

Copyright © 2012 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.

A Modular Hybrid SLAM for the 3D Mapping of Large Scale Environments

Jared Le Cras

Department of Mechanical Engineering
Curtin University
Bentley, Western Australia

Jonathan Paxman

Department of Mechanical Engineering
Curtin University
Bentley, Western Australia

Abstract—Underground mining environments pose many unique challenges to the task of creating extensive, survey quality 3D maps. The extreme characteristics of such environments require a modular mapping solution which has no dependency on Global Positioning Systems (GPS), physical odometry, *a priori* information or motion model simplification. These restrictions rule out many existing 3D mapping approaches.

This work examines a hybrid approach to mapping, fusing omnidirectional vision and 3D range data to produce an automatically registered, accurate and dense 3D map. A series of discrete 3D laser scans are registered through a combination of vision based bearing-only localization and scan matching with the Iterative Closest Point (ICP) algorithm. Depth information provided by the laser scans is used to correctly scale the bearing-only feature map, which in turn supplies an initial pose estimate for a registration algorithm to build the 3D map and correct localization drift. The resulting extensive maps require no external instrumentation or *a priori* information.

Preliminary testing demonstrated the ability of the hybrid system to produce a highly accurate 3D map of an extensive indoor space.

Keywords—SLAM, localization, 3D mapping, mining, sensor fusion, omnivision

I. INTRODUCTION

Three dimensional mapping of large scale environments is a challenging problem which has been extensively studied with reference to a range of applications. Many approaches have yielded maps of high quality in a variety of situations. The majority of 3D mapping systems do not apply well in the domain of underground mines, where the lack of satellite signals, the unreliability of physical odometry and the difficulty of introducing fiducial tags, pose particular challenges. There are several key motivations for producing a hybrid system that can overcome these challenges. Removing the dependency on external equipment such as global positioning system (GPS) satellites or targeting systems allows the device to be operated in unmonitored environments. Eliminating physical odometry measurements removes the dependence on specialized mobile platforms or vehicles. And finally, avoiding the use of *a priori* information allows the mapping of completely unknown environments. The system should also be capable of operating in six degrees of freedom to ensure the widest range of applicable environments. The production of survey quality

results is essential, as the goal application for this system is surveying tasks in underground mining environments. Reflectorless surveying has an accuracy of $\pm 6\text{mm}$, therefore highly accurate alignment of dense point clouds is required.

Previous successfully implemented examples of large scale 3D mapping have relied on GPS [5], physical odometry [12] [8], [9], *a priori* information [3] or a simplified approximation [14] [4]. The work presented in this paper demonstrates a self-contained hybrid SLAM system capable of large scale 3D mapping based on the use of a 3D laser and omnidirectional camera. The fusion of real time bearing-only vision-based localization and intermittent, single point of origin, 3D laser scans has resulted in a system that can accurately perform localization while building a detailed 3D map of the surrounding environment.

An initial stationary 3D laser scan is performed, followed by six degree of freedom bearing-only localization during transit to the next scan location. The bearing-only localization is enhanced by the incorporation of range information at the discrete 3D scan positions. This enables correct scaling of the resulting feature map, and improves the general performance of the bearing-only localization. Once the target location has been reached and the subsequent 3D scan produced, an iterative closest point (ICP) algorithm [1] [11], is applied to register the new laser data to the existing map, consequently compensating the trajectory estimate of the vehicle to accurately reflect the real world location. This process is repeated to iteratively build a large scale 3D map of the environment.

II. RELATED WORK

Large scale 3D mapping has been successfully accomplished by many research groups through the implementation of complex, non-modular systems. This non-modularity may refer to a number of design decisions such as the use of externally instrumented localization systems (e.g. GPS), external physical odometry readings, *a priori* information or a simplified motion model that removes the comprehension of higher degrees of freedom. These design choices, while appropriate for the given domains, are not well suited to our application.

The traditional commercial technique for registering multiple 3D laser scans uses artificial markers placed throughout the scene. These markers are used in post-processing to semi-automatically align the individual scans.

