Challenges of Benchmarking SLAM Performance for Construction Specific Applications

Sebastian A. Kay[1,2], Simon Julier[1,2] and Vijay M.Pawar[1,2]

***Abstract*— We have reached the stage where the feasibility of using SLAM systems are being tested to solve real-world applications. These scenarios range from the movement of materials in factories, to search and rescue tasks in unknown environments. Each application has unique requirements and we believe understanding these constraints is critical when assessing performance and guiding subsequent development. For example, in some applications, semantic classification of some parts of a map might be most important, whereas in others the need to minimise geometric error is paramount. However, prior approaches to SLAM benchmarking ignore these individual differences leading to solutions that only partially addresses end user needs; focusing on system level performance or metrics such as RMS, neither of which give an indication on how well a robot could complete an application specific goal. By introducing these application metrics within the context of SLAM benchmark tools, we intend to develop a framework that aids the adoption of SLAM solutions within real-world applications. As such, in this paper we consider the problem of developing SLAM benchmark tools within the context of the construction industry. By evaluating the differences between existing SLAM solutions against specific application constraints we propose a set of metrics and evaluation methodologies that aim to assist the development of industry-ready SLAM systems.**

I. INTRODUCTION

SLAM systems are reaching a level of maturity where they are being deployed on autonomous vehicles in unstructured environments to facilitate the completion of specific tasks to aid operational productivity and decision making. Within the construction industry there has recently been a push towards automation with brick laying robots, mobile welding robots, and additive extrusion systems being deployed in outdoor environments [1] [2] [3]. With further development of autonomous construction systems we believe the primary limiting factor will be the robot's ability to localize objects in large, highly dynamic environments to aid subsequent understanding and interaction. In this paper we will present what we believe are the requirements for SLAM systems when deployed within the construction industry.

Existing SLAM bechmarking frameworks tend to focus on either system level metrics (memory, compute speed, etc) or mapping and tracking quality (Reconstruction RMSE and odometry drift). Different SLAM algorithms can be compared on common datasets where individual algorithms can be placed on leader boards for easy comparison [4] [5] and [6]. The metrics tracked by these benchmarking systems do not however give an indication of how well an algorithm would perform given a specific task. Within construction applications the primary needs are: navigation within a cluttered environment that changes over time, differentiating between site clutter and the intended construction, comparison to project plans (for instruction and fault detection), and robust performance over large operating volumes and time periods. For mapping specific tasks the most important factor is completeness of the map, missing information is unacceptable for determining build quality, project progress, and required actions. [7].

In addition to assisting construction activities, SLAM systems can allow for automation of surveying. Throughout a construction project a site is monitored to keep track of changes in the local environment and ongoing construction. Site surveying tends to be done using GPS, LIDAR, fixed marker systems, and image based techniques (like structure from motion [8]) [7].

The requirements of surveying varies between projects and ranges from capturing 3D reconstructions, measuring seasonal changes of surrounding environment, and measure compression of a structure as it is further developed. Tolerances required for surveying tend to be +/-3mm for image based surveys [9]. On-site image based surveys tend not to be done using SLAM systems but are post-processed using structure from motion techniques in combination with known marker locations to achieve the required geometric accuracy across the full site. With this context, completeness of a survey is paramount, missing data is unacceptable and results in return trips by the survey team - making the initial survey useless leading to additional costs.

In this paper we will discuss construction specific challenges which SLAM systems need to overcome before becoming widely adopted for autonomous construction, as well as discuss a heuristic which focuses on completeness of a reconstruction penalizing missing data. The work of this paper is a precursor to a larger investigation into evaluating SLAM systems for construction applications which will be carried out in the near future.

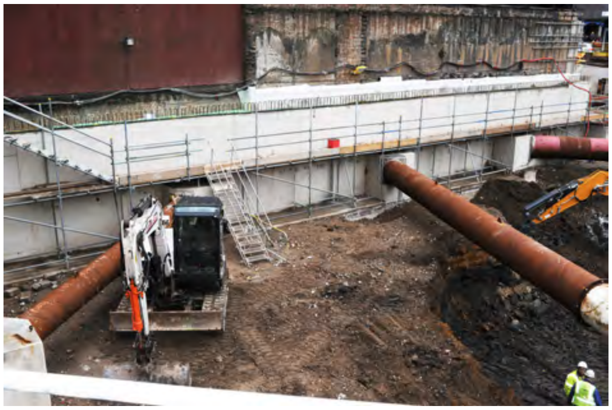
II. MOTIVATING PROBLEM

To enable automation of construction specific tasks we need to be able to evaluate how appropriate SLAM systems would be for any given challenge. We believe application specific benchmarking is needed to guide development and inform future investigations of *in the wild* deployment. Datasets are common within image processing tasks for object detection and segmentation [10], as task complexity increases dataset generation becomes more challenging.

