2012 IEEE/RSJ International Conference on
Intelligent Robots and Systems
October 7-12, 2012. Vilamoura, Algarve, Portugal

A robust RGB-D SLAM algorithm

Gibson Hu[1], Shoudong Huang[1], Liang Zhao[2], Alen Alempijevic[1] and Gamini Dissanayake[1]

Abstract— **Recently RGB-D sensors have become very popular in the area of Simultaneous Localisation and Mapping (SLAM). The major advantage of these sensors is that they provide a rich source of 3D information at relatively low cost. Unfortunately, these sensors in their current forms only have a range accuracy of up to 4 metres. Many techniques which perform SLAM using RGB-D cameras rely heavily on the depth and are restrained to office type and geometrically structured environments. In this paper, a switching based algorithm is proposed to heuristically choose between RGB-BA and RGB-D-BA based local maps building. Furthermore, a low cost and consistent optimisation approach is used to join these maps. Thus the potential of both RGB and depth image information are exploited to perform robust SLAM in more general indoor cases. Validation of the proposed algorithm is performed by mapping a large scale indoor scene where traditional RGB-D mapping techniques are not possible.**

I. INTRODUCTION

Simultaneous Localisation and Mapping (SLAM) is the process for a mobile robot to build a map of an unknown environment and simultaneously use the map to localise itself. Although many different SLAM techniques have been proposed in the past, the problem can still be very challenging especially when the quality of the sensor is poor and/or the environment has large scale and structural changes over the course of the cameras path. For instance, a RGB-D camera moving through a narrow corridor can quickly move into open space whereby the environment scale changes drastically. Clearly, a SLAM algorithm should be capable of maintaining consistent maps under these environmental changes but to our knowledge, no researchers have yet tackled this issue regarding RGB-D cameras. In this paper, our goal is to provide a framework which maximises the potential of RGB-D cameras as a viable sensor for indoor SLAM.

One of the most recent developments in sensing technology has come in a form of structured light based cameras, Microsoft Kinect and Asus Xtion Pro are both examples of sensors based upon PrimeSense's technology [1]. These sensors have been categorised under RGB-D cameras, due to the fact that they provide two image types: primary RGB based coloured image and a secondary depth image (where each pixel is representative of a range measurement). They can be further classified under active sensors [2] as the sensor must constantly emit a projected light structure. In

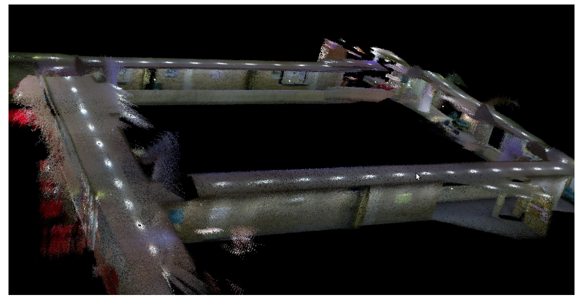

Fig. 1. 3D map generated using our proposed RGB-D SLAM algorithm

our observations the Kinect sensor's depth becomes heavily discretized and sparse after 4 metres. The overall depth uncertainty seems to grow exponentially and is completely unreliable passed this operating range.

Recent approaches to RGB-D SLAM have focused purely on using raw depth [3] or a combination of depth and RGB information [4] to predict camera ego-motion. The combination type techniques have shown generally more robustness since they take both visual and geometric information into the motion estimation. However when dealing with complex indoor environments, two scenarios can occur that cause traditional RGB-D type algorithm to fail. These are large spatial scene where little to no depth information is gained and planar scenes, where geometric structure is greatly lacking. This is the reason why these techniques have restricted themselves to office type environments and niched towards object reconstruction type scenarios. Since RGB-D sensor have a pure vision based component, the obvious approach is to see how monocular SLAM can overcome this problem.

In the field of computer vision, monocular SLAM is a technique often used to acquire camera motion over several frames of a single camera. The maps generated are typically optimised to a set scale value where the correct value can be determined by post or prior measurements [5]. Unfortunately the consistency of scale is hard to maintain using just visual odometry and scale drift is unavoidable for large scale problems when only 2D image information is used [6].

