2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)
Madrid, Spain, October 1-5, 2018

# Semi-Supervised SLAM: Leveraging Low-Cost Sensors on Underground Autonomous Vehicles for Position Tracking

Adam Jacobson[1], Fan Zeng[1], David Smith[2], Nigel Boswell[2], Thierry Peynot[1] and Michael Milford[1]

***Abstract*— This work presents Semi-Supervised SLAM - a method for developing a map suitable for coarse localization within an underground environment with minimal human intervention, with system characteristics driven by real-world requirements of major mining companies. This work leverages existing information common within a mining environment - namely a surveyed mine map - which is used to sparsely ground map locations within the mine environment, increasing map accuracy and allowing localization within a global frame. Map creation utilizes a low cost camera sensor and minimal user information to produce a map which can be used for single camera localization within a mining environment. We evaluate the localization capabilities of the proposed approach in depth by performing data collection on operational underground mining vehicles within an active underground mine and by simulating occlusions common to the environment such as dust and water. The proposed system is capable of producing maps which have an average localization error 2.5 times smaller than the next best performing method ORB-SLAM2, comparable localization performance to a state-of-the-art deep learning approach (which is not a feasible solution due to both compute and training requirements) and is robust to simulated environmental obscurants.**

## I. INTRODUCTION

Deploying localization systems within underground mining environments is still currently a difficult task due to two main factors, 1) difficulty surrounding the map building process and 2) difficulty dealing with environmental change and occlusions. Within an underground environment the mapping process is difficult due to the lack of access to GPS, the poor lighting conditions and the untextured and highly aliased walls. The localization problem is challenging due to environmental change and obscurants, like water and dust, the highly aliased nature of the environment and the high cost of maintaining localization infrastructure.

Simultaneous Localization and Mapping (SLAM) [1] is the process of generating a map of an environment while estimating the position of the vehicle within the map. Generating accurate metric maps is challenging and usually requires an expensive multi-sensor sensing apparatus, including lasers, IMUs or depth cameras. Furthermore, utilization of these metric techniques does not guarantee a suitable map of the environment can be generated (Figure 1). Topological mapping systems have been shown to generate large scale maps [2] and relax the sensor requirements enabling lower cost sensors to be used. However, without a metric grounding, these techniques cannot be used for producing metric position estimates. Work presented in [3] attempts to relax the requirements of the mapping system by enabling a human intervention into the mapping process to manually remove false loop closures while the SLAM system is operating.

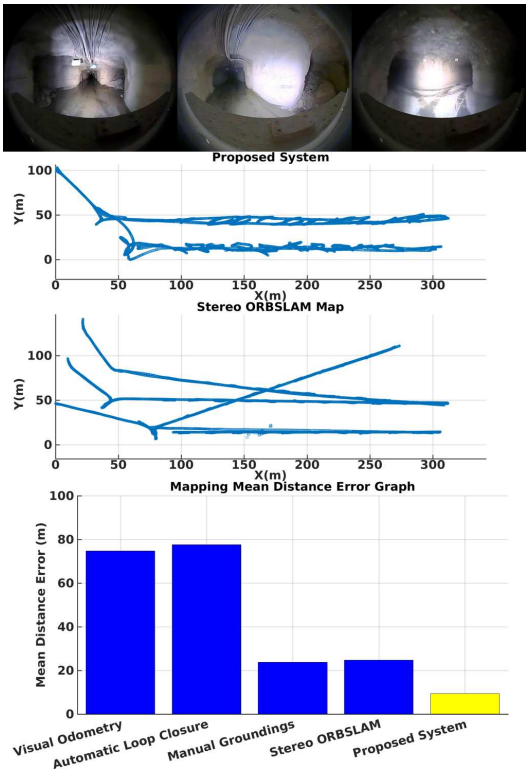


Fig. 1: (top) Illustrative example of images from the underground mining environment, (middle) maps generated with the proposed approach and Stereo ORB-SLAM2 and (bottom) mean distance error bar graph of mapping performance within the mining environment.

Visual place recognition systems for surface applications such as FABMAP [2] and SeqSLAM [4] or state-of-the-art deep learning approaches [5] have demonstrated the capability of performing place recognition despite changing environmental and seasonal conditions [4], however these systems require a manual grounding in the real world using GPS or some other technique to provide metric position estimates.

The core focus of this work is to develop a system which can provide a coarse localization estimate within an underground environment using a low cost monocular camera. The deployment of this system requires a localization system

[1]Queensland University of Technology, Australia, [2]Caterpillar, Inc.
This research was supported by an Advance Queensland Innovation Partnerships grant from the Queensland Government, Mining3, Caterpillar and the Queensland University of Technology. MM also received support from an ARC Future Fellowship FT140101229.
a1.jacobson@qut.edu.au

which can handle the vast speed changes, lighting condition changes and environmental condition changes common to the environment. Furthermore, to be widely adopted for industrial applications, such a localization system needs a convenient, robust and accurate framework for building a map of each new site.

In this work, we present Semi-Supervised SLAM, a system which allows the deployment of topo-metric SLAM maps in a real world environment enabling coarse localization. The proposed monocular SLAM approach utilizes manual map grounding to improve SLAM results and enable abstraction of the map representation. Once the map is created, monocular localization can be performed producing metric position estimates in the environment which are robust to common environmental obscurants. This work extends the prior SeqSLAM algorithm [4] to enable navigation through arbitrary paths with arbitrary environment branching and arbitrary vehicle speeds.

