarily correspond to sufficiently similar poses, the PLICP scan matching may converge to a local minima [5], especially since the initial estimate of the transform is not readily available. In addition, geometrically simple indoor environments such as hallways where scans appear as parallel lines, can lead to degenerate cases with erroneous PLICP estimates of the translation for the yaw scanner. This is particularly problematic as the yaw scanner is used to recover 3 out of 6 localization variables, regardless of which specific localization algorithm we use. Since erroneous transforms lead to large errors during the TORO optimization process, steps must be taken to prune LC candidates with poor transformations.

To detect these, we apply a set of three criteria to all candidate LC transformations; two of these are based on

statistics generated by the yaw scan matching process, as described in Section 4.1, and one is based on the consistency of the transformation with the 3D surroundings, as described in Section 4.2.

4.1. Pruning Based on Scan Matching Statistics

The first two metrics are computed based on the matched, or associated, points between the two laser scans during the scan matching process. Associated points are determined by taking into account the transformation between two successive yaw scans from the horizontal yaw scanner, and computing the distance from each point in the first scan to the nearest neighbor in the second scan. Points for which this distance is below a specified distance, or association gate, are considered to be associated. We then compute the percentage of associated points in each scan as well as the residual distance between those associated points. As no initial conditions to the scan matching process are known, other than which side camera the images came from, the process needs to be repeated across various association gates in order to avoid converging to a local minima. Throughout this paper the association gates are chosen to be 0.05m and 0.45m. Each of these association gates provides a candidate LC transformation.

4.2. Pruning Based on Image Features

We propose an error metric based on the transformation of the 3D location of features between matching LC images. We start with the two images, I_1 and I_2 corresponding to LC events, and based on their timestamps, we collect the corresponding 2D scans from the side-looking vertical geometry scanners which are approximately captured around the same time as the two images. In doing so, we ensure that laser scans and corresponding camera images both come from the same side of the backpack. Specifically, we choose 15 seconds[1] or 150 consecutive scans from the left (right) looking side scanner whereby the timestamp for the 75th scan is closest to that of the image from the left (right) camera. We then apply pre-TORO, unoptimized[2] open loop pose estimates from the algorithms in Section 3 to the scans in order to generate a small point cloud. This point cloud is then projected onto the LC images resulting in 3D depth values for the 2D pixel locations in the images. Figure 6 shows an example of such a depth map.

We also apply a feature extraction algorithm such as SIFT [11] to I_1 and I_2 and match their features to obtain a set of matching features with pixel locations y_1 and y_2. Depth values are assigned to the features according to the following methodology. In each of I_1 and I_2, the features are individually assigned to the 3D position of the nearest projected laser point only if the distance is less than 15 pixels. Furthermore, since pre-TORO, unoptimized poses are used to assemble the point clouds, discontinuities in depth often occur in slightly erroneous locations. To alleviate this, laser points near discontinuities are ignored during this process. Matching features with assigned depth values are collected into the sets $\hat{y_1}$ and $\hat{y_2}$. The candidate LC transformations are then scored according to:

$$e = \frac{1}{N}\sum_{i=1}^{N} \parallel \hat{y_2}(i) - \xi(\hat{y_1}(i), \hat{\boldsymbol{\theta}}, \hat{\mathbf{t}}) \parallel \qquad (2)$$

where $\xi(\hat{y_1}(i), \hat{\boldsymbol{\theta}}, \hat{\mathbf{t}})$ is the 3D location of feature $\hat{y_1}(i)$ roto-translated by the candidate 6 DOF transformation $\xi(\cdot, \hat{\boldsymbol{\theta}}, \hat{\mathbf{t}})$, with $\hat{\boldsymbol{\theta}}$ and $\hat{\mathbf{t}}$ denoting estimated rotations and translations respectively. In essence, this represents the mean distance between the 3D locations of features $\hat{y_1}$ and $\hat{y_2}$ under the candidate transformation. For an ideal depth map, e would be zero when using a perfect candidate transformation. However, since the depth map contains errors, e tends to be nonzero even with a perfect transformation. To account for the quality of the depth map in the scoring process, we use an SVD based approach, such as [1], to estimate a least-squares transformation, $\tilde{\xi}(\cdot, \tilde{\boldsymbol{\theta}}, \tilde{\mathbf{t}})$, from the set of matched features $\hat{y_1}$ and $\hat{y_2}$ directly. We then calculate an image based score, $\tilde{e}$, using Eq. 2 with $\tilde{\xi}(\cdot, \tilde{\boldsymbol{\theta}}, \tilde{\mathbf{t}})$. This $\tilde{e}$ is used to normalize the score for each candidate transformation e. Specifically, we use $e/\tilde{e} = e_{\text{final}}$ as the final metric to prune candidate transformations for each LC event. In essence, when the quality of the depth map as measured by $\tilde{e}$ is high, i.e. the value of $\tilde{e}$ is small, we require the residual error e due to candidate LC transformation to be lower for the transform to be accepted.

