Statement of Purpose

Akash Sharma Ph.D. in Robotics
UserID: akashsharma@cmu.edu

As a master's student at *the Robotics Institute (RI)* CMU, three main reasons guide my decision to pursue a Ph.D. in Robotics: (1) a need to continue research in *Simultaneous Localization and Mapping* (SLAM) to enable spatial understanding and ambiguity awareness in robots, (2) an interest to teach and popularize robotics in general and SLAM in specific, and (3) a desire to continue collaboration with some of the smartest people I have met.

I believe SLAM is one of the most fundamental problems in robotics. Being the first of the `sense-plan-act` robot control loop, an accurate and semantically interpretable model of the world is imperative for every downstream task required by a robot. In recent years many SLAM systems have started to mature, but only in terms of trajectory accuracy [2, 3] and reconstruction fidelity [4]. These traditional approaches typically provide only a geometric reconstruction in the form of a point cloud or an occupancy grid, which is generally limiting for downstream tasks. As roboticists, if we mean to build a robot capable of household chores such as folding clothes or even simple object retrieval, we need the robot to spatially position objects as well as recognize them in its world model. Therefore, there is a need to consider a broader viewpoint of SLAM that includes not only robot localization and *geometric* mapping but also *object* tracking and *semantic* mapping. In my Ph.D., I want to work towards solving this unified SLAM viewpoint.

One potential strategy towards a solution is the following multi-pronged attack: (1) Incorporate semantics in the loop to create meaningful maps [1], (2) relax the assumption of static landmarks to a problem formulation that handles multiple landmark tracking and robust robot pose estimation and (3) relax the optimization backend to one that considers ambiguity in environments [5, 6].

During my master's at *the Robotics Institute (RI)* CMU, advised by Prof. Michael Kaess, I made some headway towards solving the first part of the problem – incorporating semantics in SLAM. I developed an online SLAM system on GPU that reconstructs a scene as a pose graph of 3D objects, each represented as a Truncated Signed Distance Function (TSDF) volume grid. This representation replaces typically expensive drift correction operations [7] with a lightweight pose graph optimization and drastically reduces both memory usage and computational cost. Our work, therefore, can be treated as an amalgamation of sparse and dense SLAM methods. Further, in this work, I developed a novel rendering method to compose a scene from the background and the viewed objects to improve *frame-to-model* [4] tracking. This work [1] has been submitted to the *International Conference on Robotics and Automation (ICRA) 2021* and is under review. Admittedly, while our semantic SLAM system makes progress towards higher level maps, unlike humans, it does not use strong *object priors/models* to complete partially viewed objects. Model completion is relevant because it can, in turn, enable robust tracking. Recent progress in deep implicit representations [8, 9] may help in model completion and map compression. In this regard, I intend to explore novel mapping representations during my Ph.D.

The other two aspects of the problem are interrelated. If we relax the static landmarks assumption to achieve multi-object tracking, we automatically introduce multiple modes in the solution and lose convexity in the *Maximum a Posteriori (MAP)* formulation. Consider a pathological case with two beacons and a robot: given a measurement, when the robot moves, if the beacon also moves, there are an infinite number of solutions that can satisfy this ambiguous constraint. Consequently, in the frontend, we face the data-association problem – robustly estimating correspondences between moving landmarks – and in the backend, we face a non-convex optimization problem – multiple plausible hypotheses can explain the measurements obtained. In the past, researchers solved data association via the Hungarian algorithm or assumed it as known. However, those approaches made proactive decisions and did not consider cyclic consistency across measurements. The ability to correct previous associations in light of new information is critical to handle ambiguity. Incidentally, this problem of data association finds many parallels in multi-object tracking from the computer vision community. To benefit from these synergies, I also collaborate with Prof. Katerina Fragkiadaki from the Machine Learning Department at CMU. In ongoing work, I have re-implemented state-of-the-art deep learning-based methods for data association and am currently exploring dynamic object tracking and deep representations for 3D mapping. On the other end, some progress towards multi-hypotheses solvers in SLAM [5] has already been made in our (Prof. Michael Kaess') research group, but this solution has to track hypotheses explicitly, making it computationally expensive. In my Ph.D., I am interested in extending this solution with non-parametric inference methods to handle ambiguity.