We evaluate our approach using real-world data captured from within an Australian underground mine, presenting three core evaluations including:

- Evaluation of the quality of the SLAM map using an external Radio Telemetry System.
- Evaluation of the utility of this SLAM map to enable monocular localization within the mine environment.
- Evaluation of the reliability of the localization system for handling occlusions and obscurants common within an underground mine by simulating water and dust at various levels.

We also present a comparison of our Semi-Supervised SLAM approach to a state-of-the-art metric SLAM system, ORB-SLAM2 [6], and compare our localization system to a state-of-the-art deep learning place recognition approach [5].

The paper proceeds as follows. In Section II, traditional SLAM and place recognition systems are discussed. Section III presents our approach describing the map generation process and the place recognition system. In Section IV, we present the experimental setup, with the results presented in Section V. Section VI discusses the outcome of the research and areas of future work.

## II. Background

In this section, we describe the state-of-the-art of SLAM and localization algorithms, highlighting systems designed to operate in hazardous conditions such as underground mine environments.

### A. SLAM

There is a large body of work demonstrating SLAM implementations using onboard robotic sensors such as lasers and cameras. Current SLAM implementations have shown the ability to operate over tremendous distances and within a variety of situations [2], [7], [8], [9], [10], [11], [12]. Systems such as ORB-SLAM2 [6] have demonstrated mapping performance with centimeter level accuracy. However, these impressive results have been achieved within surface applications, with no demonstrated results of systems operating within underground environments capable of handling the confined space, varying lighting conditions or aliased environmental conditions.

### B. Place Recognition

*1) Underground Environment Localization:* Traditional underground positioning technologies, e.g. Wireless Sensor Networks [13] and radio-frequency identification (RFID) [14] systems, work by trilateration based on measurements on time of arrival [15] or received signal strength indicator differences [16]. The underlying hardware typically consists of a mesh of beacons such as wireless local area network access points, yielding sufficient time of arrival difference to all possible localization candidates. As shown in [17], underground mine site localization performance generally increases with beacon density. Consequently there is substantial initial cost in capital and time to set up such a network, on top of which subsequent investment on construction and maintenance is continuously required with the development of the mine site. Previous vision-based mine site localization approaches have focused more on better reliability at intersections only, as opposed to accurate localization throughout a tunnel, as adopted in the hierarchical system called "Opportunistic localization" described in [23]. This approach limits the ability to prevent accidents such as collisions between vehicles and pedestrian workers.

*2) Vision-Based Localization:* The classical visual place recognition algorithm - FABMAP [2] processes bag-of-words representation of images by implementing a probabilistic approach that places less weight on commonly-seen landmarks and accentuates unusual features. This feature-based approach is contrasted with the holistic image comparison method used in SeqSLAM [18], in which the Sum of Absolute Difference (SAD) score is compared to select the local best match of an image sequence. To allow possible variations in viewing angles, the lowest SAD score among comparisons within a search window is recorded [19] in the confusion matrix. This approach works particularly well in datasets with substantial lighting and seasonal changes, thanks to steps such as patch-normalization and the incorporation of temporal relevance in sequences of images. Research presented in [20] stores multiple experiences of a place to improve place recognition. [21] tackles the problems of night-time localization by building and incorporating models of artificial light sources. A Hidden Markov Model (HMM) sequence matching strategy is used to solve the visual place recognition problem in [22]. For these systems to provide a metric position estimate, external grounding must occur via GPS or some other metric localization system.

## III. Approach

In this section, we describe our approach for performing Semi-Supervised SLAM and performing online localization using the generated map.

### A. Semi-Supervised SLAM

The Semi-Supervised SLAM system is a monocular SLAM system for creating maps of an environment with minimal human intervention. The system is comprised of 4 steps, visual odometry, loop closure, inclusion of external map groundings and map compression for localization. The only step which requires human intervention is the manual grounding phase, the other stages operate autonomously. The SLAM backend used for this work is GTSAM [24], with all the visual odometry, loop closure and external map grounding information added to a pose graph which is then optimized and then compressed for later usage for localization.

*1) Visual Odometry:* The Visual Odometry approach used within this work tracks optical flow on the ground plane to estimate the 2D motion of the vehicle. A region of the image is manually selected which lies on the ground plane, as seen in Figure 2a). Each new image is compared to the previous image by comparing the sum of absolute differences for various patch offsets. Assuming an Ackermann Vehicle model with the camera placed over the rear axle, pixel motion in the x direction corresponds to pure rotation and pixel motion in the y direction corresponds to pure translation of the vehicle, as seen in Figure 2. The rotation of the vehicle can be estimated if the distance from the rear axle and the homography of points corresponding to the ground plane is known. We calculate the rotation of the vehicle, $\theta$, using:

$$\theta = arctan\frac{x_p k_x}{y_c + y_h} \tag{1}$$

where $x_p$ is the pixel offset between the current frame and the previous frame in the $x$ direction and $k_x$ is a scaling factor derived from the homography which transforms pixel shift in the x direction into meters(meters/pixel). $y_c$ is the distance between the real axle and the camera and $y_h$ is the distance between the camera and the point at the center of the image in meters. We calculate the translation of the vehicle, $d$, using:

$$d = y_p k_y \tag{2}$$

where $y_p$ is the pixel offset between the current frame and the previous frame in the $y$ direction and $k_y$ is a scaling factor derived from the homography which transforms pixel shift in the y direction into meters(meters/pixel).