There are many shortcomings to this approach: the scene must be artificially manipulated before scanning can commence, all registration must be completed in a post-processing phase and there is significant manual intervention required. The resulting process is slow and labor intensive.

GPS-INS (Global Positioning System / Inertial Navigation System) is often used in outdoor mobile mapping systems to provide an initial pose estimate for fine scale registration [5]. However, GPS-INS is not suitable for indoor or underground applications and can suffer from reliability problems in dense urban areas due to obstruction of satellite signals. The primary application for our own work on large scale 3D mapping is the underground mining industry [6] where there is no access to GPS positioning.

The typical response to the absence of a GPS signal in indoor environments is to make use of physical odometry readings to provide a rough pose estimate prior to fine scale registration [12]. These odometry readings are usually supplied by encoders attached to the wheels of the host platform and are prone to considerable error in environments with significant slip (e.g. gravel, sand or ice). Odometry can be used in conjunction with an Inertial Measurement Unit (IMU) or continuous scan registration to improve robustness to odometry errors and aid in the extrapolation of odometry readings to a six degree of freedom environment [8], [9]. Reliance on physical odometry precludes a truly modular and flexible mapping system due to dependence on the instrumentation of the vehicle itself. It is not possible to produce a single vehicle that is perfectly suited to all potential 3D mapping environments; therefore a modular solution is desirable. A modular mapping system can be deployed on any existing vehicle without integration into the vehicle's on-board systems. This modular proposal is examined in detail in Section III-A.

One method to remove the dependency on physical odometry is to instead use *a priori* information. Früh and Zakhor [3] use aerial photographs or Digital Surface Models to correct 2D laser based localization over large distances. Again, this is only possible in uncluttered outdoor environments and would not be feasible for the continually evolving profile of an underground mine where *a priori* information may not be available or accurate.

The mapping of indoor environments can often be simplified to a three degree of freedom problem due to the flat, level nature of many man-made environments. By removing roll, pitch and vertical translation, the six degree of freedom localization problem is reduced to three degrees of freedom. This allows localization to be produced by simple 2D laser scan registration where the reduced dimensionality considerably simplifies the computations required. The effectiveness of this approach is demonstrated by Thrun et al. [14] and Hähnel et al. [4]. We are interested in full six degree of freedom localization, so that the results can be extended to relatively complex non-flat environments.

The lack of a modular system required for the large scale mapping of underground mining environments led to the development of the approach detailed in the following section. Future testing in an underground mine will include the integration of our previous work on the compensation of vision

based localization in environments with dynamic illumination [7]. The reduction of the system to simply an omnidirectional camera and 3D laser scanner while maintaining full six degree of freedom functionality, results in a system which can produce maps in the majority of traversable environments.

III. LARGE SCALE 3D MAPPING

A. Platform and Data Collection

The key components of our large scale mapping system are a Point Grey Ladybug omnidirectional camera and a Leica ScanStation 3D laser scanner. The omnidirectional camera was mounted on the extended arm of a Cyclops robotic platform (see Fig. 1). During preliminary testing, laser data was collected from a tripod mount, independently of the camera, as the 3D laser exceeded the weight limit of the robotic deployment system (see Fig. 1). The distance between the laser and camera is minimized at scan locations so that data from both sensors can be overlaid. The sensors have subsequently been integrated into a single modular localization system which is mounted on a mine site certified vehicle.

The data collection process commences with the production of an initial stationary 3D laser scan. This scan has a 360° horizontal field of view and 270° vertical field of view, with a range of up to 300m. Once the scan is complete, the omnidirectional camera produces an image at the same location which is then used to render the 3D scan. This image has a horizontal field of view of 360° and a vertical field of view of 290°. Both sensors are then transported to the next designated scan location. During transit to the next scan location, the omnidirectional camera records a panoramic stream of images which is used for bearing-only vision based localization (discussed further in Section III-C). Upon arrival at the following scan location, another 3D scan is performed and the process is repeated.

For this preliminary dataset distances of around 20 meters were used between scans; however, line-of-sight distances of up to 40 meters were trialed successfully, assuming that the environment was feature rich (thus reducing localization drift).



