

Adaptive RGB-D Localization

Michael Paton
Computer Science
George Mason University
Fairfax, Virginia
mpaton@gmu.edu

Jana Košečka
Computer Science
George Mason University
Fairfax, Virginia
kosecka@cs.gmu.edu

Abstract—The advent of RGB-D cameras which provide synchronized range and video data creates new opportunities for exploiting both sensing modalities for various robotic applications. This paper exploits the strengths of vision and range measurements and develops a novel robust algorithm for localization using RGB-D cameras. We show how correspondences established by matching visual SIFT features can effectively initialize the generalized ICP algorithm as well as demonstrate situations where such initialization is not viable. We propose an adaptive architecture which computes the pose estimate from the most reliable measurements in a given environment and present thorough evaluation of the resulting algorithm against a dataset of RGB-D benchmarks, demonstrating superior or comparable performance in the absence of the global optimization stage. Lastly we demonstrate the proposed algorithm on a challenging indoor dataset and demonstrate improvements where pose estimation from either pure range sensing or vision techniques perform poorly.

Keywords-RGB-D; SLAM; Mapping; Visual Odometry;

I. INTRODUCTION

The ability for autonomous robots to localize in novel environments and generate accurate 3D maps has been of great interest, and heavily researched by the robotics community. Traditional localization and mapping techniques have mostly relied on range sensors for localization and mapping tasks. While successful solutions have been attained [1], they either required expensive laser range finders or they were limited to simplistic environments [2]. More recent efforts have seen use of vision based sensors, which are less expensive and take advantage of rich visual information [3]. There are many environments which contain little or no visual features, limiting the robustness of all visual based mappers.

Several recent efforts have been made to combine range finding sensors with vision sensors to overcome the weaknesses of both. Research in this area became more active with the availability of RGB-D cameras, which provide video (RGB) data, along with per-pixel depth information. Preliminary mapping efforts have been reported in [4], [5]. The previous techniques combine vision and range sensing for mapping algorithms into one coherent solution, typically using visual sensing to initialize the ICP style algorithms for matching range data. While these algorithms allow for a certain flexibility they heavily rely on either range matching

or vision based motion estimation techniques, the choice of the interaction between two techniques is typically made off-line, making it difficult to adapt to difficult dynamic environments and agile motions.

As our main contribution in this paper we show how to use visual correspondences to initialize the Generalized ICP algorithm and how to adaptively switch between different models of initialization depending on the quality of the correspondences and initial pose estimates. This results in a robust and accurate system, capable of coping with a variety of environments where either pure visual or range based matching fail.

Overview: Related work is detailed in Section II. Section III introduces the proposed RGB-D mapping algorithm. Testing and experimental evaluation with the RGB-D benchmark data set as well as results of the experiments are reviewed in Section IV. Section V provides closure on the results as well as future work.

II. RELATED WORK

In this work we focus on metric SLAM using RGB-D cameras. This problem can be broken down into following sub-problems: data association and motion estimation from two views, loop closure detection, and globally consistent motion estimation. While the globally consistent motion estimation part is often similar regardless of sensing modality, data association and motion estimation techniques differ.

Laser Mapping: There has been a large number of works published on problems of data association in the context of range sensing; ranging from matching raw scans to various feature based methods. Some discussion of the topic can be found in [6]. The methods which use raw scans achieve correspondences by means of scan matching, which is typically accomplished by the Iterative Closest Point (ICP) algorithm originally proposed by [7]. In ICP-style algorithms the correspondences are typically initialized by finding nearest neighbor point in 3D scans. The advantages of any ICP variant is the simplicity of the algorithm, and quick performance when the nearest neighbor calculations are optimized. This is typically accomplished with k-d trees. A drawback of ICP-style algorithms include the assumption that there is a full overlap between the two point clouds and

difficulty of finding reliable correspondences between two scans.

Vision Mapping: The need to robustly differentiate between scenes with similar geometry and obtain richer models has spurred research in vision based mapping. With the advances in scale invariant image feature matching such as [8] and [9], research in wide baseline matching, loop closure detection, and 3D mapping led to several successful solutions to the vision based localization problem. In work of [10] the authors proposed novel parameterization of the problem and used Extended Kalman Filter focusing on real-time issues. Works of [11], [12] both pursued a Fast-SLAM particle filtering approach using stereo and monocular models respectively. Authors in [13], [14] used full MAP estimation approach, where the estimates were initialized from two frames using linear methods, followed by a global bundle adjustment. In addition to visual odometry and sparse feature models, works of [15], [16] focus on dense 3D reconstruction.