Each new frame creates a new node in the pose graph which is bound to the previous node by a constraint defined by the measured optical flow.

*2) Loop Closure:* Loop closure is an autonomous method of determining whether two sensor readings are observations of a singular place. Identification of such observations can be utilized for correcting map corruption and reducing error accumulated in odometry calculations. Loop closure is performed by comparing each new frame to each previously recorded frame. once a loop closure is detected, a constraint is added to the pose graph between the two nodes. In this work, each image is pre-processed as described in Section IV-C and then compared using:

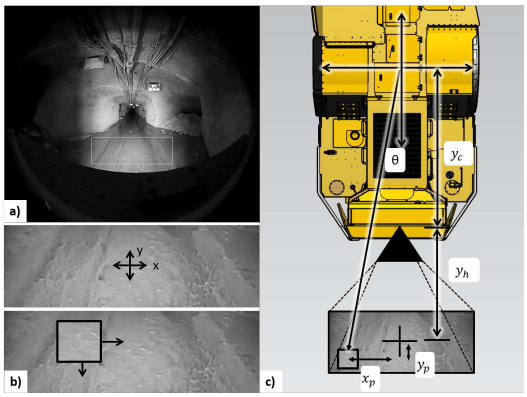


Fig. 2: a) Camera view and illustration of the selected region of ground plane for performing optical flow. b) Illustration of how patches are tracked to perform ground plane optical flow. c) Schematic of camera placement on vehicle and illustration of how translation and rotation estimates are generated.

$$d(i,j) = \frac{1}{R_x R_y} \sum |I_i - I_j(k)| \tag{3}$$

where $d(i,j)$ is the sum of absolute difference score for the current query frame $i$ and previously recorded frame $j$. $I_i$ and $I_j$ are the current query frame at timestep $i$ and node $j$, respectively. As in SeqSLAM, to account for small variations in pose, the query and reference image are laterally shifted ($\pm x_{max}$ and $\pm y_{max}$), such that the SAD score of the overlapping region is minimized [18]. $R_x$ and $R_y$ is the resolution of the images being compared. A loop closure constraint is added when $d(i,j)$ is below a threshold, $d_t$.

*3) External Map Grounding:* This approach leverages initial mine site exploration and mapping information that is required for regular mine site operation, introducing no additional requirements for deployment within a mine site environment. The external map grounding leverages existing information within a mine site to improve mapping accuracy, reducing the requirements of the visual odometry system and enabling the creation of a map which is representative of the environment within the global coordinate frame.

External grounding points are manually collected during mapping when the vehicle is at salient locations within the mine and are marked using an existing pre-surveyed map. Once the visual odometry, loop closures and external map groundings have been added to the pose graph, the graph can be optimized and compressed for localization.

*4) Stored Map for Localization:* Once the pose graph has been optimized, it is then compressed for use within our localization framework. Central to our localization approach is a graph representation of the environment, enabling the description of the connectivity of the environment. This graph representation enables our proposed system to operate in environments that contain decision boundaries which are common within underground mine environments. Each node

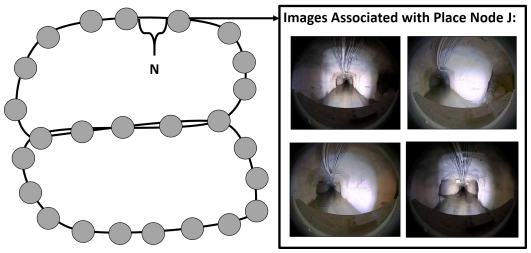


Fig. 3: Example graph representation used for localization, nodes are separated by $N$ meters and each node is associated with multiple frames to handle changes in appearance and small variations in heading.

or "unique" place in the map is separated by $N$ meters, with multiple frames associated with each node enabling the encoding of multiple view points (images) from the same position in the environment, as seen in Figure 3.

The graph structure is represented by an adjacency matrix, $A$, where $A(x, y)$ is the Euclidean distance of the shortest path between nodes $x$ and $y$. Once the map has been generated offline, the online place recognition can be performed.

### *B. Localization*

The Localization system utilizes the map constructed by the Semi-Supervised SLAM system to provide a position estimate of the vehicle while navigating through the environment. The localization system consists of three main components, image comparison, temporal filtering and the final selection of a place match.

*1) Image Comparisons:* The localization system compares input images to images stored within the map generated offline to produce a place hypothesis. The input query images are pre-processed by down-sampling the query image and performing patch normalization, as described in Section IV-C. The input query image is compared to each frame of each node, where a difference score for each node is defined as:

$$d(i,j) = min_{k \in N_j} \frac{1}{R_x R_y} \sum |I_i - I_j(k)| \tag{4}$$

where $d(i, j)$ is the minimum sum of absolute difference score for the current query frame and all the frames associated with graph node $j$. $I_i$ and $I_j(k)$ are the current query frame at timestep $i$ and the $k^{th}$ image of node $j$, respectively. $N_j$ is the number of frames stored within node $j$. As above, the query and reference image are laterally shifted (in the range of $\pm x_{max}$ and $\pm y_{max}$), such that the SAD score of the overlapping region is minimized [18]. $R_x$ and $R_y$ are the width and height of the images being compared.