Figure 1. (Left) The Point Grey Ladybug Omnidirectional Camera mounted on the Cyclops robotic platform. (Right) The Leica ScanStation on its tripod.

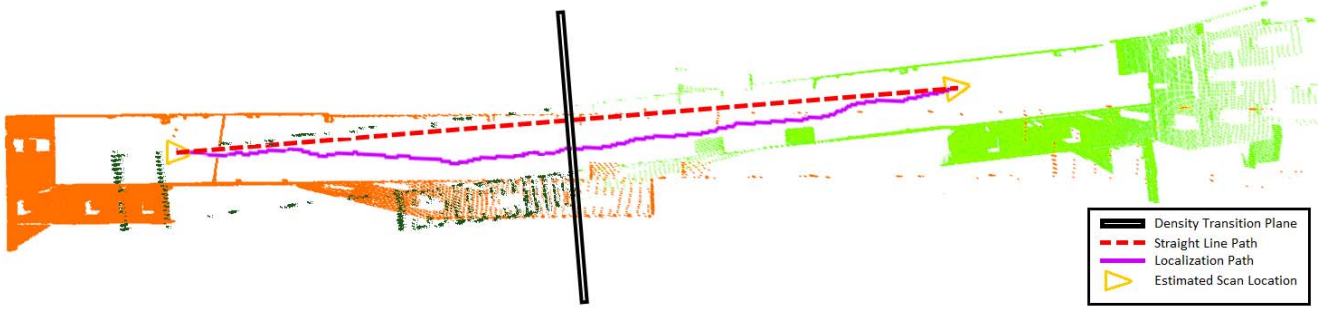


Figure 2. The cropping technique used to improve registration for scans with a separation greater than 20 meters. A density transition plane, located at the midpoint of the straight line path between scan locations, segments the new laser data into a dense segment (light green) and a sparse segment (dark green). The sparse segment then undergoes coarse registration onto the existing map (orange). The two point clouds above are scans 1 (orange) and 3 (light and dark green) of the architecture building dataset, taken approximately 40 meters apart.

The modifications to the technique necessary to produce accurate registration at intervals of 40+ meters are detailed in Section III-B.

B. Scan Registration and Pose Correction

To produce a high quality large scale 3D map, registration based on localization alone is not sufficient; a converging registration algorithm such as Iterative Closest Point (ICP) is required. ICP is an optimization algorithm which attempts to minimize the mean squared error (E) between two sets of points. The secondary point set (p) is iteratively translated and rotated, then compared to the primary point set (x).

$$E(R, t) = \frac{1}{N_p} \sum_{i=1}^{N_p} \|x_i - Rp_i - t\|^2 \quad (1)$$

For ICP to reach the correct convergence, an adequate initial pose estimate must be supplied otherwise the algorithm

is unlikely to find the global minimum [1]. In our system, this pose estimate is supplied by the vision based localization technique described in Section III-C. To correctly align the latest scan to the existing map, a coarse ICP algorithm is executed based on the initial pose estimate. Random seeding is used during the coarse ICP step to reduce dependence on initial pose accuracy. Coarse ICP is executed 20 times, each time using an initial pose estimate with a random offset of up to $\pm 30\%$ for each of the six degrees of freedom. The process is executed on reduced density point clouds (through voxel grid reduction) to keep processing time within acceptable limits.

Coarse registration of scans recorded up to 20 meters apart can be performed using the entirety of each scan due to the large areas of overlap present. However, for larger distances of over 20 meters, difficulties arise due to the reduced overlap and non-uniform point density of the scans. These factors produce the phenomenon seen in Fig. 2, where the sparse section of the