Along with the vision sensing modality comes a list of unique drawbacks. Maps generated by matching visual features are sparser than those generated by range finders. An exception to these are full 3d reconstruction systems which are usually computationally more expensive [15]. Many texture-less environments where the quantity and reliability of scale-invariant keypoints is low are often problematic for visual odometry systems. An example of this is the common office hallway, white walls and often featureless carpets. In this situation it is very difficult for a vision system to match features between frames, especially in the presence of large motions.

Multi Sensor Mapping: The common drawbacks of both range sensing and vision based modalities, spurred research into techniques that utilize both laser and vision sensors. A recent example of this is [17], who uses both a laser scanner and a camera installed on a mobile robot to incrementally build a 3D metric map of the environment.

The combination of both sensors allows for highly accurate odometry estimation from the laser data coupled with robust and quick loop closure detection from the camera, effectively making the fusion approach more effective than either sensor alone. These techniques may overcome some of the obstacles of range finding and vision sensors, but they still rely on expensive laser sensors, additionally it can be difficult to associate 3D range data with the collected vision data.

RGB-D sensors solve the problem of data correspondence found in multi sensor mapping by providing RGB images synchronized with per-pixel depth information. Current examples of RGB-D mapping techniques are [5] and [4]. In the approach of [5] SURF feature matching is used to robustly estimate initial motion using RANSAC. The motion is then refined by Generalized ICP, followed by global optimization performed by the HOGMAN method. The



Figure 1. Limited matching features.

approach is evaluated on the TUM benchmark dataset [18]. In [4], SIFT features are used instead and motion is refined using Generalized ICP instead of point-to-plane ICP. Both methods use visual matches, RANSAC and motion estimates to initialize the ICP algorithms.

III. ADAPTIVE RGB-D LOCALIZATION

Our method combines matching and motion estimation strategies found in both visual and range sensing modalities. The general outline of a common RGB-D mapping consists of a) feature detection and matching b) initial rigid body motion estimation with RANSAC and c) refinement of motion estimates with an ICP-style algorithm. There are two cases where step b) can fail: a lack of visual features in the environment, or a lack of common features between two frames. The former case can be seen in common indoor environments such as office hallways. The latter can be attributed to rapid movement, small overlap between consecutive views or possible data loss due to communication and threading issues. An example of this can be seen in Figure 1, the only overlap between the two frames is the corner wall which is essentially featureless. In all of these scenarios the motion estimate derived from RANSAC may not be representative of the true motion and initializing ICP with this motion can cause ICP not to converge. We propose a method to overcome these weaknesses by using Generalized ICP algorithm [19] not only as a motion refinement tool but also as a fallback method if RANSAC provides an unsatisfactory result. A general flow of our algorithm is detailed in Figure III.s

There are three possible paths in our algorithm. The first two start with feature extraction and matching with SIFT features, motion estimation and outlier rejection using visual correspondences and RANSAC algorithm. After this step the performance of RANSAC is analyzed, in terms of the error returned from the estimation, as well as the amount of visual correspondences that were found to be inliers. If the result is satisfactory then the motion estimate is accepted. In the case of the estimate's error being high generalized ICP is used to refine the estimation. In the third case the motion is estimated solely using Generalized ICP on a random subset of the RGB-D depth data. This is triggered when any of the following happens:

- 1) There are not enough visual correspondences to con-

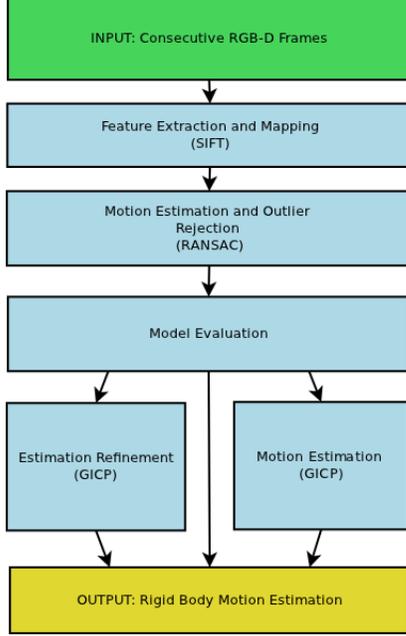


Figure 2. plane-to-plane algorithm flow

fidently compute an estimate with RANSAC.

- 2) The motion estimate generated by RANSAC has high residual error not suitable for ICP initialization.
- 3) RANSAC exceeded maximum number of iterations indicating high outlier ratio.

In any of these cases the motion estimate generated by RANSAC would hinder Generalized ICP from converging to the correct solution. In this case Generalized ICP estimates the motion without the initialization from RANSAC. This path also disregards the visual correspondences from the feature detection and matching stage in place of a random subset of points in the image and their respective 3D coordinates in consecutive scans. Generalized ICP does not require exact correspondences to converge to a correct motion estimate. This initialization enables correct registration, even when there is a lack of full overlap between point clouds. More details about Generalized ICP algorithm can be found in [19]. The pseudocode of the proposed adaptive algorithm is in Figure III.