*2) Temporal Filtering:* This work extends the SeqSLAM algorithm to enable navigation through arbitrary paths with arbitrary environment branching and arbitrary vehicle speeds through the use of the Hidden Markov Model [25]. The current state of the agent is $\mathbf{X_i}$ and we want to estimate the current location of the robot using:

$$\mathbf{X_i} = \mathbf{O_i} \pi \mathbf{X_{i-1}} \tag{5}$$

where $\mathbf{O_i}$ is the observation matrix, $\pi$ is the transition matrix and $\mathbf{X_{i-1}}$ is the previous state. $\mathbf{X_0}$ is initialized to the starting position of the robot which is known prior to vehicle operation. $\mathbf{X_i}$ is then normalized using:

$$\mathbf{X_i} = \frac{\mathbf{X_i}}{\sum(\mathbf{X_i})} \tag{6}$$

The Observation matrix is calculated using the image comparison results above.

$$\mathbf{O_i} = \begin{bmatrix} \gamma(d(i,0), \pi X_{i-1}(0)) \\ \vdots \\ \gamma(d(i,N_j-1), \pi X_{i-1}(N_j-1)) \end{bmatrix} \tag{7}$$

where $\gamma(k)$ is defined as:

$$\gamma(k,\alpha) = \begin{cases} \frac{d_{max}-k}{d_{max}-d_{min}}, & \text{if } \frac{d_{max}-k}{d_{max}-d_{min}} < p_t \\ \epsilon, & \text{if } \alpha < X^t_{(i-1)} \\ \epsilon, & \text{otherwise} \end{cases} \tag{8}$$

where $d_{max}$ and $d_{min}$ are the maximum and minimum prior $d$ scores respectively. We set all values less than $p_t$ to a small value $\epsilon$ as these values are deemed to be poor place matches. Furthermore, we only evaluate place matches where the prior state estimate is above threshold $X^t_{(i-1)}$ which is defined dynamically as the maximum $N_t$ estimates.

The state transition matrix utilizes the environment adjacency matrix described above and is created using a uniform distribution over a range of velocities to set the probabilities of transitioning from state to state. The transition matrix is defined as:

$$\pi = \begin{bmatrix} \beta(A(x,y)) & \dots & \beta(A(x,y)) \\ & \ddots & \\ \beta(A(x,y)) & \dots & \beta(A(x,y)) \end{bmatrix} \tag{9}$$

where $A(x, y)$ is the length of the shortest path between nodes $x$ and $y$ (as above) and $\beta(k)$ is defined as:

$$\beta(k) = \begin{cases} 1, & \text{if } k < A_t \\ 0, & \text{otherwise} \end{cases} \tag{10}$$

and $A_t$ is the maximum distance a vehicle can travel per frame. The transition matrix is normalized such that all columns sum to 1.

*3) Selecting a Place Match:* To determine if a query location matches a reference location, a search is performed for the place hypothesis with the largest score:

$$b(i) = \underset{d \in D}{\text{argmax}}(X_i(d)) \tag{11}$$

where $b(i)$ is the best place hypothesis for query location $i$ and $D$ is the set of all locations in the reference map.

The current best matching sensory snapshot is determined to be a place match if the difference score is above a global matching threshold $X_{thresh}$:

$$m = \begin{cases} 1, \text{ if } X_i(b(i)) \geq X_{thresh} \\ 0, \text{ if } X_i(b(i)) < X_{thresh} \end{cases} \tag{12}$$

It is this threshold, $X_{thresh}$, that determines if a particular location is a place match. $X_{thresh}$ is swept over a range of values to generate the Precision-Recall curves presented

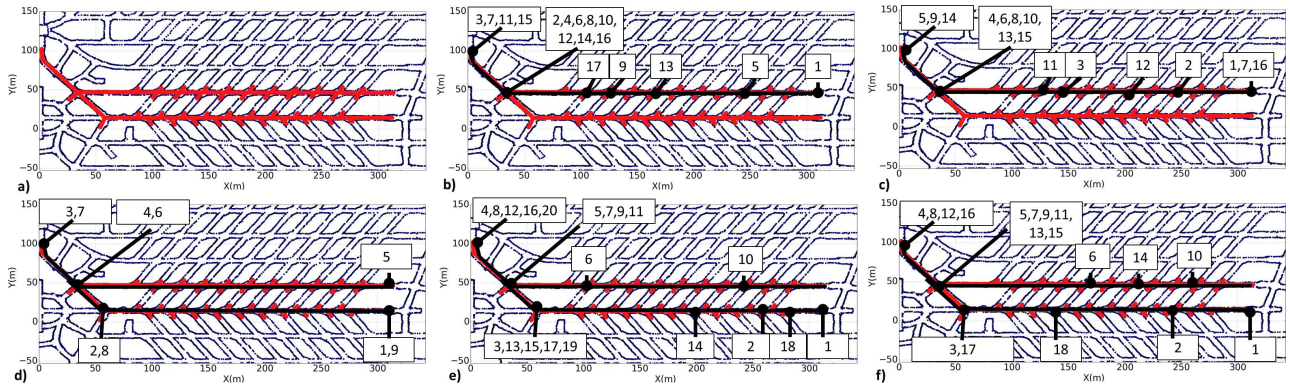


Fig. 4: a) Surveyed map of the mine environment highlighting the mapped path in red, Maps of the traveled path for the five localization trials presented (b-f), with the numbers within each map denoting the order of motion of the vehicle through the environment during the dataset.