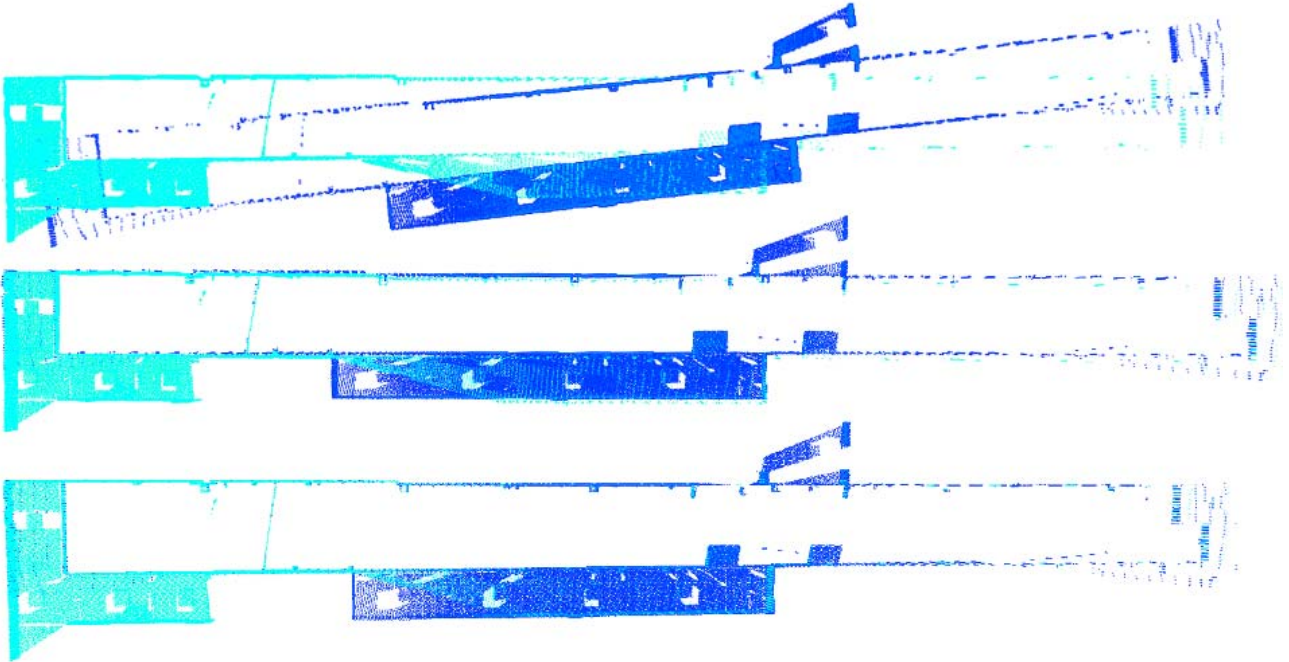


Figure 3. Alignment of scans 1 (light blue) and 2 (dark blue) from the architecture building dataset. (Top) Scan alignment based on initial pose estimate, mean squared error of 0.675m^2 . (Middle) Scan alignment after coarse registration, mean squared error of 0.178m^2 . (Bottom) Scan alignment after fine registration, mean squared error of 0.165m^2 .

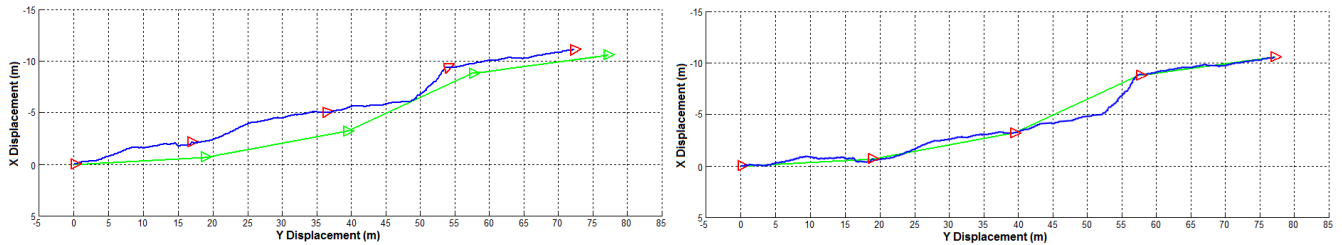


Figure 4. (Left) Comparison of uncompensated long term localization to the ground truth. (Right) Comparison of ICP-compensated long term localization to the ground truth. Both images show the 5 scan location estimates as red triangles.

