DCPAR: Dynamic Collaborative 3D Reconstruction by maintaining privacy and safety implications in active construction sites

Jerrin Bright
Researcher, Department of Aerospace Engineering,
Indian Institute of Science, Bangalore, India.

November 2021

Abstract

Inspection and monitoring systems in an active construction site are a rather expensive and time-consuming process that requires expensive sensors and types of equipment. The system designed should not affect the privacy of the stewards working on the site. Also, an active construction site tends to be most dynamic in terms of the infrastructure and working stewards and moving of components or tools frequently. Considering these issues, a Dynamic Collaborative Perception Aware Reconstruction (DCPAR) technique is proposed which ultimately solves all the common issues in reconstruction and inspection in an effective manner. Current multi-agent SLAM techniques fail to address the real-time scene dynamics. Thus, a distributed, scalable multi-agent system is designed with minimal human intervention leveraging monocular vision and inertial sensors capable of working efficiently in dynamic environments constantly updating the global map keeping in mind the safety and privacy of the stewards.

Objective

Some of the major contributions of the proposed system include:

- Multi-agent dynamic real-time reconstruction technique working in a distributed collaborative manner.
- The system doesn't rely on texture variances considering the use of depth maps.
- Use of virtual loop closure for better odometry corrections by reducing re-projection error using key-frame from other agents.
- Communicating via a map module which is a database storing key-frames aids in saving computational cost significantly.
- Dynamic Class Removal (DCR) for privacy and safety implications which also gives more accurate depth maps.

Preliminary Literature Review

Benchmarked multi-system collaborative scene recreation techniques include [2] (two agents focusing on handheld/ wearable devices), [10] (three to twelve agents for handheld devices) resulting in one global map and techniques like [12] (three agents focusing on UAVs) and [6] (four agents on sequences of EUROC dataset) resulting in multiple local maps. There has been a lot of monocular depth algorithm including [3], [4] and [1] to name a few. The depth technique adapted by the proposed system is inspired from [7]. Image in-painting was done using [8] which is a generative image in-painting technique using adversarial edge learning. Algorithms like [11] and [9] work on low overlap feature matching which could be used to attain more robust global maps.

Methodology

In this section, the architecture of the proposed system is explained in detail. Figure 1 visualizes the architecture workflow for two agents. The architecture can also be extended for multiple agents communicating with the 'map module' which will contain the key-frames estimated by the respective agent. Communication via key-frames also aids in less consumption of computation. Firstly key-frames are selected and sent to the decoders for depth estimation. This solves redundancy issues of frames and thereby saves computational cost significantly.