The thresholds τ_1 , τ_2 , and μ are the parameters of the algorithm. If the average residual error of the estimated motion generated from the RANSAC algorithm is less than or equal to τ_1 and the number of inliers is greater than μ , then the RANSAC estimate is deemed successful and is used as the final answer, otherwise Generalized ICP is used. If the RANSAC error is greater than the user provided threshold τ_2 , then the motion estimate provided by RANSAC is deemed too unstable and generalized ICP starts from scratch using a randomized subset of the full 3D range data.

```

1:  $[\{F_A\}, \{F_B\}] \leftarrow ExtractSIFT(RGBD_A, RGBD_B)$ 
2:  $[2DP_A, 2DP_B] \leftarrow MatchFeatures(F_A, F_B)$ 
3:  $[3DP_A, 3DP_B] \leftarrow ConvertTo3DWorld(2DP_A, 2DP_B)$ 
4:  $T_0 \leftarrow I$ 
5:  $[T, inliers, err] \leftarrow RANSAC(3DP_A, 3DP_B, T_0)$ 
6: if  $(sizeof(inliers) \leq \mu) \vee (err \geq \tau_1)$  then
7:   if  $err \geq \tau_2$  then
8:      $[3DP_A, 3DP_B] \leftarrow GetSubset(3DP_A, 3DP_B)$ 
9:      $T \leftarrow GICP(3DP_A, 3DP_B, I)$ 
10:  else
11:     $T \leftarrow GICP(3DP_A, 3DP_B, T)$ 
12:  end if
13: end if

```

Figure 3. Algorithm: Adaptive RGB-D localization

In this case generalized ICP registers the new point clouds using only the identity matrix as an initialization. If the error is less than τ_2 but greater than τ_1 , generalized ICP uses the motion estimate and inliers obtained by RANSAC to enhance the transformation \mathbf{T} . We have used values of $\tau_1 = 5$, $\tau_2 = 20$ and $\mu = 20$ across all our experiments.

IV. EXPERIMENTS

In order to test the validity of our algorithm, we ran two separate experiments. The goal of the first was to compare our results to the current state of the art RGB-D mapping algorithms [4] [5]. This was accomplished by utilizing the RGB-D mapping benchmarks provided by Technische Universität München [18]. The advantages to using these benchmarks are twofold: each trajectory provides accurate ground truth data measured by a third party sensor, and results on each respective algorithm have been collected, allowing for a way to directly compare algorithms. Benchmark data was taken with a Microsoft Kinect sensor, providing 640 x 480 RGB and depth frames at a 30Hz rate, the ground truth data was taken with a highly accurate motion capture system, composed of eight 100Hz cameras. The second experiment was set up with the purpose of testing our algorithm in a setting that is sparse in visual features, for the RGB-D benchmarks lacked such environments. This was accomplished by collecting RGB-D data in the barren office hallway of our university's engineering building.

A. RGB-D Benchmarks

In the benchmark experiments, we focused on five of the provided trajectories. The chosen trajectories range in length, presence of loops, purpose, and difficulty. An example is the XYZ benchmark shown in Figure 4. This benchmark is rich with visual features, is of short duration and contains an extraordinarily large amount of loops which can be exploited in global optimization. This benchmark is almost trivial to solve and serves mainly as a tool to debug camera calibration

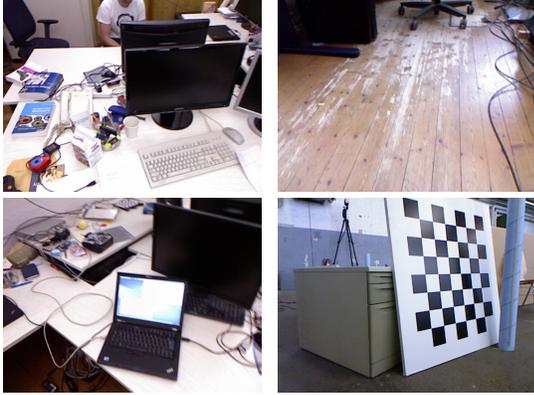


Figure 4. **Benchmark Examples:** Example frames from varying trajectories. Clockwise starting from top right (xyz,floor,room,pioneer).

errors. In contrast the SLAM3 benchmark is nearly three times the length, and contains no loops.

The main goal of the experiments is to provide an empirical comparison of our algorithm to a current state of the art RGB-D mapping method, RGB-D SLAM [5] as well as the component algorithms of Adaptive RGB-D localization (RANSAC, RANSAC refined w/ GICP).