new scan overlaps the dense section of the existing map. The sparse section of the existing map, in turn, overlaps with the dense section of the new scan. During the ICP registration process, the dense section of the data set carries significantly more weight than the sparse section due to the quantity based preference of mean square optimization. This weighting is applied despite the fact that the sparse section (usually containing less than 20% of the points) contains about half of the scanned surfaces. The overlap produced by large point cloud origin offset causes the ICP algorithm to focus on the alignment of the dense section of the new scan with the poorly distinguished sparse section of the existing map. The low number of target points during alignment with the existing map causes many of the new scan points to be associated with the same existing map point. This sharing of target points reduces the efficiency of the registration and makes it more prone to falling into local minima that satisfy a small, dense section of the new scan.

To correctly register point clouds with a large offset between scan origins, a technique was developed to segment the new scan before registration. A dividing plane called the ‘density transition plane’ is used to segment the newly acquired laser data. The density transition plane is a plane located at the midpoint of the straight line distance between scans, supplied by the vision-based localization. The plane is perpendicular to the straight line path and roughly represents the point at which the newly acquired laser scan becomes more dense than the existing map. The sparse segment of the new scan data is then used for coarse registration onto the existing dense map (see Fig. 2). The resulting transformation is then applied to the entirety of the new scan data. Superior results are obtained via this segmentation technique and processing time is also significantly reduced.

Upon coarse registration convergence, a fine ICP algorithm is run to provide the final alignment. For the results shown in Fig. 3, the alignment based on initial pose estimate had a mean squared error of 0.675m^2 , the alignment after coarse registration had a mean squared error of 0.178m^2 and the alignment after fine registration had a mean squared error of 0.165m^2 .

Once the fine registration step has successfully converged, the resulting transformation is also applied to the vehicle pose estimate. Without the use of a registration algorithm or loop closure, the monocular vision based localization algorithm tends to drift during long term localization tasks. The pose correction step results in a long term localization path that is highly correlated to the ground truth (see Fig. 4). The region of greatest distortion in the pose-corrected localization path occurs between scans 3 and 4. This region of the environment was paved in exposed brickwork, causing the suspensionless Cyclops robotic platform to shake significantly as it moved along. The resulting noise in the recorded image stream did not prevent localization; however there was a reduction in accuracy.

C. Bearing-Only Vision Based Localization

For a registration algorithm such as Iterative Closest Point (ICP) to successfully register two 3D point clouds, an initial pose estimate must be supplied (see Section III-B). This pose estimate is usually supplied through physical odometry or GPS. To allow our large scale mapping system to remain modular, an alternative technique was developed.

Bearing-only vision based SLAM has been demonstrated to be a highly robust localization tool over large distances when scaled using odometric information [2]. For application in an underground mining environment, an alternate scaling

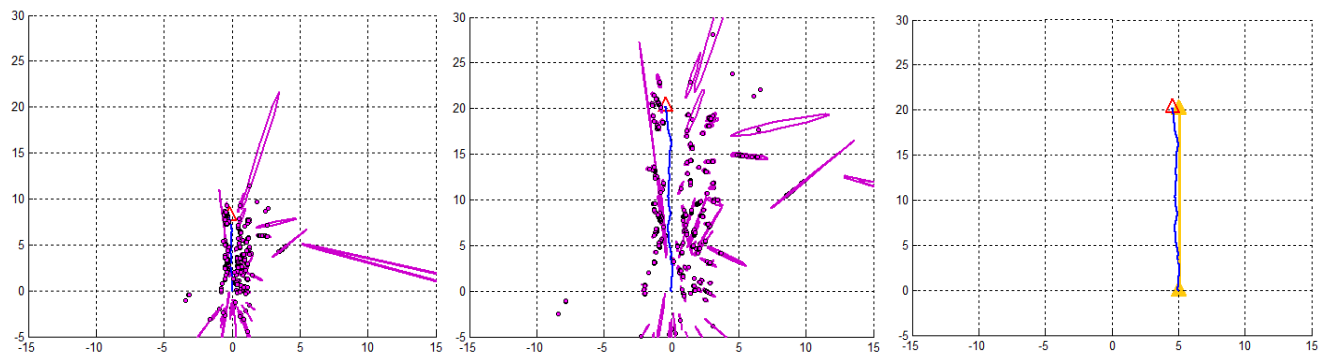


Figure 5. (Left) Feature mapping and localization before depth scaling. (Middle) Feature mapping and localization using depth scaling. Displacement in meters. (Right) Comparison of scaled localization path to ground truth. Displacement in meters.