This paper proposes a new RGB-D SLAM technique which will selectively choose either an RGB-D or monocular SLAM based local map building architectures. The key is to obtain accurate visual odometry while maintaining correct metric scale for our maps. We choose to use the popular Sparse Bundle Adjustment (SBA) implementation together with eight point RANSAC [7] for computing our monocular

[1]Centre for Autonomous Systems, Faculty of Engineering and Information Technology, University of Technology, Sydney, Australia. {Gibson.Hu, Shoudong.Huang, Alen.Alempijevic, Gamini.Dissanayake}@uts.edu.au. [2]Institute of Remote Sensing and GIS, School of Earth and Space Science, Peking University, Beijing China. lzhao@pku.edu.cn

978-1-4673-1736-8/12/S31.00 ©2012 IEEE

maps. In the alternative, RGB-D SLAM maps are created by applying SBA on a two step re-projection RANSAC (RE-RANSAC), iterative closest point (ICP) approach. A heuristic switching algorithm constantly monitors camera motion, image features, environmental structure to determine which visual odometry case is best suited. Furthermore, we can guarantee global map consistency in terms of scale and uncertainty by applying scale based local map joining [8]. Fig. 1 is an example of the type of maps generated by our algorithm.

The remainder of the paper is organized as follows. Section II discusses related work. Section III presents the proposed approach in details. Experimental results are provided in Section IV and finally, conclusions are drawn and future work proposed in Section V.

II. RELATED WORK

Mapping algorithms using RGB-D information have been an active field of research for some time, brought back into the spotlight after the recent advancements of RGB-D technology, mainly the PrimeSense structured light camera [1]. A significant body of work has gone into utilizing these sensors for addressing the problem of SLAM using both feature based and appearance based approaches. The pioneering work done by Nister et al. [9] combined feature extraction and matching with RANSAC to obtain accurate pose estimation using stereo cameras. Izadi et al. [3] tackled SLAM by splitting the tracking and mapping into separate problems. In this work the camera is tracked over a global model of the environment through an ICP algorithm. The environment model is made denser and refined over several observations such that tracking is improved. Unfortunately pure ICP relies on very structured environments, thus, limiting this technique to only a room type scenario and not a general indoor case. Furthermore, keeping track of a dense model can be computationally expensive and will not scale very well.

Henry et al. [4] proposes a method to jointly optimizes visual features based RANSAC with ICP. This approach is simplistic in nature but able to work very intuitively with RGB-D cameras. During optimisation, bundle adjustment is applied to obtain consistent maps. Audras et al. [2], argued against feature based method due to their uncertain nature in matching and detection by introducing an appearance base approach. There dense stereo tracking technique, converted to work with the RGB-D camera, predicts relative camera motion by minimizing a warping function between frames given 3D point cloud information.

For all these methods to work successfully in the general indoor case, sufficient depth information must be provided. Rather than treating this as a sensor limitation, our approach allows for pure RGB based maps (such as monocular SLAM) to be executed in areas where failure in depth based mapping could occur.

Work in monocular SLAM has typically diverged into two categories, filtering and optimisation. In filters, Civera et al. [10] proposed a feature based approach to improve efficiency and data association in which inverse depth is used in an Extended Kalman Filter (EKF) for feature parameterization. Doing so allows uncertainties, for distant features, to be represented accurately. Features are only brought into XYZ domain after sufficient certainty has been reached. This was extended by applying 1 point RANSAC [11], further reducing computation cost from data association and matching. The problem with filter based methods is that they suffer from inconsistency and require good model selection. However, they are efficient and heavily coupled with data association, which makes them very popular in monocular SLAM.

In the field of optimisation, the bundle adjustment algorithm [6] is a highly regarded technique. Essentially it minimizes the re-projection error between feature observations by solving the nonlinear least squares problem. Over time many researchers have improved on this idea by exploiting the sparseness of the Jacobians to reduce complexity and tailoring Levenberg-Marquardt (LM) algorithms to guarantee convergence. Klein and Murray [12] addressed the problem of rapid camera motion by separating the tracking and mapping problem. Bundle adjustment is done on key frames and the camera location is tracked using the current feature map. However, an estimate of feature uncertainties can not be obtained, thus, the algorithm cannot handle large loop closures.