B. Pioneer Hallway Data

In order to test RGB-D mapping in feature limited environments, we took data of the hallway of our University’s engineering building. Data that consists mainly of white walls and grey carpets, with only doorways and light fixtures providing reliable visual features. There are sections of the trajectory that involve tight corners with large rotations between frames. This environment is challenging for systems using vision or range sensing equipment. The data was taken with a Microsoft Kinect attached to a Pioneer robot. Example frames from the dataset can be seen in Figure 1.

V. RESULTS

Presented in this section are results from the RGB-D benchmark and GMU Hallway experiments. Due to the nature of each data set, the results from both are presented in different formats. The TUM benchmarks provide highly accurate ground truth data collected from a third party sensor thus we are able to display the error between measured and truth data. Due to time constraints we were unable to provide ground truth data for our hallway collection. As a result benchmark data is presented with respect to truth data, while the hallway data is merely a comparison of trajectory plots between algorithms.

A. Benchmark Results

Results of the benchmark experiments were calculated using the evaluation tools provided with the benchmark. There are two means of evaluation: absolute trajectory error and relative pose error. The absolute trajectory error (ATE)

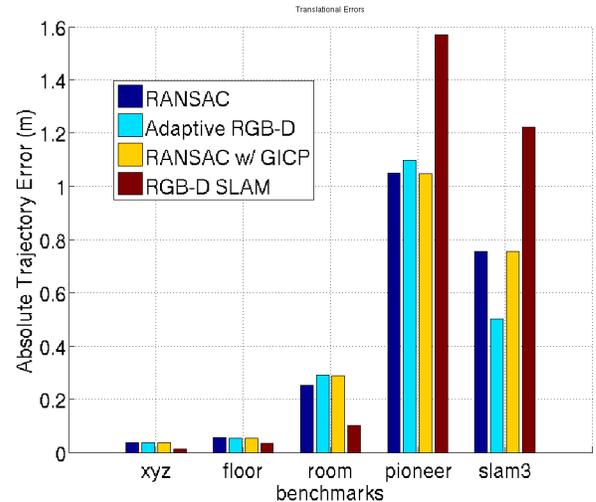


Figure 5. Benchmark evaluation Absolute Trajectory Error

evaluation method directly compares the difference between poses in the ground truth and measured trajectory. Measuring the absolute position between poses, this error evaluation is especially useful for evaluating the performance of visual SLAM systems. The end result of the ATE evaluation method is the root mean squared error (RMSE) of the per pose errors summed over the entire trajectory, it is calculated in the following way:

$$ATE_{rmse} = \sqrt{\frac{1}{\eta} \left(\sum_{i=0}^{\eta} \|\hat{t}_i - t_i\|^2 \right)}$$

where $\hat{t}_i = [\hat{x}, \hat{y}, \hat{z}]^T$ and $t_i = [x, y, z]^T$ and \hat{t}_i is the measured pose at position i , and t_i is the truth pose at position i . Results of the ATE evaluation method on selected benchmarks can be seen in Figure 5.

The relative pose error (RPE) evaluation method directly compares the error in relative motion between all timestamps in the trajectory. This method is very useful for evaluating the effect of accumulated drift. Results from this method are separated into translational and rotational errors. The RPE results of a 1.0 second drift on all benchmarks for Adaptive RGB-D Localization and RGB-D Slam can be seen in Figure 6.

As with the ATE method, results from the RPE method will be presented in terms of the root mean squared error over the entire trajectory. The RPE method calculates both translational and rotational errors in the following way:

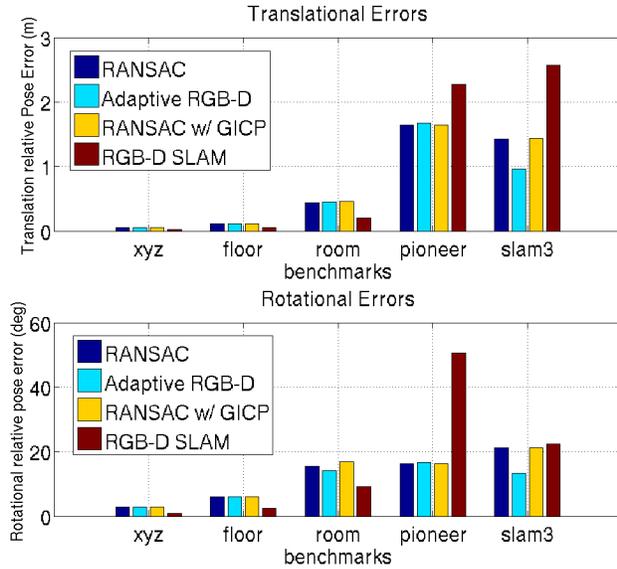


Figure 6. Benchmark Evaluation relative pose error

$$Trans_{RPE} = \sqrt{\frac{1}{\eta} \left(\sum_{i=0}^{\eta} \|\hat{t}_i - t_i\|^2 \right)} \quad (1)$$

$$Rot_{RPE} = \sqrt{\frac{1}{\eta} \left(\sum_{i=0}^{\eta} \|R_i^T \hat{R}_i\|^2 \right)} \quad (2)$$

Where (\hat{t}_i, \hat{R}_i) and (t_i, R_i) are the relative translations and rotations between frames of a specified time apart, for measured and truth data respectively.