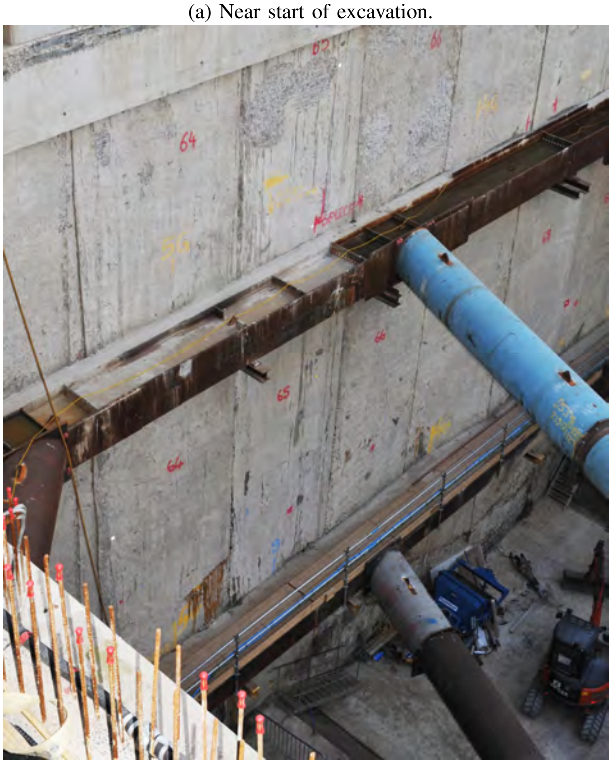
The presented work was supported by the Engineering and Physical Sciences Research Council (EP/N018494/1).

[1] Faculty of Engineering at University College London, UK

[2] UCL Robotics Institute, UCL, London, UK

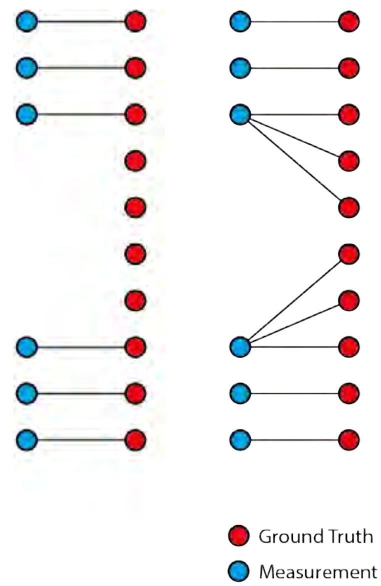


(a) Near start of excavation.



(b) 6 months into excavation project.

Fig. 1: 2 images from a construction project in central London. 1a shows site near beginning of project and 1b shows towards end of excavation.



(a) d_i when measuring R.

(b) s_i when measuring T.

Fig. 2: A graphical representation of process described by equations 1 and 2.

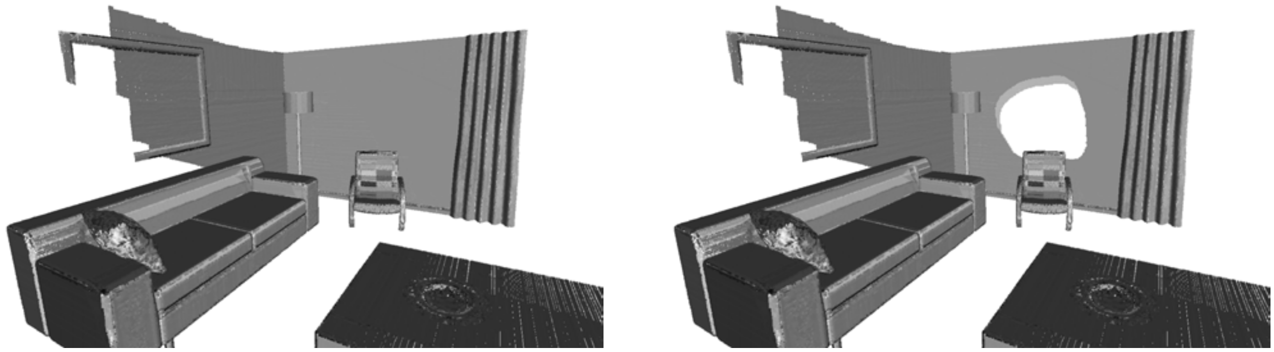
Application specific benchmarks have been produced to help determine hardware and software suitability for specific tasks, the YCB Object and Model Set is used for robotic grippers and grasping [11], MAVBench is used to evaluate various applications of mobile aerial vehicles [12]. These two example application focused benchmark tools present a set of task specific challenges for other research groups to attempt allowing for comparisons between implementations on a common problem.