Indoor monocular SLAM can be considered a very different problem to outdoor, as features tend to be more erratic in nature and seldom remain in the field of view long enough to be properly applied in an observation model. In addition, poor data association can often occur due to large scene differences. Furthermore, scale drift is still a key issue in large scale SLAM when purely using a single camera.

In our proposed approach we utilize complimentary advantages of both RGB-D and monocular SLAM to build an overall framework in RGB-D mapping. Doing so, we are able to overcome some of the fundamental problems as described above.

III. THE PROPOSED RGB-D SLAM ALGORITHM

This section describes, in detail, the proposed algorithm.

A. RGB-D SLAM framework

The components required to perform the proposed robust RGB-D SLAM in a general indoor scene is shown in Fig. 2.

The global map is generated from the combination of two types of local maps. First is RGB-D Bundle Adjustment (RGB-D-BA), the name suggest that the map is created using both depth and image information. Alternatively, the second map (RGB-BA) is created only using image. When the local maps are being built, a monitoring components looks at the status of one of the visual odometers, and makes predetermination of which technique is more preferred. When a switch occurs a local map is built and placed into the global map building framework. Each local map builds up a set of key frame in which loop closure can be performed later. Map joining is applied incrementally between the current global

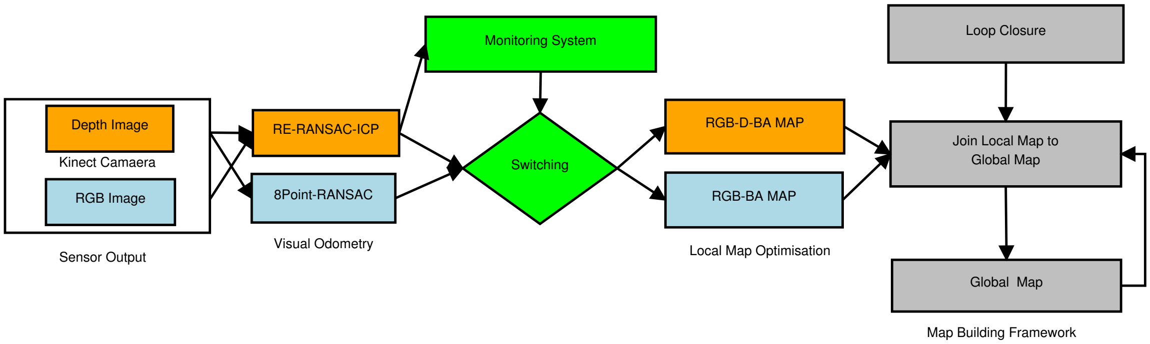

Fig. 2.   Overview of the proposed RGB-D SLAM algorithm

and local map, with the result being the new global map. When the algorithm ends, the final map can be visually represented by its set of camera poses.

B. RGB-D-BA SLAM local map

For this map we adapt Henry's [4] two stage RGB-D ICP method for creating RGB-D maps. In their paper they state the method to be more accurate compared to Re-Projection or Euclidean error type RANSAC techniques. Also it is only slightly less accurate compared to the joint optimization strategy [4], which doubles the computation time.

The steps to this technique are as follows. First feature are found between two successive images. In our case FAST features [13] are used. Since the Kinect camera provides us with one to one depth and RGB correlation, only features with depth values associated are used. Currently a hard threshold of 4 metres is applied to the depth values.

Matching is performed on the features using a simple K Nearest Neighbor search on feature descriptors. This type of matching criteria is heuristic in nature and can often lead to wrong matches. Therefore outliers are removed using re-projection error RANSAC (RE-RANSAC) explained in [4].