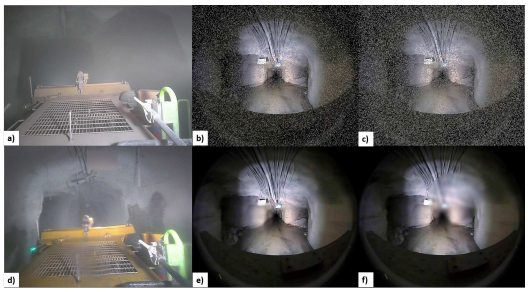

Fig. 5: Illustrations of real world dust (a) and water (d) based occlusions and examples of simulated occlusions including dust occlusions level 1 (b) and 2 (c) and water occlusions level 1 (e) and level 2 (f).

in the results. Autonomous methods for calibrating this threshold include [26], [27]. A metric position estimate is recovered by extracting the metric location of the best place match from the stored pose graph.

## IV. Experimental Setup

In this section, we discuss the experimental setup used to evaluate the proposed technique.

### A. Dataset Description

The datasets used for experimentation were captured in an Australian underground hard rock mine. The vehicle used for dataset collection was a Load Haul Dump (LHD) vehicle with sensors attached to the rear of the vehicle. The core sensor used for this experimentation is a rugged monocular camera with a 180° field of view and image resolution of 944x800. We also captured data with a MultiSense S21 Stereo camera for evaluation using ORB-SLAM2.

There are two core experimentations presented below, evaluation of the accuracy of our Semi-Supervised SLAM approach and evaluation of the utility of our mapping approach for localization within an underground environment. Data was recorded from two complete tunnels with the vehicle traveling between drawpoints and the dumping station. We captured data for the SLAM component in four parts, allowing mapping of each tunnel with both directions of travel and with data collected from digging at each drawpoint; the path of the vehicle can be seen in Figure 4a.

The data collected for evaluating the localization component consists of the LHD vehicle maneuvering through the tunnels as would be typical during normal operations, with varying speed, collecting ore from drawpoints and delivering the load to the dumping station. There are a total of five datasets collected for evaluating the performance of the localization system, these have been shown in Figure 4b-f), with the numbers within the figures denoting the traveled path of the vehicle during the dataset.

We also include a study of localization performance containing simulated obscurants that are common within the environment. Figure 5 illustrates the real world occlusions we are attempting to simulate (aerosolized dust and water occluded lenses) and our four simulations which are applied to our five localization datasets. Our four simulations included two levels of dust and two levels of water occlusions. Dust level 1 and dust level 2 in which 5% and 10% of the pixels in each frame were randomly set to a grey color to simulate aerosolized dust. Water level 1 and water level 2 consists of a static circular Gaussian blur applied to each image, with water level 1 having a diameter of 90 pixels and water level 2 having a diameter of 150 pixels.

### B. Performance Metrics

The performance metrics utilized within this work are the mean distance error metric and the Precision-Recall metric. The mean distance error metric used for this work is the mean of the distance from each point to the correct ground truth location. Precision is the number of correct place matches divided by the total number of place matches reported. Recall is the fraction of correct place matches divided by all of the recallable place matches within a dataset. The ground truth system used for this work is a radio-telemetry based system which utilizes an IMU to propagate position estimates between radio position corrections.

### C. Image Pre-Processing

Each image is pre-processed before place matching can be performed. Images are converted to grayscale, downsampled and patch normalized, as in [28], to improve lighting invariance and pose variance. Patch normalization takes each pixel

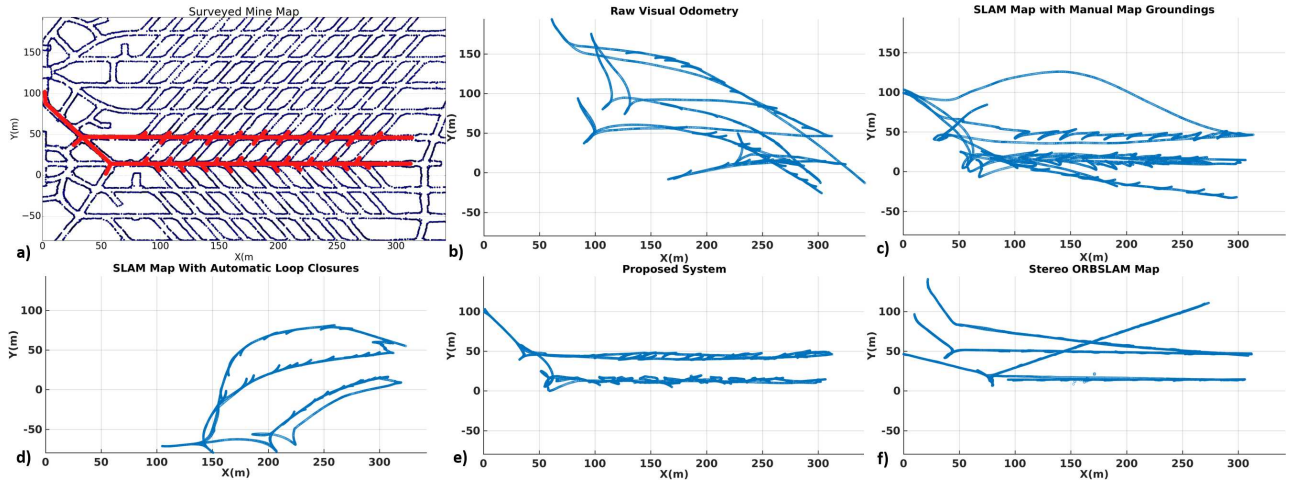


Fig. 6: a) Surveyed map of the mine environment highlighting the mapped path in red, b) Map generated using ground plane visual odometry, c) Map generated using manual map groundings, d) Map generated with automatic loop closure, e) Map generated using the proposed approach and f) a map generated using a MultiSense S21 and ORB-SLAM2. It can be seen that the proposed approach outperforms the ORB-SLAM2 map and is a faithful and topologically correct representation of the environment.

of an image and centers a patch around that pixel, then each pixel score has its patch mean subtracted and is divided by its patch standard deviation.