The primary features required for a construction based SLAM system are: map completeness, track changes, operate over extended periods of time and in changing light conditions, separate between structure and clutter/potential hazards, and assurance of geometrical accuracy [7]. A common heuristic for map accuracy is the reconstruction error (RER) which is the mean distances from each point to the nearest surface in the ground truth 3D model [13]. The issue with using RER to measure the quality of reconstruction is that missing data will not impact benchmarking scores - but an incomplete map with some regions closely aligned with the ground truth is still an incomplete map.

To illustrate these requirements, we will present images from a deep excavation project within central London (an excavation carried out by Bouygues UK near Warren Street Station). The purpose for surveying in this case was to ensure that the walls were not deforming beyond acceptable margins - there was an expected displacement on the retaining wall of approximately 350mm over the course of excavation which took place over a 6 month period. Figure 1 shows the same area of the construction site near the start and end of the excavation work. Figure 1 highlights some of the challenges associated within construction. In Figure 1b the retaining wall is clearly visible, during the course of the project these shift slightly causing the wall to deform. This deformation is particularly pronounced at the corners but occurs along the length of the wall.

Figure 1a shows obstruction of the back wall due to construction machinery being in the way, reconstructions from this dataset would need to differentiate between the vehicle and the rear wall and support beams. Construction projects tend to be outdoors and hence subject to severe changes in lighting conditions.

Tolerances required within the construction vary but tend to require site wide coherency in the order of a few mm, which is below results posted by ElasticFusion [13] using ICL-NUIM dataset [14] which achieved 7-28mm reconstruction error. Requiring an accuracy of below published SLAM



(a) (b)

Fig. 3: Test dataset from ICL-NUIM [5], a subset of the ‘Livingroom’ dataset was used. Figure 3a is not missing any information when compared to the ground truth, but Figure 3b has a missing section in the back wall.

map accuracies does not necessarily prohibit the use of SLAM for these applications but encourages post survey processing without the real time requirement imposed on SLAM systems.

III. METHODOLOGY

We will define a set of situations which a SLAM system would encounter during construction based activities which are required for autonomous operation. These tasks are defined in the following list:

- Initial Scan, capturing full site geometry.
- Navigate in environment to complete task (such as deliver material).
- Take updated scan in modified environment, differentiating between site clutter and incomplete structure.
- Evaluate collected data to determine any problems with the project (scheduling, damaged masonry, etc.).

When benchmarking with a focus on reconstruction completeness we need a heuristic which penalizes missing data from a map. Holes are likely to occur around areas of low texture which are common in construction. When monitoring a construction site a missing wall is a large issue, as is misdiagnosing a missing wall. Missing information can lead to poor estimates on level of completion, allow for poor navigation, and lead to a misallocation of resources.

We define the set of reconstructed points as $\mathcal{M}_p$ and the set containing ground truth geometry as $\mathcal{M}_g$. For each point $p_i \in \mathcal{M}_p$ we define a distance d_i which is the separation from p_i to its nearest point in $\mathcal{M}_g$. From this we can define the reconstruction error as:

$$R = \frac{1}{N} \sum_{i=1}^{N} \sqrt{d_i^2} \tag{1}$$

Where N is the number of points in $\mathcal{M}_p$. If we also take the mean distance from each point in $\mathcal{M}_g$, q_i and find the distance to its nearest neighbour in $\mathcal{M}_p$, s_i rejecting any point pair which has a separation greater than R we obtain another measurement for reconstruction accuracy. Here M is the number of points in s.

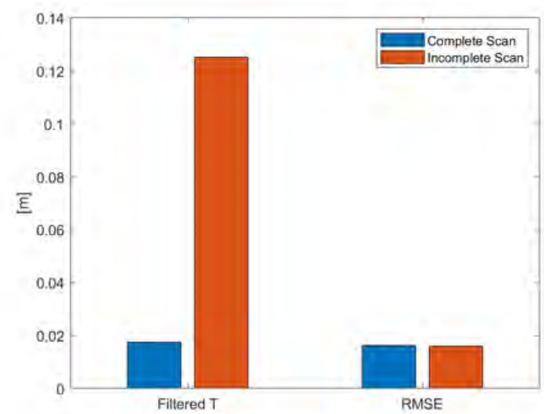


Fig. 4: Comparison between RMSE and Filtered T for two datasets one with missing data and one without missing data.

$$T = \frac{1}{M} \sum_{i=1}^{M} \sqrt{s_i^2} \tag{2}$$

Comparing R with T gives us a measurement of map completeness and presence of holes in the reconstruction.