For the final set of inliers, an LM least squares minimization is applied to obtain the final transform T.

$$T^* = \underset{T}{\operatorname{argmin}} \left(\frac{1}{n} \sum_{i \in A} (proj(x'_i), proj(\hat{x}_i)) \right) \tag{1}$$

In the equation (1), for the set of inliers A, $\hat{x} = (\hat{x}, \hat{y}, \hat{z})$ is the estimated feature location and x' is the observed. $proj$ is the function to re-project into camera image frame (u, v, d), where d is the disparity. The second stage simply uses T as an initial estimate for Iterative Closest Points (ICP). By running this, ICP becomes the stronger geometric constraint that refines the image based RANSAC solution. To prevent erroneous depth measurements by the sensor we have a preemptive step in which we apply a bilateral filter on the depth image [14]. Doing so removes ICP measurements from noisy edge readings without augmenting any other depth values.

For our ICP algorithm we employ the state of the art Generalised ICP [15], which is a type of hybrid point-to-point and point-to-plane implementation that can overcomes limitations present in both methods. We found that the point clouds can be down sampled by a factor of 3 to speed up processing without any affect on accuracy.

In our local map we will jointly optimize between camera poses and feature poses. To do this, Sparse Bundle Adjustment (SBA) is applied. Since T found from ICP is disconnected from T obtained from RE-RANSAC, in SBA, new features should be found which are relevant to the ICP algorithm. These are found and matched similarly to [4].

The SBA algorithm has higher information consistency and accuracy over popular pose graph based methods [16], in which features are marginalized. Also, since our initial estimates of scale is inherited from the poses and features, the final local maps are metrically correct.

C. RGB-BA SLAM local map

Firstly, feature detection and matching is identical to RGB-D-BA. However features without depth are not filtered out and inliers are achieved through an epipolar geometry model.

The RANSAC approach, in this case involves randomly sampling 8 points and solving the direct linear solution for the fundamental matrix [7]. Then using the gold standard algorithm [7], we can estimate the best fundamental matrix which fits the resultant inliers.

Lastly, given that the camera is calibrated, decomposition is applied to recover the relative camera transform. Note these transforms obtained are only up to a scale where the scale consistency between multiply camera poses can only be maintained through optimisation. Once we have all the camera poses, the feature locations can be determined by linear triangulation [7].

To optimize this local map, the same SBA algorithm [6] is used.

D. Heuristic switching

To plan how to switch between the two techniques for local map building, an analysis of their properties and failure modes is made. For simplicity, we look out how the visual odometries compare against vehicle encoder. The critical assumption in our current implementation is that there are always enough features observed for at least one mapping approach to continue functioning. We have considered this is not always the case, and see this as a shortcoming to be addressed in future research.

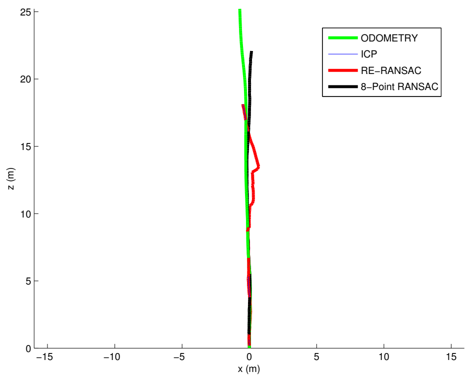

Fig. 3. Depth drop off

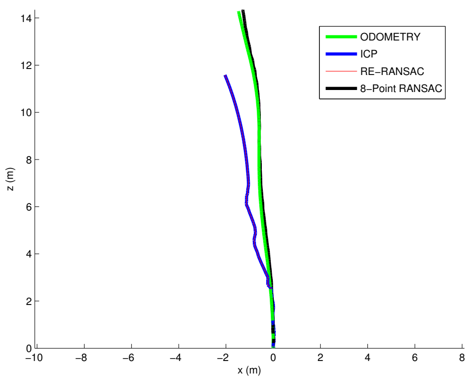

Fig. 4. Poor structure

1) Case 1: Depth drop off: When features are detected in the RGB image and little or no depth association is available, The first conclusion, is that the scene is now too far away or too close for the camera to pick up any reliable depth. Fig. 3 shows the case where depth information slowly drops until no depth is available. We can see that both RE-RANSAC and ICP cannot recover from this error, however 8 Point RANSAC stays consistent as compared with odometry. We can also see that ICP begins performs significantly poorer due to bad initial values.