technique must be used due to the likelihood of odometric errors caused by wheel slip and the desire to maintain a modular sensor package. Since the vision based localization commences at the same location as a 3D laser scan, features identified in the first few frames of the video feed can have a highly accurate depth estimate associated with them. In a similar way, the tracked features still active upon reaching the destination can also be scaled based on the subsequent laser scan.

At the initial scan location, features in the bearing-only vision based localization which have become well established in the first few frames of motion are situated in an initially dimensionless map. This map is then calibrated to real-world distances via the depth measurements obtained from the associated 3D laser scan. The resulting scaling factor is calculated for each well-established feature and then the mean of the scaling factors for the overall scan location is determined

(μ_{s1}). The same procedure is also performed to find the mean scaling factor at the destination scan location (μ_{s2}). These two mean scaling factors are then weighted based on the number of well-established features at each location (N_{f1}, N_{f2}), to produce the final scaling factor (s) for the localization path (see Equation 2). An example of the scaling factor being applied to a localization result can be seen in Fig. 5.

$$S = \frac{\mu_{s1}N_{f1}}{N_{f1}N_{f2}} + \frac{\mu_{s2}N_{f2}}{N_{f1}N_{f2}} \quad (2)$$

The rotation and translation matrices resulting from the vision-based localization (R_v, t_v), combined with the final scaling factor (s), are then used as the initial coarse transformation for the secondary laser scan (denoted L2), before it is correctly registered to the primary laser scan using a registration algorithm (see Equation 3).