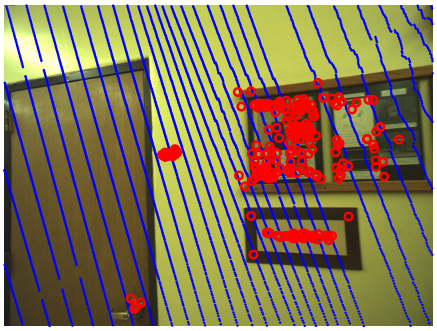


Figure 6. Example depth map using projected laser points to estimate depth. Projected laser points are shown in blue and SIFT features are shown as red circles.

4.3. Applying Loop Closures

Using the metrics described in Sections 4.1 and 4.2, we select from the candidate LC transforms by applying the

[1] We have empirically found 15 seconds of laser data to be sufficient to describe the entire field of view of the cameras.

[2] The use of the unoptimized poses is not problematic in this application because they tend to be locally accurate. Besides, prior to TORO optimization, open loop, dead reckoning poses are the only ones available to use.

Dataset	Detected Loop Closures		
	MSPO	SCA	DCA
2	4	2 (2)	12 (7)
3	1	2 (1)	28 (12)
4	1	2 (2)	6 (4)

Table 2. Number of LCs detected for each dataset and detection mode (MSPO, SCA, DCA), with the number of LCs after pruning in parentheses. MSPO LCs are not pruned.

following logic. For each LC event, all candidate LCs are scored according to the above three criteria. From those, we select the transformation with the smallest association gate that satisfies these three conditions: (a) the proportion of matching points to be significantly high, i.e. 27%, (b) the residual error between matching points to be sufficiently low, i.e. 40 cm, (c) e_{final} to be smaller than 50. If multiple candidate transforms pass all criteria, the transform with the smallest association gate is selected so as to bias the algorithm towards small translations.

4.4. Results on Multiple Cameras LCs

We test the approach described in this section on three datasets 2, 3, and 4, shown in Figures 5(a), 5(b), and 5(c) respectively. Dataset 4 shown in Fig. 5(c), is a single stairwell where the operator starts at the bottom, traverses up, and then returns back to the starting location at the bottom. We compare the localization error for 3 datasets in 3 cases: the first case corresponds to manual detection of LCs with similar position and orientation, namely $\Delta\mathbf{t} \approx 0$ and $\Delta\psi \approx 0$, which we refer to as "Manual Similiar Position and Orientation," or MSPO. The second and third cases correspond to automatic detection of LCs using the approach in [6]. In the second case, which we refer to as "Single Camera Automatic," or SCA, one camera is used and $\Delta\mathbf{t} \approx 0$ and $\Delta\psi \approx 0$. In the third case, which we refer to as "Dual Camera Automatic," or DCA, one or two cameras are used and $\Delta\mathbf{t} \approx 0$ and $\Delta\psi \approx 0$ or $180°$. Table 2 shows the number of loop closures used in each case with and without pruning. As expected, DCA has a larger number of LCs than SCA or MSPO even after pruning.