2) Case 2: Poor geometric structure: When the environment does not have enough structure, for example the camera only observing a wall while rotating or translating (occurring in Fig. 4), we see RGB-D visual odometry induces incorrect translations. This can be associated with a phenomenon known as sliding [17], where ICP cannot determine where the best fit lies along a plane. Although a good initial estimate can sometimes prevent sliding, upon observations it seems the two step approach seems heavily susceptible. We can associate this with ICP and Re-projection-error RANSAC not being jointly optimised.

3) Case 3: Feature clustering: Even though there is enough depth information for RGB-D visual odometry to proceed, the features with depth can often clustered into one region. Due to the nature of RANSAC, more clustered features will be picked over spatially spread features, leading to poor transformation estimates. The main cause of this is feature distribution in the RGB scene. Since our assumption is that the scene is always feature rich, we found that case 3 and case 1 can be somewhat related, since clustering of features typically means reduced coverage of depth information, a cluster can often mean a depth drop off may be imminent.

4) Case 4: Accuracy: One question to pose is why not use RGB-BA only, since there are far less failure modes? The most important aspect to consider when dealing with RGB-D sensors like the Kinect, is that depth information within the workspace of the sensor, is typically more reliable than the image information. The RGB camera on both the Kinect and Xtion are of poor quality. With resolutions of 640x480, there are high levels of noise due to motion blur, lighting and sensor quality. In a normal scene where both maps are able to work simultaneously, ICP should always be preferred. In this analysis we do not address the failure mode of our RGB-BA map, however, further investigation is warranted.

5) The Heuristic: We identified 4 cases which needed to be addressed. The quantity of depth features, motion through planar structures, feature cluster based on spatial locations and mapping technique preference. Fig. 3 and Fig. 4 shows a plot between how each degeneracy has on transformation error. With this in mind, we propose thresholding equations to determine when failure has happened.

In our heuristic we want to address all 4 cases by looking directly at RE-RANSAC output. Case 1 is simply handled by $n < \gamma_1$ where γ is the thresholding value and n is the number of inliers. For Case 2, we threshold to a cluster score α. First a centroid C is calculated from our image inliers, α is then found using a distributed average over the normal distances to C, expressed in (2). Then our threshold becomes $\alpha < \gamma_2$. To handle Case 2, we fit a plane to all inlier points and then calculate β according to the residual to the normal N, expressed in (3). Another thresholded is applied to this value $\beta < \gamma_3$. If any thresholds are violated, a switch is made to RGB-BA mapping.

$$\alpha = \frac{1}{n}\sum_{i=1}^{n} d(C, x_i) \tag{2}$$

$$\beta = \frac{1}{n}\sum_{i=1}^{n} d(N, x_i) \tag{3}$$

The switch criteria is monitored over a set of frames F before a determination is made, so as to guarantee noisy image readings directly affecting the decision. We have found $F = 5$ frames to be suitable.

E. Map joining and global map consistency

Initial we define our global map as being M, this map can be either. M_{rgbd} map made by RGB-D-BA or M_{rgb} map made by RGB-BA. Note that M_{rgb} can only build maps which is accurate up to a scale. In joining, if either one or both maps are of M_{rgb}, then the relative scale should be estimated.

The overall architecture for map joining is a modified 3D Iterative Sparse Local Submap Joining Filter (I-SLSJF) [8].