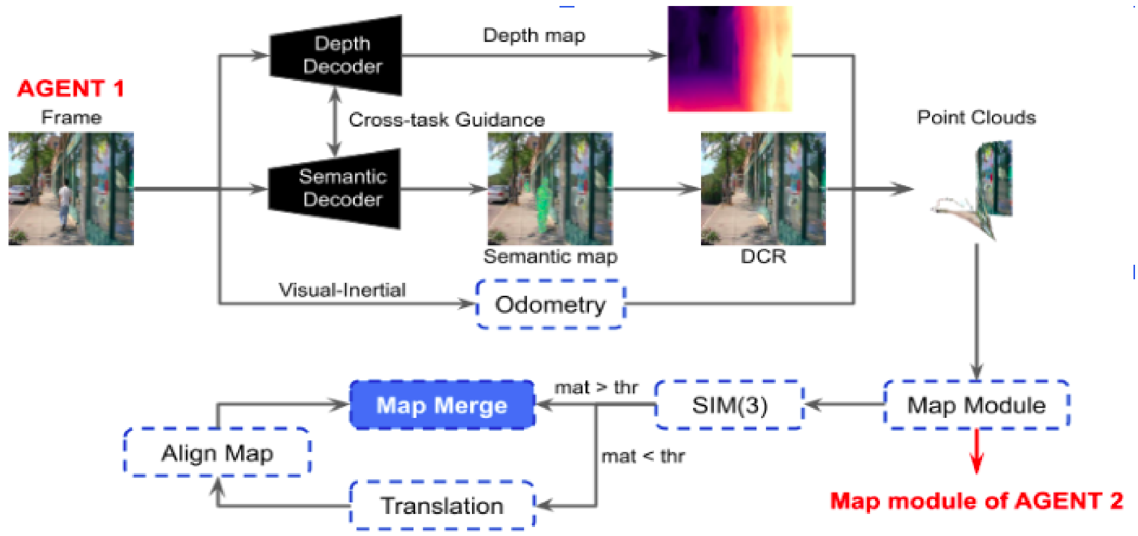

Figure 1: Architecture of the proposed system

Dynamic Class Removal (DCR) is considered to be a segmentation and image in-painting problem. Semantic Segmentation using dense U-Net backbone was leveraged and adversarial edge learning and connecting technique was used for image in-painting. Dilated convolution layers and residual blocks were used for in-painting the 'hole' as a result of semantic identification. In the below figure 2, the DCR algorithm is tested with people to be the only dynamic class to be screened considering the privacy of stewards in construction sites.

Figure 2: Dynamic Class Removal (DCR)

Monocular Depth Estimation is done using the semantic map from the DCR framework which guides and estimates the depth maps in an unsupervised manner by establishing cross-task guidance with semantic segmentation (U-Net). In the below figure 3, the proposed depth algorithm is trained and tested in the frames fed by the DCR framework which was a by-product of the desired screening of dynamic classes depending on the use-case. Once the output key-frame from the DCR is obtained, pre-processing is done. Pre-processing includes depth estimation and conversion of point clouds from the depth image.

Open3D based point clouds are obtained from the depth map. These point clouds are then sent to the map module. Fusion of local maps takes place in respective map modules in a decentralized manner thereby reducing latency. In the map module, virtual loop closure between the respective agent takes place upon the high matching percentage of key features. This will in turn be used to correct the re-projection error giving more accurate pose estimation resulting in better map fusion.

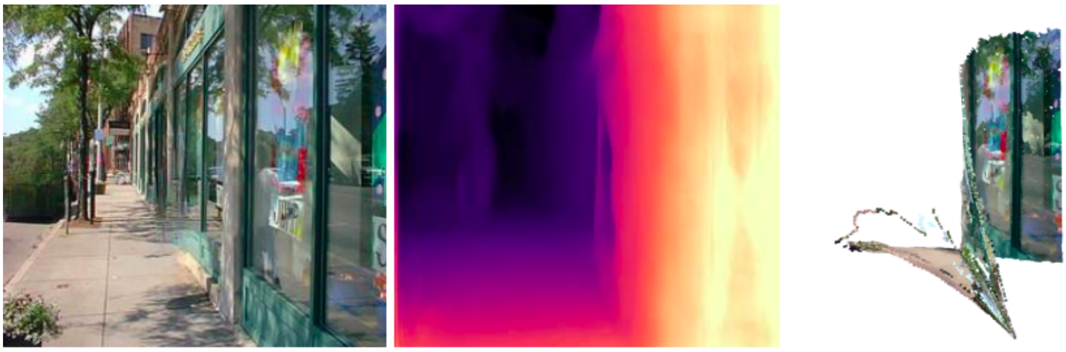

Figure 3: Pre Processing

The matching probability will be compared with a threshold and if it exceeds, the maps will be merged instantaneously. If the overlap is very less, the virtual loop closure will be triggered and map merge will take place after aligning with odometry information. This way, computation, and efficient mapping are established.

Preliminary Results

Pre-processing including dynamic class in-painting, depth map and point cloud estimation from key-frames has been implemented and tested using real-time data. The key-frames where first sent into the in-painting module to remove dynamic classes thereby ensuring privacy and depth accuracy is maintained. Testing was done using Logitech C930e monocular camera and Jetson Xavier NX development board. The entire pre-processing module runs at 10 key-frames per second when tested in on-board Jetson board. Also, multi-session SLAM techniques including CORB and RTABMAP where exploited and studied extensively to bring forth a robust map merging technique.