Figure 7 shows global position error for 3 datasets and the 3 scenarios described above. As seen, global position errors are generally improved with DCA, in particular for dataset 3. For dataset 3, global x (y) errors for DCA are about 65% (25%) lower than MSPO or SCA. In addition, DCA LCs are able to constrain the path well along the hallway. As seen in Figure 8, unlike MSPO and SCA, the path due to DCA overlaps significantly with that of the ground truth. Dataset 4 shows a slight loss of performance when adding DCA LCs, as shown in Figure 7(c); in particular, when the single MSPO LC which represents the start and end of the path is replaced with its equivalent SCA or DCA LC, there is a slight loss in accuracy. This can be attributed

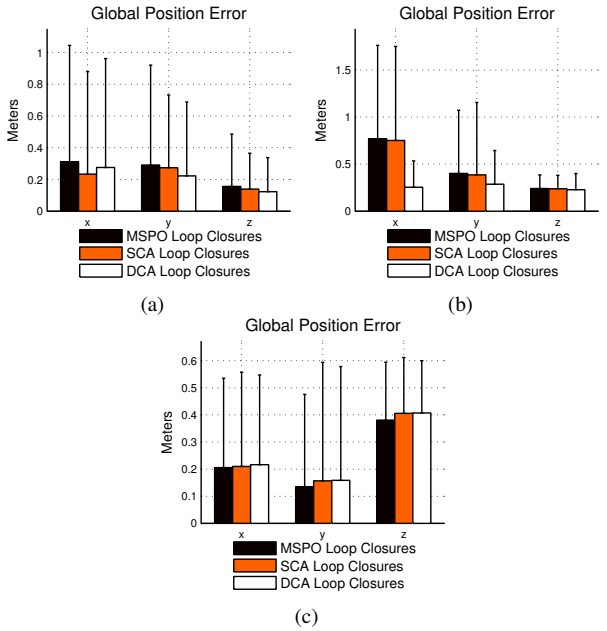


Figure 7. Global RMS Position error characteristics using MSPO, SCA, and DCA LCs for 2×ICP+OS for datasets (a) 2, (b) 3, and (c) 4. Markers above each bar denote peak errors.

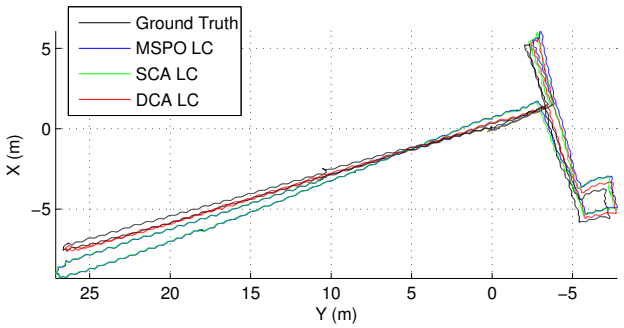


Figure 8. Reconstructed paths for MSPO, SCA, DCA LCs for 2×ICP+OS for dataset 3. SCA and MSPO are overlapping each other at the end of the hallway; same is true for SCA and ground truth.

to the fact that this SCA or DCA LC does not occur at the exact same instant as that of the MSPO LC, and its estimated pose happens to be slightly more erroneous than that of the manually defined one.

Average position error for 3 datasets are composed in Table 3. As expected, DCA outperforms MSPA by 14% and 47% for datasets 2 and 3, respectively, and is outperformed by MSPO by 9% for dataset 4. For dataset 3, there is a 47% improvement in average position error from SCA to DCA indicating the effectiveness of using 2 camera LC over 1 camera LC.

The global rotation errors for the three approaches are nearly identical. As the LCs only represent a few transformations compared to the large number of total transforma-

Dataset	Average Position Error		
	MSPO	SCA	DCA
2	0.389 m	0.349 m (-10.3%)	0.333 (-14.4%)
3	0.819 m	0.789 m (-3.7%)	0.422 (-47.4%)
4	0.426 m	0.458 m (+7.5%)	0.463 (+8.7%)

Table 3. Average position error for each dataset using MSPO, SCA, and DCA loop closure sets. The % change between SCA/DCA and MSPO is shown in parentheses.

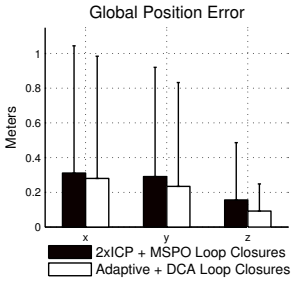


Figure 9. Global RMS position error characteristics using adaptive localization and DCA LCs on dataset 2. Markers above each bar denote peak errors.