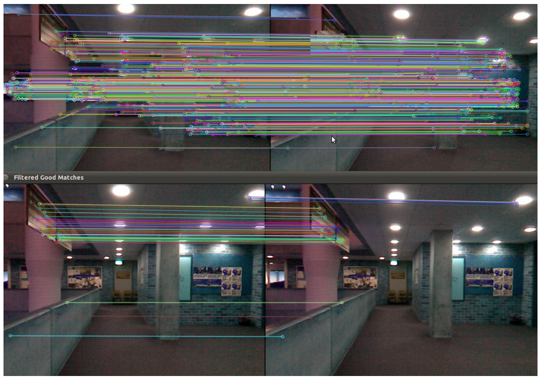

Fig. 5. Top: Inliers found by 8point RANSAC, Bottom: Inliers found by RE-RANSAC, example of depth loss and feature clustering

3D I-SLSJF algorithm is an extension of the 2D I-SLSJF (code available on OpenSLAM.org) and uses an Extended Information Filter (EIF) to fuse the local maps in sequence. This approach is computationally more efficient than the typical maximum likelihood method. The modified algorithm is able to optimize relative scale. Since scale can be arbitrary in M_{rgb}, the metric scale for the global map cannot be recovered until a M_{rgbd} map is joined.

Map joining in EIF can be expressed in this way. Suppose local map j is given by $(\hat{X}_j^L, I_j^L)$ where $\hat{X}_j^L$ is the state estimate and I_j^L is the associated information matrix. The local map state estimate $\hat{X}_j^L$ can be regarded as an integrated "observation" of the true relative poses from the cameras start pose and features of the local map to the other poses and features involved in the local map. That is,

$$\hat{X}_j^L = H_j(X_{MJ}) + w_j \tag{4}$$

where X_{MJ} is the state vector including the poses involved in all the local maps, $H_j(X_{MJ})$ is the relative pose/feature "observation" function and w_j is the zero-mean Gaussian "observation noise" whose covariance matrix is $P_j^L = (I_j^L)^{-1}$.

So the problem of joining local maps 1 to k is to estimate the global state X_{MJ} using all the local map information (4) for $j = 1, \cdots, k$. This problem can be formulated as a least squares problem. That is, finding X_{MJ} such that

$$\sum_{j=1}^{k} (\hat{X}_j^L - H_j(X_{MJ}))^T (I_j^L)(\hat{X}_j^L - H_j(X_{MJ})) \tag{5}$$

is minimized [18].

Since the local maps are of good quality and the uncertainties of the local maps are taken into account in map joining, the consistency of the final global map is maintained.

During map joining, the accuracy of the maps are influence heavily by accurate data association between observed features. We propose a duel directional search scheme to handle this. Firstly starting at the point where the algorithm has switched we simultaneously look forward and backwards in camera frames to find associated features. When the feature overlap drops we stop the search. Since optimisation does not affect redundant features (features that are not in both maps), we marginalize these out to reduce computation cost.

F. Loop Closure (LC)

Loop closure is handled using key frame selection [4]. During which local maps are checked for scene difference between consecutive frame. The first frame is selected as the current key frame, then every new frame is compared with the current. If the feature matches fall below a set threshold a new key frame is created.

After a new local map is joined to the global map, the key frames within the local map is matched to a database. If there is a significant overlap, a relative pose constraints is added into the existing global map. Once new relative pose constraints are known applying a standard 3D I-SLSJF to optimize the global map with the new constraints is trivial. Identifying a loop will lock the map into the correct configuration as well as maintain global consistency.

The current drawback to our loop closure is that key frame features must have depth information, otherwise this will disrupt the scaling of the global map.

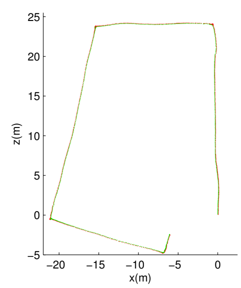

Fig. 6. Before LC

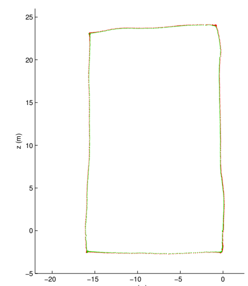

Fig. 7. After LC

IV. EXPERIMENTAL RESULTS

To evaluate the proposed approach we conducted an experimental over a large-scale scene where the three failure modes of RGB-D-BA are present. Triggering the algorithm to switch between the two methods for local maps building.