1) **Overall Results:** The first set of benchmarks to be tested were the so called short duration benchmarks: xyz, floor, and room. These are trajectories under fifty seconds long. Results between all three trajectories were similar in nature: overall generally low errors below .5 meters for all algorithms. RGBD-SLAM slightly outperforms Adaptive RGB-D in these benchmarks. An explanation of this might be the lack of global optimization in our algorithm, considering all of the short duration benchmarks contain loops. In addition to the short duration benchmarks we also tested with two longer trajectories: pioneer and SLAM3. These trajectories ranged between 100 and 160 seconds in length, with a more open environment than previous trajectories. The purpose of testing against these benchmarks was to get an idea of how our algorithm is performing with respect to error due to accumulated drift with the absence of a global optimization solution to correct it. Overall results of all benchmarks are detailed in Figures 5 and 6, for absolute trajectory errors and relative pose errors respectively. The

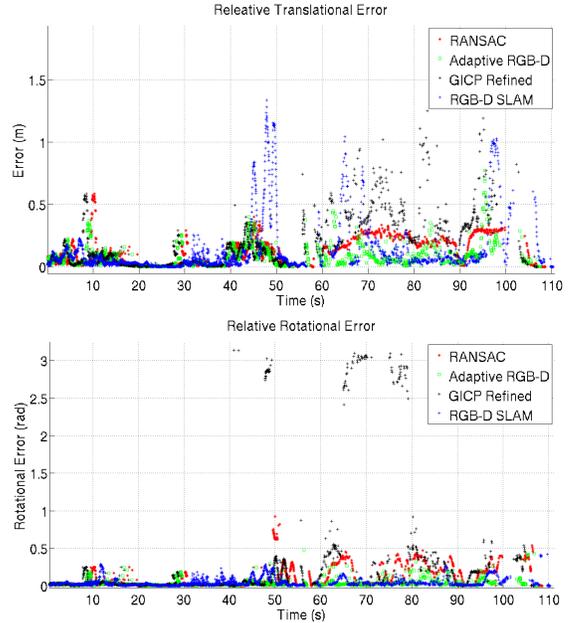


Figure 7. Pioneer RPE 1.0 second drift

following subsections explore the detailed results of both medium distance trajectories: pioneer and SLAM3.

2) **Detail Results: Pioneer:** At 155.72 seconds in length, the pioneer trajectory is the longest tested in these experiments. There are many loops and opportunities for loop closure in this trajectory. Relative pose errors with a 1.0 second drift are detailed in Figure 7. The adaptive RGB-D algorithm show consistently lower errors throughout the trajectory. Results also show that Adaptive RGB-D Mapping generally outperforms both standalone RANSAC as well as RANSAC refined with GICP.

The Pioneer SLAM benchmark contains many loops, making an ideal dataset for global optimization. Despite this our algorithm outperforms RGB-D SLAM. A possible explanation might be the introduction of false loops into their loop closure algorithms. The absolute trajectory plots (Figure 8) affirm that the Adaptive RGB-D method’s trajectory is closer to the ground truth data.

3) **Detail Results: SLAM3:** The SLAM3 trajectory is 111.9 seconds in length, and contains no loops, making it markedly different from the Pioneer trajectory, yet the results are similar. Details of the relative pose errors are in Figure 9. It shows that there are times when the RGB-D SLAM’s translational error is larger by a full meter. This can be seen near frame 50, frame 65, and frame 98. Visualization of the map generated by Adaptive RGBD and RGBD-SLAM can be seen in figure 10.

In both medium distance benchmarks our RGB-D localization algorithm produced errors significantly smaller than RGB-D SLAM. Our algorithm not only outperforms a mapping algorithm with global optimization in a trajectory

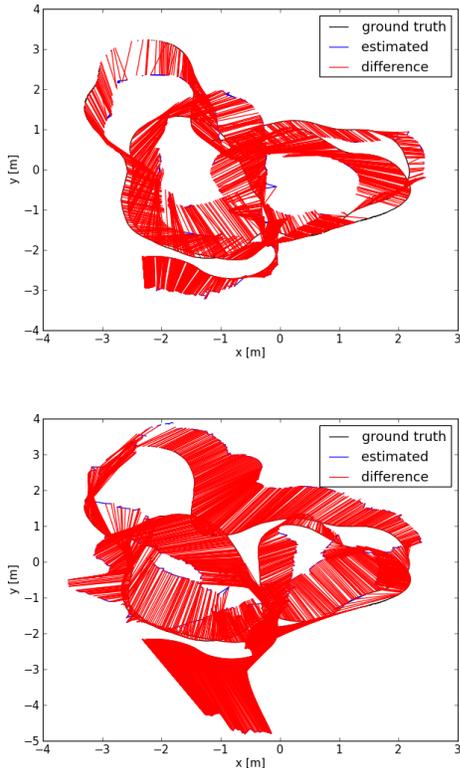


Figure 8. Pioneer Absolute Trajectory Plots (Adaptive RGB-D, RGB-D SLAM)