### D. External Map Grounding Points

The external map grounding points used for the semi-supervised mapping experimentation were captured at the beginning and end of every tunnel and while the vehicle is stationary at every drawpoint. These points were tagged using a DXF map of the mine when the vehicle was at salient points within the mine structure.

## V. Results

We present the evaluation of our proposed SLAM system and the performance of our localization system built using the SLAM map, along with an evaluation of the localization performance with simulated environmental occlusions.

### A. SLAM Map Evaluation

Figure 6 shows (a) the pre-surveyed mine map illustrating the two tunnels being mapped, along with the results from the ground plane visual odometry map (b), map generated using manual map groundings (c) and automatic loop closures (d). Figure 7 presents a graph of the mean distance error of each approach. Individually, it can be seen that these three maps are poor estimates of the environment. These methods are combined within the proposed approach, seen in Figure 6e), to generate an accurate map of the environment, with a mean distance error of 9.44m (Table I). The results of generating a map of the environment using a MultiSense S21 and ORB-SLAM2 can be seen in Figure 6f), this map suffers from poor loop closures and scale error, producing a map with a mean distance error of 24.73m (Table I). Table I shows the mean distance error in meters for each mapping result, with the most accurate mapping result being produced by the proposed Semi-Supervised SLAM system, having an

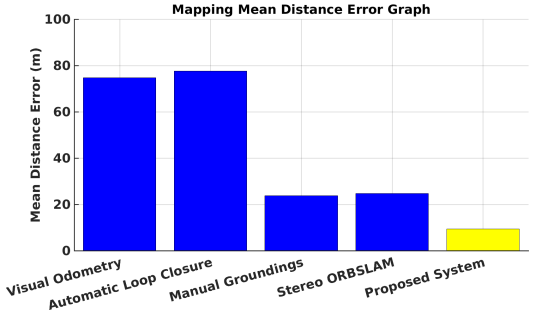


Fig. 7: Mean distance error graph for the five different mapping approaches. The proposed approach outperforms the four other techniques and the state-of-the-art stereo ORB-SLAM2 approach.

average localization error 2.5 times smaller than the next best performing method ORB-SLAM2.

TABLE I: Evaluation of Mean Distance Error for the Mapping Techniques

| Mapping Technique | Mean Error |
|---|---|
| Stereo ORBSLAM | 24.73m |
| Visual Odometry | 74.67m |
| Automatic Loop Closure | 77.57m |
| Manual Groundings | 23.76m |
| **Proposed System** | **9.44m** |

### B. Localization Performance

Figure 9 presents the Precision-Recall results for the localization experiment using our proposed approach and a state-of-the-art deep learning approach [5] utilizing the conv3 layer of a pre-trained AlexNet ConvNet [29] for place recognition comparison. The map used for both our proposed localization approach and the deep learning approach is the Semi-Supervised SLAM map generated above. The mean

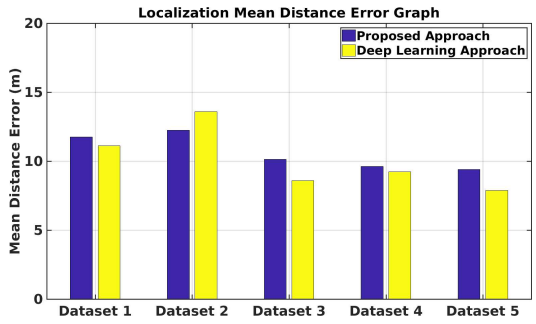


Fig. 8: Mean distance error graph for the five different localization trials. The proposed approach has comparable performance to the state-of-the-art deep learning approach, outperforming the deep learning approach in dataset 2.

error for the SLAM map is 9.44m, for evaluating the localization systems we used a ground truth threshold of 30m from the ground truth position to calculate Precision and Recall.

The results shown in Figure 9(a-e) demonstrate that the performance of the deep learning approach and our proposed approach are comparable in all localization experiments with our proposed approach outperforming the deep learning approach in localization trial 2 and 5 (Figure 9(b,e)).

We also present mean distance error results in Figure 8, demonstrating comparable performance in all five experiments with the proposed approach outperforming the deep learning approach in trial 2.

### C. Localization Performance with Occlusions

The results to the extended localization trial featuring simulated environmental occlusion can be seen in Figure 9(f-j). The results illustrate that, across all trials, the proposed approach can successfully localize with the dust level 1, dust level 2 and water level 1 occlusions, with water level 2 occlusions causing the system to not function reliably. The results show that environmental dust occlusions is not a significant limitation for the proposed system, however water accumulation causing occlusions on the lens will need to be monitored to ensure operation of the system in real-world conditions. This characterization is driven directly by industry requirements and provides them with a guide to what active cleaning requirements would be required for camera sensors and under what conditions.

Furthermore, it can be seen in all trials that the proposed approach outperforms the deep learning approach with dust level 1 and dust level 2, indicating that the deep learning approach would not be suitable within a dusty mining environment. It can also be seen that the water occlusions do not significantly effect the deep learning approach, maintaining similar performance to the original localization trials.