The test was carried on a robotics platform named Robot Assist [19]. The image data from the Kinect camera was logged as the robot traversed through a large loop on level 6, Engineering Building located at the University of Technology, Sydney. The rectangular loop is 25m across and 15m wide as seen in Fig. 1 and Fig. 8. A total of 1670 Kinect images were collected along the loop. No other sensor data is used in the proposed SLAM algorithm.

For this dataset, the proposed switching algorithm is invoked 8 times and ten local maps are built. The thresholds are set to $\gamma_1 = 50$, $\gamma_2 = 70$ and $\gamma_3 = 2$. Five are built by RGB-BA and the other five are built by RGB-D-BA. As seen in Fig. 8, when camera approaches corners more depth features are found and RGB-D-BA are used (indicated by red solid box); when the camera is moving along the corridors, RGB-BA is triggered (indicated by blue dashed box). This is because features with depth are typically clustered tightly along the right wall and sometimes very planar. The loop closure via key frame matching is indicated by green dashed box. The final map shown in Fig. 1 is created by re-rending the coloured points clouds on the optimised trajectory. To check the accuracy of the map built, we overlay the point cloud projected by our final trajectory on top of a 2D floor architectural plan as shown in Fig. 8. It can be seen that the estimated map is visually accurate.

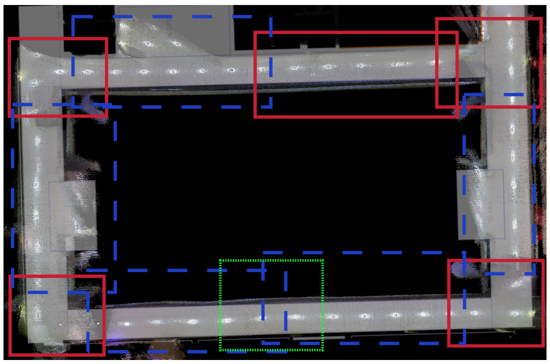

Fig. 8. Comparison of RGB-D SLAM MAP (background layer) and Floor Plan (transparent layer). Local maps are represented by, RGB-BA (Blue Dashed), RGB-D-BA (Red Solid), Loop Closure (Green Dashed)

V. CONCLUSIONS

In this paper, we have proposed a robust algorithm for SLAM using RGB-D sensors. The algorithm first builds local maps either using vision only (RGB-BA) or vision and range depending (RGB-D-BA) on the different scenarios, then a map joining algorithm is applied to combine all the local maps. By applying the heuristic switching, the algorithm is able to handle various failure modes associated with RGB-D-BA. The map joining strategy reduces the computational cost significantly and the algorithm has great potential to be applied to RGB-D SLAM in more general large-scale environments.

There are still some shortcoming in which we will factor into future work. Data association, currently we are applying a naive based approach where features descriptors are matched between consecutive frames. These features are tracked based on how well they flow through these frames. However due to image noise, features tracks are quickly lost and never recovered. Secondly, handling degenerate cases when the camera only observes planar surfaces has been challenging. Due to the nature of the 8 point algorithm, it is incapable of handling this case properly when RGB-BA is used. However work on image homography may give some insight into how best to tackle this problem. Finally, many features are marginalize after map joining, so they cannot be used in loop closure. However without this the complexity of optimisation is heavily increased. This is the main reason why we have resorted to use pose constraints instead.

Overall we feel that RGB-D type sensor are well suited for mapping both small and large scale indoor environments. The next step is towards a fully operational robust real time mapping system.

REFERENCES

[1] B. Freedman, A. Shpunt, M. Machline, and Y. Arieli, "Depth mapping using projected patterns," 2007, uS Patent App. 11/899,542.

[2] C. Audras, A. Comport, M. Meilland, and P. Rives, "Real-time dense RGB-D localisation and mapping," in *Australian Conference on Robotics and Automation*, Monash University, Australia, 2011.

[3] R. A. Newcombe, A. J. Davison, S. Izadi, P. Kohli, O. Hilliges, J. Shotton, D. Molyneaux, S. Hodges, D. Kim, and A. Fitzgibbon, "Kinectfusion: Real-time dense surface mapping and tracking," in *Mixed and Augmented Reality (ISMAR), 2011 10th IEEE International Symposium on*, oct. 2011, pp. 127 –136.