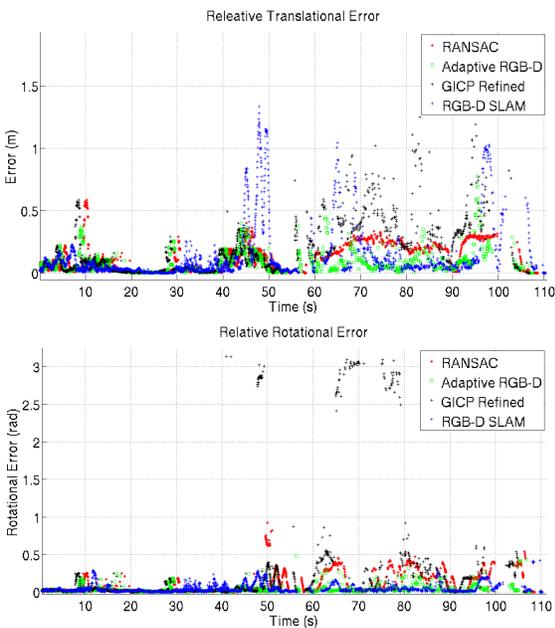


Figure 9. SLAM3 RPE 1.0 second drift

with many loops, but it also does so in a no loop situation as well. There are two main differences between our method and RGB-D SLAM; Global optimization and our algorithms ability to chose which algorithm to use depending on the circumstances. The strength of this can be seen in the SLAM3 results, where Adaptive RGB-D localization outperforms RGB-D SLAM, stand-alone RANSAC, and RANSAC refined with GICP. Considering this a few possible explanations might be made on RGB-D SLAM’s behalf. It is possible the use of global optimization is hindering their result by introducing false loops. It is also a possibility that generalized ICP is providing incorrect results when initialized with a poor motion estimate from RANSAC based off of visual features.

B. Hallway Results

The intent of the hallway experiments was to see how the Adaptive RGB-D method performed in a feature limited environment. This was accomplished by collecting data of some simple trajectories in the hallways of our University’s engineering building. An example of one of the data sets can be seen in Figure 1. The sequence is approximately 600 frames and features blank grey walls and solid carpets. it features a straight trajectory down a hallway, one 90° turn to the right and a small distance down the next hallway. The format of our results for this experiment are quite different from the benchmark data. We lacked the resources and time to collect detailed truth data as was provided with the RGB-D benchmarks, so we compare the resulting trajectories and associated 3D points visually comparing trajectories between Adaptive RGB-D and the pure RANSAC/ICP method. Results can be seen in Figure 11. It shows the pure RANSAC/ICP algorithm on the left in red compared with the Adaptive RGB-D method on the right in green. Comparing these trajectories it is evident that the pure RANSAC algorithm does not perform very well. The first straight away starting from point (0,0) it veers to the right by at least five meters, while the 90° turn is highly inaccurate. The Adaptive RGB-D trajectory shows an accurate approximation of the trajectory. While more work needs to be done to quantify the errors between these algorithms, it is evident that the Adaptive RGB-D algorithm can cope with these environments where one who relied too heavily on visual features and RANSAC would have difficulties.

VI. CONCLUSIONS

While we only tested against a limited number of benchmarks we made several observations. The first is that in the face of many loops, global optimization can improve local motion estimates. The observations made in the short duration benchmark results (Section V-A) showed that the RGB-D SLAM algorithm consistently outperformed the Adaptive RGB-D mapping in simple trajectories that are