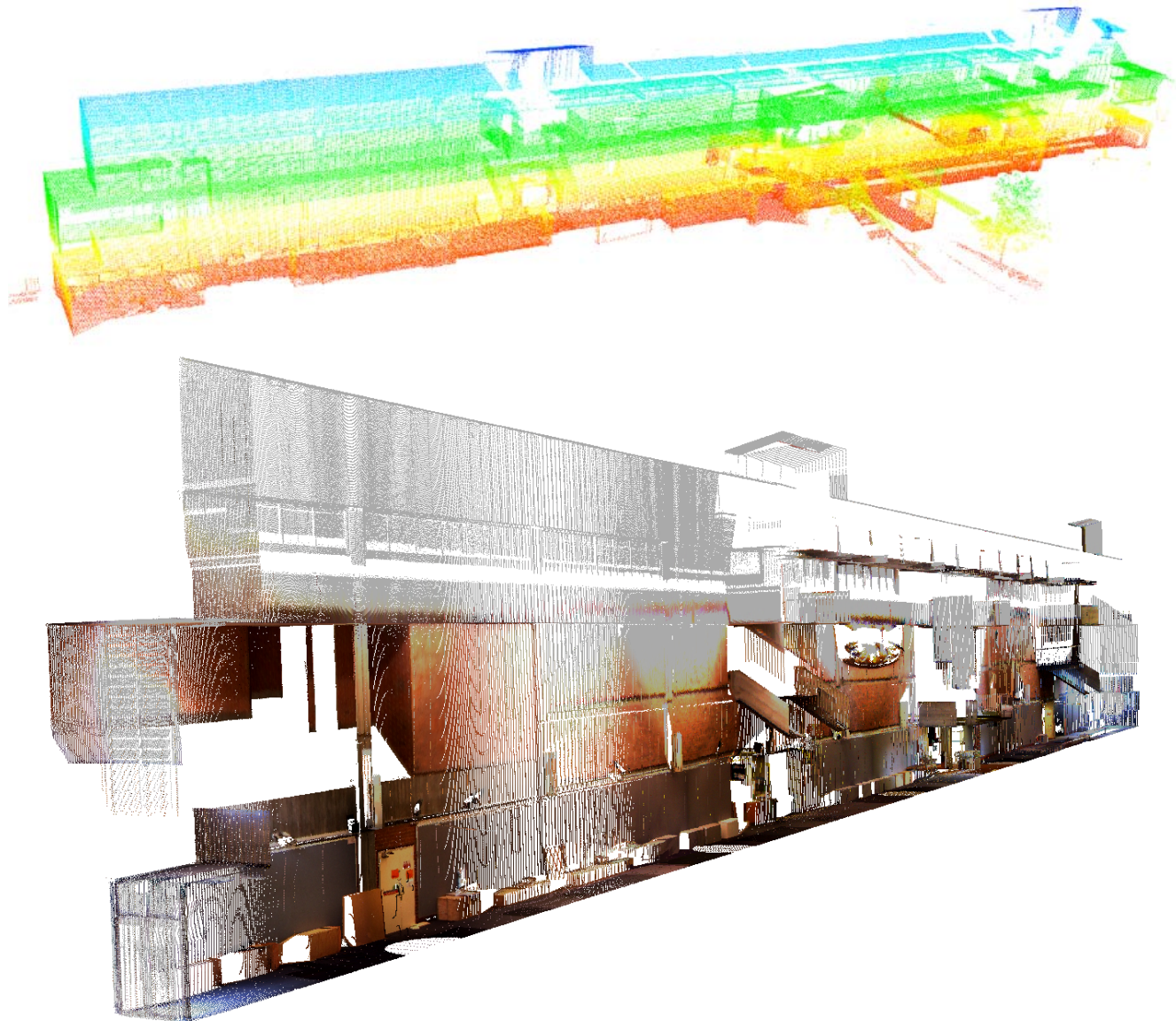


Figure 6. (Top) Final large scale 3D map of architecture building. (Bottom) Cross section of rendered map.

$$\begin{bmatrix} x_{L2'} \\ y_{L2'} \\ z_{L2'} \\ 1 \end{bmatrix} = \begin{bmatrix} sR_V & t_V \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x_{L2} \\ y_{L2} \\ z_{L2} \\ 1 \end{bmatrix} \quad (3)$$

The bearing-only vision-based SLAM algorithm applied in this work is based on the pinhole camera model version created by Civera et al. [2] which was converted to a spherical model [9] and then adapted to work with the Ladybug omnidirectional camera.

IV. MAPPING RESULTS

To evaluate the effectiveness of our large scale 3D mapping system, a dataset was collected within the architecture building on the Curtin University Bentley campus. A total of five 3D laser scans were recorded along the internal length of the building, roughly 20 meters apart. The scans were linked by localization produced from a stream of images collected between scan locations. The individual scans were then registered and rendered to produce the large scale 3D map seen in Fig. 6.

The results from this preliminary dataset demonstrated that the described technique produces a 3D map nearly indistinguishable from the map created by the commercial standard marker-based approach. An average error of 15mm was observed, well within the expected tolerance for a map of a 20m x 18m x 107m structure. A factor contributing to the long range accuracy of our technique is the registration of only a small number of high detail 3D scans as opposed to copious numbers of smaller 3D or 2D scans. Combining occasional large scans requires far fewer registration steps than combining frequent smaller scans. Reducing the number of registration steps reduces the accumulative registration error in the 3D map, improving long distance accuracy. An improvement in long distance accuracy is highly desirable for the mapping of underground mining environments as it is often not possible to perform six degree of freedom loop closure [13] due to the layout of the mine.

Although no online processing was completed during the collection of this dataset, the monocular vision-based SLAM algorithm has the ability to run in real time [2]. The average time required to perform registration on a new scan was 5.5 minutes when running on a laptop with a 2.67GHz processor. This is not prohibitive as the time required to perform a medium density scan is 9.8 minutes. If the registration of a scan is performed during the collection of a subsequent scan, no additional processing time is required.