	Avg. Pos. Error	% Change
2×ICP + MSPO LCs	0.389 m	—
Adaptive + DCA LCs	0.334 m	-14.1%

Table 4. Average position errors for dataset 2 using adaptive localization and DCA LCs.

tions, the incremental position and rotation error statistics also remain largely unchanged.

5. Combining Adaptive Localization and Multiple Camera LC

In this section, we present results combining the two techniques discussed in Sections 3 and 4, namely adaptive localization and DCA. The combination algorithm is applied only to dataset 2, because the adaptive algorithm closely mimics 1×ICP+OS+Planar on dataset 3 and 2×ICP+OS on dataset 4, and as such the results are not different from those in previous sections. The combined method is compared to the baseline 2×ICP+OS method with MSPO LCs: Figure 9 shows the global RMS errors for the translation parameters, and Table 4 shows the average position errors. As seen, the combined method reduces error for x, y, and z as well as average position error as compared to the baseline.

6. Conclusions

In this paper we have presented an adaptive localization algorithm which detects and exploits floor planarity in order to increase accuracy in mixed environments with planar and non-planar floors. We have presented a method for the integration of automatic image based LCs from diametrically opposing cameras for situations where data acquisition paths are re-traversed. We have shown that in general, after pruning the resultant erroneous LC transformations, the addition of the extra LCs resulting from using left and right cameras decreases localization error.

References

[1] K. S. Arun, T. S. Huang, and S. D. Blostein. Least-squares fitting of two 3-D point sets. *IEEE Trans. Pattern Anal. Mach. Intell.*, 9:698–700, 1987. 6

[2] D. Borrmann, J. Elseberg, K. Lingemann, A. Nuchter, and J. Hertzberg. Globally consistent 3D mapping with scan matching. *Robotics and Autonomous Systems*, 56:130–142, 2008. 1

[3] M. Bosse and R. Zlot. Continuous 3D scan-matching with a spinning 2D laser. In *Proc. of the IEEE Int. Conference on Robotics and Automation*, pages 4244–4251, 2009. 1

[4] A. Censi. An accurate closed-form estimate of ICP's covariance. In *Proc. of the IEEE Int. Conference on Robotics and Automation*, 2007. 3

[5] A. Censi. An ICP variant using a point-to-line metric. In *Proc. of the IEEE Int. Conference on Robotics and Automation*, 2008. 3, 5

[6] G. Chen, J. Kua, S. Shum, N. Naikal, M. Carlberg, and A. Zakhor. Indoor localization algorithms for a human-operated backpack. In *3D Data Processing, Visualization, and Transmission*, 2010. 1, 2, 3, 7

[7] A. J. Davison, I. D. Reid, N. D. Molton, and O. Stasse. MonoSLAM: Real-time single camera SLAM. *IEEE Trans. Pattern Anal. Mach. Intell.*, 29(6):1052–1067, 2007. 1

[8] K. Granstrom, J. Callmer, F. Ramos, and J. Nieto. Learning to detect loop closure from range data. In *Proc. of the IEEE Int. Conference on Robotics and Automation*, 2009. 2

[9] G. Grisetti, S. Grzonka, C. Stachniss, P. Pfaff, and W. Burgard. Efficient estimation of accurate maximum likelihood maps in 3D. In *Proc. of the IEEE/RSJ Int. Conference on Intelligent Robots and Systems*, 2007. 2, 3

[10] T. Liu, M. Carlberg, G. Chen, J. Chen, J. Kua, and A. Zakhor. Indoor localization and visualization using a human-operated backpack system. In *Int. Conference on Indoor Positioning and Indoor Navigation*, 2010. 1, 2

[11] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *Int. J. Comput. Vision*, 60(2):91–110, 2004. 2, 6

[12] F. Lu and E. Milios. Robot pose estimation in unknown environments by matching 2D range scans. *Journal of Intelligent and Robotic Systems*, 18:249–275, 1994. 1

[13] P. Newman and M. Cummins. FAB-MAP: Probabilistic localization and mapping in the space of appearance. *The Int. J. of Robotics Research*, 27(6):647–665, 2008. 2

[14] V. Pradeep, G. Medioni, and J. Weiland. Visual loop closing using multi-resolution SIFT grids in metric-topological SLAM. In *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition*, 2009. 1, 2