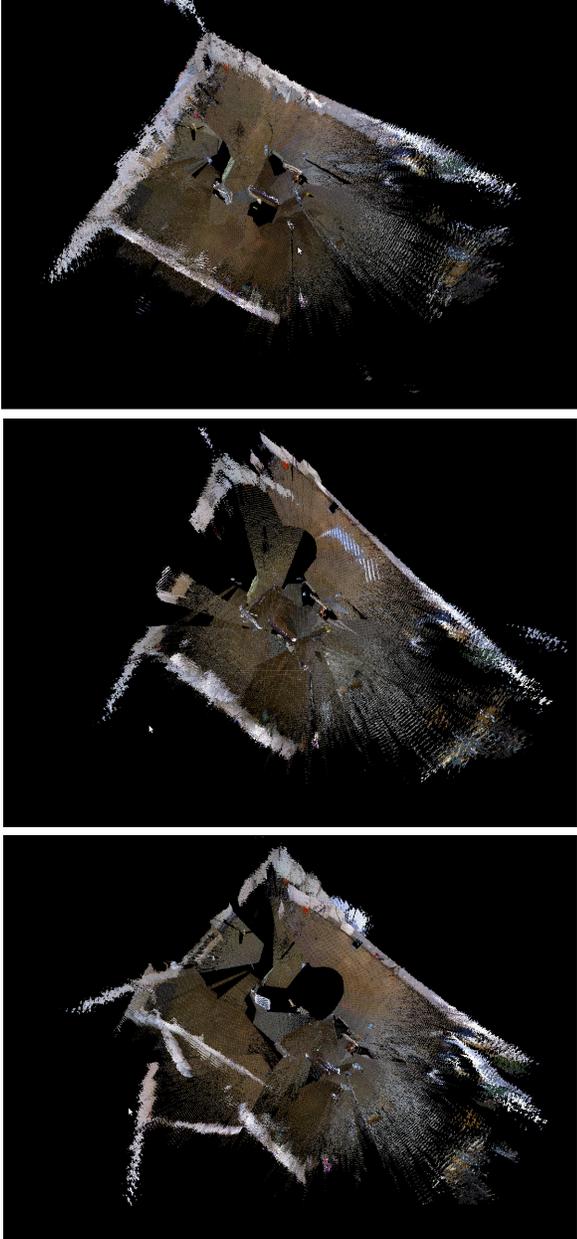


Figure 10. SLam3 Benchmark Visualization (top: truth middle: Adaptive R-GBD; bottom: RGB-D SLAM)

rich in visual features, contain many loops and produce very small RANSAC errors. This evidence leads us to believe that generalized ICP would do little to refine these trajectories, and an adaptive solution is not necessary. This leaves only one real difference between RGB-D SLAM and adaptive RGB-D, global optimization.

The second hypothesis is that as trajectories grow longer in length and become increasingly difficult the need for a method of dynamically alternating between both vision and depth based algorithms increases, as the need for global optimization decreases. This was observed when comparing our

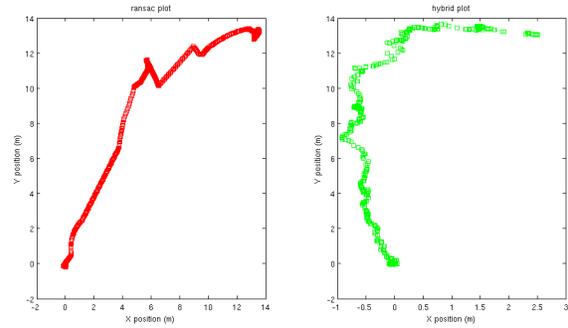


Figure 11. Hallway Comparison

Adaptive method with RGB-D SLAM in the two medium distance benchmark experiments. Despite our algorithm lacking global optimization we were able to outperform RGB-D SLAM in a trajectory that contained many loops and one that contained none. Since the results of both trajectories are similar, it would appear that global optimization is doing little to assist RGB-D SLAM. In the RGB-D SLAM method, ICP acted as a mandatory refinement step picking up where the RANSAC algorithm leaves off. Our method is able to adapt to dynamic environments, utilizing generalized ICP according to observed errors from the initial RANSAC motion estimate. Based on the results of the medium trajectory experiments, this allows us to hypothesize that there are certain conditions when the RANSAC motion estimate is off to a degree where it will hinder generalized ICP from converging and may even cause further deviation from the correct transformation. This makes it necessary at times to abandon the RANSAC estimate and have generalized ICP start from scratch. Our Adaptive algorithm does this and has proven to outperform a state of the art RGB-D mapping algorithm, even in absence of global optimization.

The benchmark results proved that our Adaptive RGB-D method was able to successfully approximate our hallway trajectory dataset, compared to the standalone RANSAC algorithm which did not. The initial results show that the Adaptive RGB-D method has potential to solve the current problems of mapping with RGB-D sensors.

Future work: The results of the short and medium distance trajectory experiments seem to contradict one another, both reaching separate conclusions on the usefulness of global optimization. The hallway results are a first step in determining if our adaptive method can accurately map difficult trajectories, but it currently lacks the ground truth reference, and global optimization stage. This will allow us to better compare our algorithm with current state of the art methods, as well as gauge the importance of global optimization and the need for an adaptive method.