[4] P. Henry, M. Krainin, E. Herbst, X. Ren, and D. Fox, "Rgb-d mapping: Using kinect-style depth cameras for dense 3d modeling of indoor environments." SAGE Publications, 2012.

[5] D. Scaramuzza and F. Fraundorfer, "Visual odometry [tutorial]," *Robotics and Automation Magazine, IEEE*, vol. 18, no. 4, pp. 80–92, 2011.

[6] M. Lourakis and A. Argyros, "The design and implementation of a generic sparse bundle adjustment software package based on the levenberg-marquardt algorithm," *Institute of Computer Science-FORTH, Heraklion, Crete, Greece, Tech. Rep*, vol. 340, 2004.

[7] R. Hartley and A. Zisserman, *Multiple view geometry in computer vision*. Cambridge Univ Press, 2000.

[8] L. Zhao, S. Huang, L. Yan, J. Wang, G. Hu, and G. Dissanayake, "Large-scale monocular slam by local bundle adjustment and map joining," in *Control Automation Robotics and Vision (ICARCV), 2010 11th International Conference on*. IEEE, 2010, pp. 431–436.

[9] D. Nister, O. Naroditsky, and J. Bergen, "Visual odometry," in *Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on*, vol. 1, 2004, pp. I–652.

[10] J. Civera, A. Davison, and J. Montiel, "Inverse depth parametrization for monocular slam," *Robotics, IEEE Transactions on*, vol. 24, no. 5, pp. 932–945, 2008.

[11] J. Civera, O. Grasa, A. Davison, and J. Montiel, "1-point ransac for ekf-based structure from motion," in *Intelligent Robots and Systems, 2009. IROS 2009. IEEE/RSJ International Conference on*. Ieee, 2009, pp. 3498–3504.

[12] G. Klein and D. Murray, "Parallel tracking and mapping for small AR workspaces," in *Proc. Sixth IEEE and ACM International Symposium on Mixed and Augmented Reality (ISMAR'07)*, Nara, Japan, November 2007.

[13] E. Rosten and T. Drummond, "Machine learning for high-speed corner detection," *Computer Vision–ECCV 2006*, pp. 430–443, 2006.

[14] C. Tomasi and R. Manduchi, "Bilateral filtering for gray and color images," in *Computer Vision, 1998. Sixth International Conference on*. IEEE, 1998, pp. 839–846.

[15] A. Segal, D. Haehnel, and S. Thrun, "Generalized-icp," in *Proc. of Robotics: Science and Systems (RSS)*, 2009.

[16] G. Grisetti, R. Kummerle, C. Stachniss, U. Frese, and C. Hertzberg, "Hierarchical optimization on manifolds for online 2d and 3d mapping," in *Robotics and Automation (ICRA), 2010 IEEE International Conference on*. IEEE, 2010, pp. 273–278.

[17] N. Gelfand, L. Ikemoto, S. Rusinkiewicz, and M. Levoy, "Geometrically stable sampling for the icp algorithm," in *3-D Digital Imaging and Modeling, 2003. 3DIM 2003. Proceedings. Fourth International Conference on*. IEEE, 2003, pp. 260–267.

[18] G. Hu, S. Huang, and G. Dissanayake, "3d i-slsjf: A consistent sparse local submap joining algorithm for building large-scale 3d map," in *Decision and Control, 2009 held jointly with the 2009 28th Chinese Control Conference. CDC/CCC 2009. Proceedings of the 48th IEEE Conference on*, 2009, pp. 6040 –6045.

[19] N. Kirchner, A. Alempijevic, S. Caraian, R. Fitch, D. Hordern, G. Hu, G. Paul, D. Richards, S. Singh, and S. Webb, "Robotassist-a platform for human robot interaction research," in *Proc. of the 2010 Australasian Conference on Robotics and Automation (ACRA)*, vol. 10, 2010, pp. 1–10.