## VI. Discussion and Future Work

The results presented above illustrate that the proposed Semi-Supervised SLAM system can generate an accurate map of an environment where current state-of-the-art SLAM systems fail to accurately represent the world and create accurate loop closures. The Semi-Supervised SLAM system utilizes a monocular camera and information which is present within a mining environment a priori to generate maps which are more than 2.5 times more accurate than the map generated using state-of-the-art approaches.

The map generated using the Semi-Supervised SLAM approach enables accurate position estimation within an underground environment. Furthermore, the proposed localization approach is demonstrated to be robust to the presence of simulated dust. The proposed approach was capable of producing comparable localization performance to a state-of-the-art deep learning place recognition system (a system which is not a feasible solution due to both compute and training requirements).

Future work will investigate mapping and localizing in a larger tunnel network, integrating multiple sensing modalities to improve mapping and localization performance. Furthermore, future work will integrate multiple sensors such as lasers and radar into the HMM framework to improve positioning accuracy and precision.

## References

[1] G. Dissanayake, P. M. Newman, S. Clark, H. Durrant-Whyte, and M. Csorba, "A solution to the simultaneous localisation and map building (SLAM) problem," *IEEE Transactions on Robotics and Automation*, vol. 17, no. 3, pp. 229–241, 2001.

[2] M. Cummins and P. Newman, "Highly scalable appearance-only SLAM - FAB-MAP 2.0," in *Robotics: Science and Systems*, Seattle, United States, 2009.

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

[4] M. Milford and G. Wyeth, "SeqSLAM: Visual Route-Based Navigation for Sunny Summer Days and Stormy Winter Nights," in *Proc. - IEEE Int. Conf. Robot. Autom.*, St Paul, United States, 2012.

[5] N. Sünderhauf, S. Shirazi, F. Dayoub, B. Upcroft, and M. Milford, "On the performance of convnet features for place recognition," in *Intelligent Robots and Systems (IROS), 2015 IEEE/RSJ International Conference on*. IEEE, 2015, pp. 4297–4304.

[6] R. Mur-Artal and J. D. Tardós, "ORB-SLAM2: an open-source SLAM system for monocular, stereo and RGB-D cameras," *IEEE Transactions on Robotics*, vol. 33, no. 5, pp. 1255–1262, 2017.

[7] A. J. Davison, I. D. Reid, N. D. Molton, and O. Stasse, "Monoslam: Real-time single camera slam," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2007.

[8] H. Andreasson, T. Duckett, and A. Lilienthal, "A minimalistic approach to appearance-based visual slam," *IEEE Transactions on Robotics*, 2008.

[9] L. M. Paz, P. Pinies, J. D. Tardos, and J. Neira, "Large-scale 6-dof slam with stereo-in-hand," *IEEE Transactions on Robotics*, 2008.

[10] K. Konolige and M. Agrawal, "Frameslam: From bundle adjustment to real-time visual mapping," *IEEE Transactions on Robotics*, 2008.

[11] S. Thrun and M. Montemerlo, "The graphslam algorithm with applications to large-scale mapping of urban structures," *The International Journal of Robotics Research*, 2006.

[12] A. Kawewong, N. Tongprasit, S. Tangruamsub, and O. Hasegawa, "Online and incremental appearance-based slam in highly dynamic environments," *The International Journal of Robotics Research*, 2010.

[13] M. A. Moridi, Y. Kawamura, M. Sharifzadeh, E. K. Chanda, M. Wagner, H. Jang, and H. Okawa, "Development of underground mine monitoring and communication system integrated ZigBee and GIS," *Int. J. Min. Sci. Technol.*, vol. 25, no. 5, pp. 811–818, 2015.

[14] D. Kent, "Digital networks and applications in underground coal mines," *11th Undergr. Coal Oper. Conf.*, pp. 181–188, 2011.

[15] T. Sathyan, D. Humphrey, and M. Hedley, "WASP: A system and algorithms for accurate radio localization using low-cost hardware," *IEEE Trans. Syst. Man Cybern. Part C Appl. Rev.*, vol. 41, no. 2, pp. 211–222, 2011.