The hallway experiments and results proved that using generalized ICP along with RANSAC is a viable method

for coping with feature limited environments. More experiments with a better means of evaluating results needs to be performed. We propose collecting data and performing experiments much in the same vein as the TUM RGB-D dataset, with a focus on feature limited environments and difficult trajectories where current RGB-D methods will fail. This will serve as a way to test the robustness of RGB-D mapping algorithms and perhaps assist in the development of new algorithms to overcome these difficult situations.

ACKNOWLEDGMENT

We would like to thank Peter Henry, for his wonderful advice and contribution to RGB-D mapping and Jürgen Sturm and his group at TUM for providing the ever helpful RGB-D benchmarks.

REFERENCES

- [1] G. Grisetti, R. Kuemmerle, C. Stachniss, and W. Burgard, "A tutorial on graph-based slam." in *IEEE Transactions on Intelligent Transportation Systems Magazine.*, vol. 2(4), 2010, pp. 31–43.
- [2] K. O. Arras, J. A. Castellanos, and R. Siegwart, "Feature-based multi-hypothesis localization and tracking for mobile robots using geometric constraints," in *Proc. IEEE International Conference on Robotics and Automation (ICRA'02)*, Washington DC, USA, 2002.
- [3] K. Konolige and M. Agrawal, "Frame-frame matching for realtime consistent visual mapping," in *Proc. International Conference on Robotics and Automation (ICRA)*, 2007.
- [4] P. Henry, M. Krainin, E. Herbst, X. Ren, and D. Fox, "Rgb-d mapping: Using depth cameras for dense 3d modeling of indoor environments." Intel Labs Seattle, Seattle, WA: International Symposium on Experimental Robotics, 2010.
- [5] N. Engelhard, F. Endres, J. Hess, J. Sturm, and W. Burgard, "Real-time 3d visual slam with a hand-held rgb-d camera," in *Proc. of the RGB-D Workshop on 3D Perception in Robotics at the European Robotics Forum*, Vasteras, Sweden, April 2011.
- [6] R. Siegwart, I. Nourbakhsh, and D. Scaramuzza, *Autonomous Mobile Robotics*, 2nd ed. MIT Press, 2011.
- [7] Z. Zhang, "Iterative point matching for registration of free-form curves," 1992.
- [8] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. of Computer Vision*, vol. 60, no. 2, pp. 91–110, 2004, <http://www.cs.ubc.ca/~lowe/keypoints/>.
- [9] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool, "Speeded-up robust features (surf)," *Comput. Vis. Image Underst.*, vol. 110, pp. 346–359, June 2008. [Online]. Available: <http://dl.acm.org/citation.cfm?id=1370312.1370556>
- [10] A. Davison, "Real-time simultaneous localisation and mapping with a single camera," in *Computer Vision, 2003. Proceedings. Ninth IEEE International Conference on*, oct. 2003, pp. 1403–1410 vol.2.
- [11] T. Barfoot, "Online visual motion estimation using fastslam with sift features," in *Intelligent Robots and Systems, 2005. (IROS 2005). 2005 IEEE/RSJ International Conference on*, aug. 2005, pp. 579–585.
- [12] E. Eade and T. Drummond, "Scalable monocular slam," in *Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on*, vol. 1, june 2006, pp. 469–476.
- [13] F. Fraundorfer, C. Engels, and D. Nistér, "Topological mapping, localization and navigation using image collections," in *Proc. of IEEE/RSJ IROS*, 2007, pp. 3872–3877.
- [14] K. Konolige and M. Agrawal, "Frameslam: From bundle adjustment to real-time visual mapping," *Robotics, IEEE Transactions on*, vol. 24, no. 5, pp. 1066–1077, oct. 2008.
- [15] R. Newcombe, S. Lovegrove, and A. Davison, "Dtam: Dense tracking and mapping in real-time," in *Computer Vision (ICCV), 2011 IEEE International Conference on*, nov. 2011, pp. 2320–2327.
- [16] A. Geiger, J. Ziegler, and C. Stiller, "Stereoscan: Dense 3d reconstruction in real-time," in *Intelligent Vehicles Symposium (IV), 2011 IEEE*, june 2011, pp. 963–968.
- [17] P. Newman, D. Cole, and K. Ho, "Outdoor slam using visual appearance and laser ranging," in *Proc. of IEEE ICRA*, 2006, pp. 1180–1187.
- [18] J. Sturm, S. Magnenat, N. Engelhard, F. Pomerleau, F. Colas, W. Burgard, D. Cremers, and R. Siegwart, "Towards a benchmark for rgb-d slam evaluation," in *Proc. of the RGB-D Workshop on Advanced Reasoning with Depth Cameras at Robotics: Science and Systems Conf. (RSS)*, Los Angeles, USA, June 2011.
- [19] A. V. Segal, D. Haehnel, and S. Thurn, "Generalized-icp." In *Robotics: Science and Systems*, 2009.