Timeline

The research has to start with conducting extensive thorough literature review to identify gaps in knowledge and the scope of the research. Basic experimentation especially with the frame pre-processing has been done. But modifying its performance and reducing computational cost is mandatory requirement. Guided Depth map is a very novel technique and it will approximately take quarter of a year to complete it with good accuracy. Another two to three months to optimize the pre-processing modules with the best algorithms. Then local map has to be made from the attained point cloud and stored in a database followed by map merging. Map merging is a very intricate process which will drastically change the efficiency and computation of the system. There is lot of approaches to do it with different techniques, constraints and assumptions which will take the most number of duration in this research estimated to be about half a year. The proposed architecture can also be enhanced with the future works mentioned in the upcoming sections. Coupling the guided depth map with optical flow information resulting in optical maps capable of comprehending both depth and dynamic scene understanding can be done in a month or two. A total of one and a half years will be perfect for the completion of the proposed system with few additions of the future works.

Conclusion

In this proposal, a novel architecture is proposed for collaborative image reconstruction in continually changing environments (infrastructure) with non-dynamic scene constraints. With very less sensor usage, that is monocular vision sensor and inertial sensor, the proposed system can work effectually in a robust manner. Safety and privacy implications were also taken into account while formulating the architecture. The proposed system was focused on construction sites primarily but has a wide range of applications especially in the inspection domain and in agile system navigation in unknown cluttered dynamic environments.

Future Works

Some of the immediate future add-ons to this architecture that will ultimately make a significant difference to the performance of this system include:

- Feature matching and registration in low overlap conditions, inspired from the works of [5].
- Unsupervised or Semi-Supervised semantic segmentation approach by coupling in optical flow information like Lucas Kanade Technique from the monocular vision sensor.
- Bring in precise constraints to limit updating the map thereby significantly reducing the computational cost of the architecture.
- Design and implement a more modular system, wherein each module can be interchanged and tested with different algorithms with ease to improve the efficiency of the system.

References

[1] Ibraheem Alhashim and Peter Wonka. High quality monocular depth estimation via transfer learning. *CoRR*, abs/1812.11941, 2018.

[2] Robert Castle, Georg Klein, and David W. Murray. Video-rate localization in multiple maps for wearable augmented reality. In *2008 12th IEEE International Symposium on Wearable Computers*, pages 15–22, 2008.

[3] Clément Godard, Oisin Mac Aodha, and Gabriel J. Brostow. Unsupervised monocular depth estimation with left-right consistency. *CoRR*, abs/1609.03677, 2016.

[4] Clément Godard, Oisin Mac Aodha, and Gabriel J. Brostow. Digging into self-supervised monocular depth estimation. *CoRR*, abs/1806.01260, 2018.

[5] Shengyu Huang, Zan Gojcic, Mikhail Usvyatsov, Andreas Wieser, and Konrad Schindler. PREDATOR: registration of 3d point clouds with low overlap. *CoRR*, abs/2011.13005, 2020.

[6] Marco Karrer, Patrik Schmuck, and Margarita Chli. Cvi-slam—collaborative visual-inertial slam. *IEEE Robotics and Automation Letters*, 3(4):2762–2769, 2018.

[7] Marvin Klingner, Jan-Aike Termöhlen, Jonas Mikolajczyk, and Tim Fingscheidt. Self-supervised monocular depth estimation: Solving the dynamic object problem by semantic guidance. *CoRR*, abs/2007.06936, 2020.

[8] Kamyar Nazeri, Eric Ng, Tony Joseph, Faisal Z. Qureshi, and Mehran Ebrahimi. Edgeconnect: Generative image inpainting with adversarial edge learning. *CoRR*, abs/1901.00212, 2019.

[9] Milos Prokop, Salman Shaikh, and Kyoungsook Kim. Low overlapping point cloud registration using line features detection. *Remote Sensing*, 12:61, 12 2019.

[10] Patrik Schmuck and Margarita Chli. CCM-SLAM: Robust and efficient centralized collaborative monocular simultaneous localization and mapping for robotic teams. In *Journal of Field Robotics (JFR)*, 2018.

[11] John Stechschulte, Nisar Ahmed, and Christoffer Heckman. Robust low-overlap 3-d point cloud registration for outlier rejection. In *2019 International Conference on Robotics and Automation (ICRA)*, pages 7143–7149, 2019.

[12] Danping Zou and Ping Tan. Coslam: Collaborative visual slam in dynamic environments. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(2):354–366, 2013.