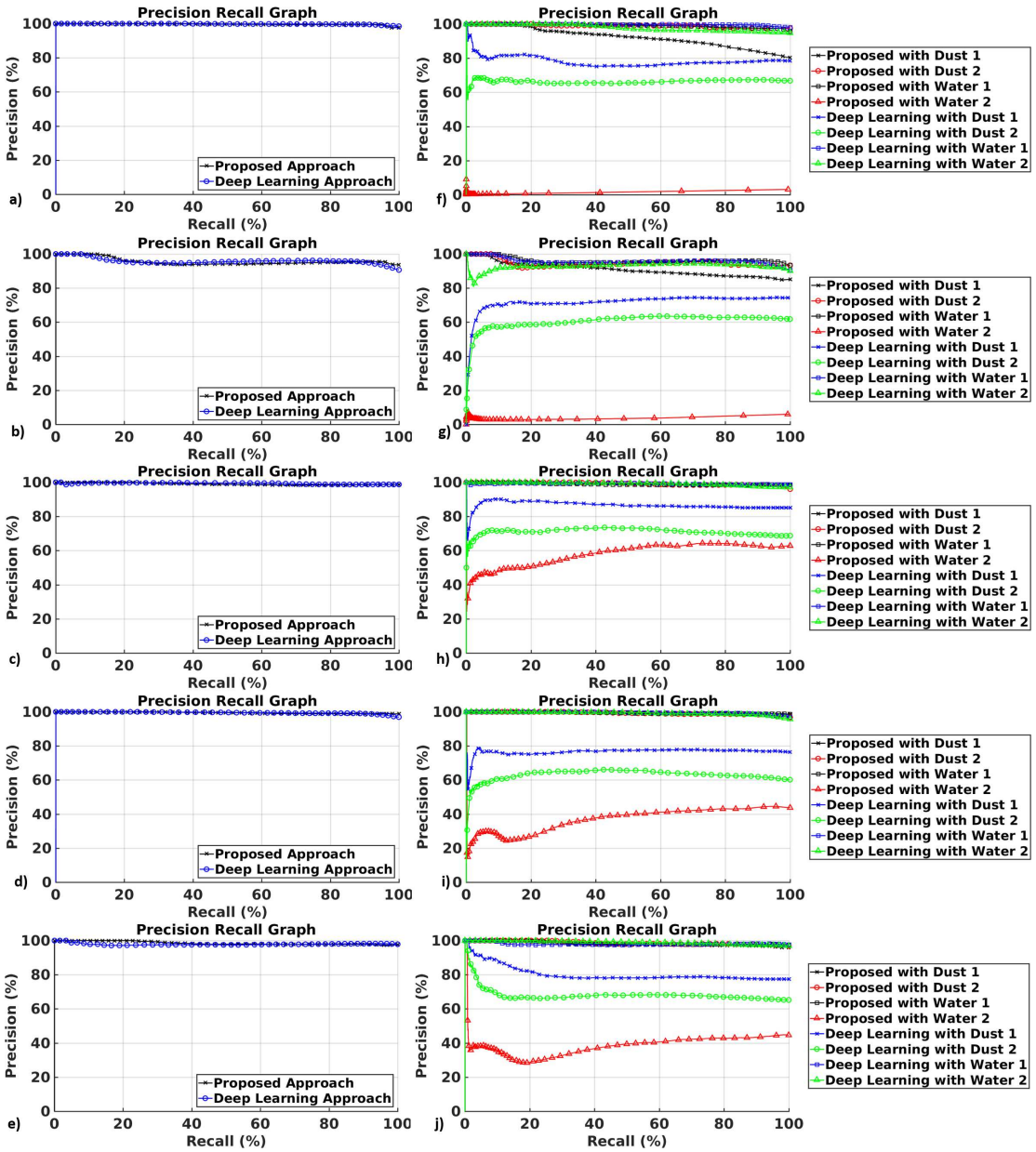


Fig. 9: Precision recall of localization datasets 1-5. The graphs in the left column present the results of the proposed approach and the deep learning approach (a-e). The graphs in the right column present the results of the proposed approach and the deep learning approach during the occlusion study with varying levels of simulated dust and water occlusions (f-j).


[16] C. Rohrig and M. Muller, "Localization of Sensor Nodes in a Wireless Sensor Network Using the nanoLOC TRX Transceiver," *VTC Spring 2009 - IEEE 69th Veh. Technol. Conf.*, pp. 1–5, 2009.

[17] N. J. Lavigne and J. A. Marshall, "A landmark-bounded method for large-scale underground mine mapping," *J. F. Robot.*, vol. 29, no. 6, pp. 861–879, 2012.

[18] M. Milford, "Visual Route Recognition with a Handful of Bits," in *Robotics: Science and Systems VIII*. Sydney, Australia: MIT, 2012.

[19] M. Milford and A. Jacobson, "Brain-inspired Sensor Fusion for Navigating Robots," in *Proc. - IEEE Int. Conf. Robot. Autom.*, Karlsruhe, Germany, 2013.

[20] W. Churchill and P. Newman, "Experience-based navigation for long-term localisation," *The International Journal of Robotics Research*, vol. 32, no. 14, pp. 1645–1661, 2013.

[21] P. Nelson, W. Churchill, I. Posner, and P. Newman, "From dusk till dawn: Localisation at night using artificial light sources," *Proc. - IEEE Int. Conf. Robot. Autom.*, pp. 5245–5252, 2015.

[22] P. Hansen and B. Browning, "Visual place recognition using HMM sequence matching," in *IEEE Int. Conf. Intell. Robot. Syst.*, 2014.

[23] E. Duff, J. Roberts, and P. Corke, "Automation of an underground mining vehicle using reactive navigation and opportunistic localization," *IEEE Int. Conf. Intell. Robot. Syst.*, vol. 4, pp. 3775–3780, 2003.

[24] F. Dellaert *et al.*, "Gtsam," *URL: https://borg. cc. gatech. edu*, 2012.

[25] L. R. Rabiner, "A tutorial on hidden markov models and selected applications in speech recognition," *Proceedings of the IEEE*, vol. 77, no. 2, pp. 257–286, 1989.

[26] A. Jacobson, Z. Chen, and M. Milford, "Online place recognition calibration for out-of-the-box SLAM," in *IEEE Int. Conf. Intell. Robot. Syst.*, 2015, pp. 1357–1364.

[27] ——, "Autonomous Multisensor Calibration and Closed-loop Fusion for SLAM," *J. F. Robot.*, vol. 32, no. 1, pp. 85–122, 2015.

[28] D. Ball, S. Heath, J. Wiles, G. Wyeth, P. Corke, and M. Milford, "OpenRatSLAM: an open source brain-based SLAM system," *Autonomous Robots*, pp. 1–28, 2013.

[29] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in neural information processing systems*, 2012, pp. 1097–1105.