V. CONCLUSION AND FUTURE WORK

A combination of bearing-only vision-based localization and the ICP-based registration of 3D laser scans has produced a surveying system capable of the production of highly accurate large scale maps. The system is designed to operate effectively without the use of external targeting systems, physical odometry, *a priori* information or motion model simplification. A modular design allows the sensor payload to be used with any available mobile platform or vehicle, greatly increasing the

potential applications. Although not fully integrated at this preliminary stage, a fully modular payload will be deployed in future work. Functionality in six degrees of freedom makes the system appropriate for the majority of indoor, outdoor and underground surveying tasks. Preliminary results obtained in a highly structured, level, man-made environment confirmed the potential of this technique and delivered a near survey quality map of a 20m x 18m x 107m structure. Future testing will take place in an extensive underground mining environment where the capabilities of the system in a less structured environment will be evaluated.

REFERENCES

- [1] Besl, P., McKay, N. 1992. A Method for Registration of 3-D Shapes. IEEE Transactions on Pattern Analysis and Machine Intelligence - Special issue on interpretation of 3-D scenes. Vol. 14 Issue 2. pp. 239-256.
- [2] Civera, J., Grasa, O., Davison, A., Montiel, J. 2010. 1-Point RANSAC for EKF Filtering: Application to Real-Time Structure from Motion and Visual Odometry. Journal of Field Robotics. Vol. 27 no. 5, pp. 609-631.
- [3] Früh, C., Zakhor, A. 2004. An Automated Method for Large-Scale, Ground-Based City Model Acquisition. International Journal of Computer Vision. Vol 60. pp. 5-24
- [4] Hähnel, D., Burgard, W., Fox, D., Thrun, S. 2003. An Efficient FastSLAM Algorithm for Generating Maps of Large-Scale Cyclic Environments from Raw Laser Range Measurements. Proceedings of the International Conference on Intelligent Robots and Systems 2003. Vol. 1, pp 206-211
- [5] Howard, A., Wolf, D., Sukhatme, G. 2004. Towards 3D Mapping in Large Urban Environments. In Proceedings of the International Conference on Intelligent Robots and Systems. Vol. 1 pp. 419-424
- [6] Le Cras, J., Paxman, J., Saracik, B., Jarosz, A. 2009. An inspection and surveying system for vertical shafts. In Proceedings Australasian Conference on Robotics and Automation 2009.
- [7] Le Cras, J., Paxman, J., Saracik, B. 2012. Improving Robustness of Vision Based Localization under Dynamic Illumination. To appear in Recent Advances in Robotics and Automation, Springer series on Studies in Computational Intelligence.
- [8] Nüchter, A., Surmann, H., Lingemann, K., Hertzberg, J., Thrun, S. 2004. 6D SLAM with an Application in Autonomous Mine Mapping. In Proceedings International Conference on Robotics and Automation. Vol. 2 pp. 1998-2003.
- [9] Nüchter, A., Lingemann, K., Hertzberg, J., Surmann, H. 2007. 6D SLAM—3D Mapping Outdoor Environments. Journal of Field Robotics. Vol 24, Issue 8-9, pp. 699-722.
- [10] Rituerto, A., Puig, L., Guerrero, J. 2010. Visual SLAM with an Omnidirectional Camera. In Proceedings International Conference on Pattern Recognition. Vol 1 pp. 348-351.
- [11] Rusu, R., Cousins, S. 2011. 3D is here: Point Cloud Library (PCL). In Proceedings of the International Conference on Robotics and Automation. pp. 1-4.
- [12] Surmann, H., Nüchter, A., Hertzberg, J. 2003. An autonomous mobile robot with a 3D laser range finder for 3D exploration and digitalization of indoor environments. Robotics and Autonomous Systems. Vol. 45 Issue 3-4. pp. 181-198.
- [13] Surmann, H., Nüchter, A., Lingemann, K., & Hertzberg, J. 2004. 6D SLAM A Preliminary Report on Closing the Loop in Six Dimensions. In Proceedings of the 5th IFAC Symposium on Intelligent Autonomous Vehicles.
- [14] Thrun, S., Burgard, W., Fox, D. 2000. A Real-Time Algorithm for Mobile Robot Mapping With Applications to Multi-Robot and 3D Mapping. In Proceedings of the International Conference on Robotics and Automation. Vol. 1 pp. 321-328