We intend to evaluate various SLAM systems using the metric T defined in Equation 2 to and compare performance. One caveat of T is any evaluated algorithm will be penalized if the ground truth data incorporates unavailable information for reconstruction, however this can be overcome by careful preparation of used datasets.

Figure 3a and 3b show two sets of data from ICL-NUIM dataset [5] but one has a section of missing data, both datasets have had Gaussian noise $\sigma = 0.02m$ applied to each point. Figure 4 shows that the two datasets score nearly identically for RMSE, 0.0161m for both datasets when compared to the ground truth. However the filtered T value shows a large discrepancy being 0.0175 for the complete dataset and 0.125 for the incomplete point cloud.

The difference in RMSE values and filtered T Value can give a much clearer indication of whether sections of the reconstruction are missing.

IV. CONCLUSIONS

In this paper we have made a case for application specific benchmarking with specific reference to SLAM within the construction industry. Currently results are yet to be fully evaluated but it is our intention to collect and curate a dataset of construction environments which incorporate challenges discussed earlier in this document such as obscured structure, changing geometry, and varying environmental conditions. We hope the work laid out in this document, and our future work will be of benefit to the SLAM and construction communities.

REFERENCES

[1] S. J. Keating, J. C. Leland, L. Cai, and N. Oxman, "Toward site-specific and self-sufficient robotic fabrication on architectural scales," *Science Robotics*, 2017.

[2] N. Kumar, N. Hack, A. Kathrin, Doerfler Walzer, G. Rey, F. Gramazio, M. Kohler, and J. Buchli, "Design, Development and Experimental Assessment of a Robotic End-effector for Non-standard Concrete Applications," *IEEE International Conference on Robotics and Automation (ICRA)*, no. June, pp. 1707–1713, 2017.

[3] G. Pritschow, M. Dalacker, J. Kurz, and M. Gaenssle, "Technological aspects in the development of a mobile bricklaying robot," *Automation in Construction*, vol. 5, no. 1, pp. 3–13, 1996.

[4] A. Geiger, "Are we ready for autonomous driving? the kitti vision benchmark suite," in *Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, CVPR '12, (Washington, DC, USA), pp. 3354–3361, IEEE Computer Society, 2012.

[5] A. Handa, T. Whelan, J. McDonald, and A. Davison, "A benchmark for RGB-D visual odometry, 3D reconstruction and SLAM," in *IEEE Intl. Conf. on Robotics and Automation, ICRA*, (Hong Kong, China), May 2014.

[6] J. Sturm, N. Engelhard, F. Endres, W. Burgard, and D. Cremers, "A benchmark for the evaluation of rgb-d slam systems," in *Proc. of the International Conference on Intelligent Robot Systems (IROS)*, Oct. 2012.

[7] *BS 5606: 1990*. 1990.

[8] J. L. Schonberger and J.-M. Frahm, "Structure-from-Motion Revisited," *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 4104–4113, 2016.

[9] D. Andrews, J. Bedford, and P. Bryan, *Metric Survey Specifications for Cultural Heritage*. 2015.

[10] L. Van Gool, M. Everingham, A. Zisserman, J. Winn, and C. K. I. Williams, "The Pascal Visual Object Classes (VOC) Challenge," *International Journal of Computer Vision*, vol. 88, no. 2, pp. 303–338, 2009.

[11] B. Çalli, A. Walsman, A. Singh, S. Srinivasa, P. Abbeel, and A. M. Dollar, "Benchmarking in manipulation research: The YCB object and model set and benchmarking protocols," *CoRR*, vol. abs/1502.03143, 2015.

[12] B. Boroujerdian, H. Genc, S. Krishnan, W. Cui, A. Faust, and V. Reddi, "MAVBench: Micro aerial vehicle benchmarking," *Proceedings of the Annual International Symposium on Microarchitecture, MICRO*, vol. 2018-October, pp. 894–907, 2018.

[13] T. Whelan, S. Leutenegger, R. Salas Moreno, B. Glocker, and A. Davison, "ElasticFusion: Dense SLAM Without A Pose Graph," *Robotics: Science and Systems XI*, 2015.

[14] A. Handa, T. Whelan, J. McDonald, and A. J. Davison, "A benchmark for RGB-D visual odometry, 3D reconstruction and SLAM," *Proceedings - IEEE International Conference on Robotics and Automation*, pp. 1524–1531, 2014.