In my short experience in research, I have observed that many novel ideas become irrelevant in the absence of *accessible* implementations. Consequently, I think that in computer science or robotics, a usable codebase and a well-implemented library is often a more valuable contribution than the idea itself. Aspiring to this goal, I have strived for easy-to-use open-source code. For instance, I implemented *spatial hashing* and *submap based registration* for the OpenCV library and contributed to a few improvements in the Open3D library. Further, I am currently working on re-designing and open-sourcing my semantic SLAM pipeline as well as the GPU based 3D reconstruction library our research group internally uses [10]. I believe that my experience as a software engineer for two years before my master's shapes my perspective to equally prioritize research ideas as well as their implementation, reproducibility, and usability.

I enjoy robotics not only from a research standpoint but also from a pedagogical perspective. I love talking about robotics, and I enjoy breaking down complex concepts to their first principles to make the subject matter understandable. In my undergrad, I volunteered to teach at a robotics workshop that drew over 200 freshers each academic year and taught introductory topics such as microcontroller programming and sensor-effector control basics. More recently, during my master's, as a teaching assistant for Robot Localization and Mapping, I delivered a lecture on dense SLAM methods and re-designed a few homeworks to make the course more enjoyable.

Robotics drew my interest rather typically. In my undergrad, I started by participating in the inter-college line-follower robot competition and did well. After that, I took relevant courses such as Control Systems, Linear Algebra, and Image Processing. To obtain a deeper understanding of the subjects, I implemented a few robotics-related projects focusing on state estimation and control. Subsequently, when I had the opportunity to join the Robotics Institute (RI) CMU, there was no looking back! Research is a collaborative task and relies on a strong peer group. Over the past year, I was fortunate to have collaborated and worked with experts and some of the smartest people I know. I am humbled, and I have learned a lot. A huge motivation to join the RI Ph.D. program at CMU is due to a yearning to continue collaboration with them.

Finally, considering the research progress I have made in the past year, my belief in implementation focused research, and my interest in teaching, I gather that I can contribute to and uphold the CMU standard. A Ph.D. from *the Robotics Institute (RI)* CMU will provide me with rigorous research experience, access to exceptional faculty and peers, and plentiful resources I seek. Given the opportunity, I intend to pursue research towards a unified SLAM solution to make household robots feasible. I am interested in continuing my association with Prof. Michael Kaess and Prof. Katerina Fragkiadaki at CMU, and I also envision collaborating with Prof. Sebastian Scherer for real-world applications of state estimation.

I look forward to attending Carnegie Mellon as a Ph.D. student!

References

[1] A. Sharma, W. Dong, and M. Kaess, "Compositional Scalable Object SLAM," *arXiv:2011.02658 [cs]*, Nov. 2020.

[2] T. Whelan, S. Leutenegger, R. S. Moreno, B. Glocker, and A. Davison, "ElasticFusion: Dense SLAM Without A Pose Graph," in *Robotics: Science and Systems XI*, vol. 11, July 2015.

[3] 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, pp. 1255–1262, Oct. 2017.

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

[5] M. Hsiao and M. Kaess, "MH-iSAM2: Multi-hypothesis iSAM using Bayes Tree and Hypo-tree," in *2019 International Conference on Robotics and Automation (ICRA)*, (Montreal, QC, Canada), pp. 1274–1280, IEEE, May 2019.

[6] D. Fourie, J. Leonard, and M. Kaess, "A nonparametric belief solution to the Bayes tree," in *2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, (Daejeon, South Korea), pp. 2189–2196, IEEE, Oct. 2016.

[7] A. Dai, M. Nießner, M. Zollöfer, S. Izadi, and C. Theobalt, "BundleFusion: Real-time globally consistent 3D reconstruction using on-the-fly surface re-integration," *ACM Transactions on Graphics 2017 (TOG)*, 2017.

[8] J. J. Park, P. Florence, J. Straub, R. Newcombe, and S. Lovegrove, "DeepSDF: Learning Continuous Signed Distance Functions for Shape Representation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 165–174, 2019.

[9] B. Mildenhall, P. P. Srinivasan, M. Tancik, J. T. Barron, R. Ramamoorthi, and R. Ng, "NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis," in *Computer Vision – ECCV 2020* (A. Vedaldi, H. Bischof, T. Brox, and J.-M. Frahm, eds.), vol. 12346, pp. 405–421, Cham: Springer International Publishing, 2020.

[10] W. Dong, J. Park, Y. Yang, and M. Kaess, "GPU Accelerated Robust Scene Reconstruction," in *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, (Macau, China), pp. 7863–7870, IEEE, Nov. 2019.