EDCPAR: An Event-based Dynamic Heterogeneous Collaborative Dense Mapping and Task-Specific Control in Ubiquitous Environments

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

December 2021

Abstract

Inspection and monitoring using ground robots in rough terrains is less efficient (in terms of speed, computationally exhaustive process). In this research, we propose an Event-driven Dynamic Collaborative Perception Aware Reconstruction and Task-Specific Control (EDCPAR) system using Anymal (Agent-I) - legged robot and DJI Matrice 100 UAV (Agent-II). Agent-I primarily performs dense mapping of the environment and Agent-II does the assigned task leveraging the global map. Thus, a distributed, scalable heterogeneous multi-agent system is designed with minimal human intervention leveraging stereo event vision and inertial sensors for Agent-I in dynamic environments constantly updating the global map. Both the agents are instructed to perform a predefined set of tasks by the user. Agent-II is equipped with monocular camera to perceive the environment using perception aware reconstruction pipeline and navigate through it in an agile manner.

Objective

Some of the major contributions of the proposed system include: Dynamic class removal from reconstruction resulting in more accurate occupancy grid. Agent-II uses the perception aware reconstruction pipeline to perform dense reconstruction and sense and avoid of static and dynamic obstacles. Use of virtual loop closure for better odometry corrections by reducing re-projection error using key-Landmarks from other agents. Agent-I is equipped and uses the Neuromorphic/Event camera, a bio-inspired vision sensor with high temporal performance and very low latency. The aforementioned sensor has HDR capabilities making it a viable option for work-case in both dark and bright conditions. Optimal trajectory planning of Agent-II is estimated leveraging the global map. Use of barrier functions in Agent-I increases the performance of the system while reducing the computation cost significantly with very less latency. Kinodynamic planning of Agent-II enables the system to perform agile maneuver and enable navigation in very cluttered and narrow environments.

Preliminary Literature Review

Benchmarked multi-system collaborative scene recreation techniques include [1] (two agents focusing on handheld/ wearable devices), [11] (three to twelve agents for handheld devices) resulting in one global map and techniques like [14] (three agents focusing on UAVs) and [5] (four agents on sequences of EUROC dataset) resulting in multiple local maps. There has been a lot of monocular depth algorithm including [2] and [3] to name a few. The depth technique adapted by the proposed system is inspired from [6]. Image in-painting was done using [8] which is a generative image in-painting technique using adversarial edge learning. Algorithms like [12] and [9] work on low overlap feature matching which could be used to attain more robust global maps. Barrier functions have been implemented for ground and aerial robots by [7] and [10]. An kinodynamic path planning [4] version proposed by [13], leverages hybrid A*, B-Spline and time adjustment.

Methodology

In this section, the architecture of the proposed system is explained in detail. Figure 1 visualizes the architecture workflow for two agents - Anymal (Agent-I) and UAV (Agent-II). The architecture can also be extended for multiple agents communicating with the 'map module' which will contain the key-landmarks and global dense map generated by the respective agent. Here in the proposed architecture, for ease of understanding only two agents are used. Agent-II can navigate in a very agile manner with low latency and kinodynamic abilities thereby collecting real-time data in a very robust manner.

Agent-II uses this map to optimally decide the navigation trajectory using cost efficient algorithms including Hybrid A* and optimization strategy including B-Spline algorithm. Model Predictive Controller based Control Barrier Function (MPC-CBF) is used for optimized trajectory planning for Agent-I which has a very good balance between computational efficiency and trajectory following with precision. 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.

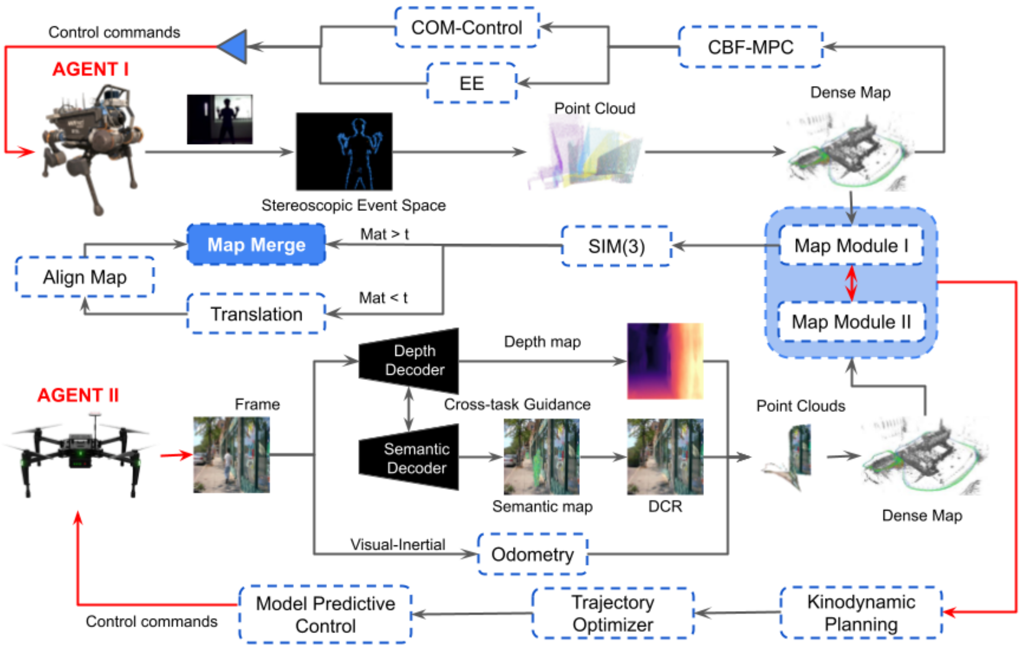

Figure 1: Architecture of the proposed system

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, can be extended to 'N' number of classes depending on the scope of the environment. The main agenda of inhibiting DCR is to result in more meaningful scene reconstruction without any dynamic obstacles.

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 figure 3, the proposed depth algorithm is trained and tested in frames fed by the DCR framework which was a by-product of the screening of dynamic classes.

Point clouds are extracted from the depth maps and 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 be used to correct the re-projection error resulting accurate pose estimation. The matching probability will be compared with a threshold

Figure 2: Dynamic Class Removal (DCR)

and if it exceeds, the maps will be merged instantaneously. If overlap is less, loop closure will be triggered and map merge will take place.

Figure 3: Pre-Processing (Depth map, Point Clouds)

Agent I uses event camera for perception module. Event camera is an asynchronous vision sensor that measures per-pixel brightness changes and outputs the stream of events. Event cameras work way better than traditional cameras especially in high dynamic low lighting conditions considering it has high spatio-temporal clarity.

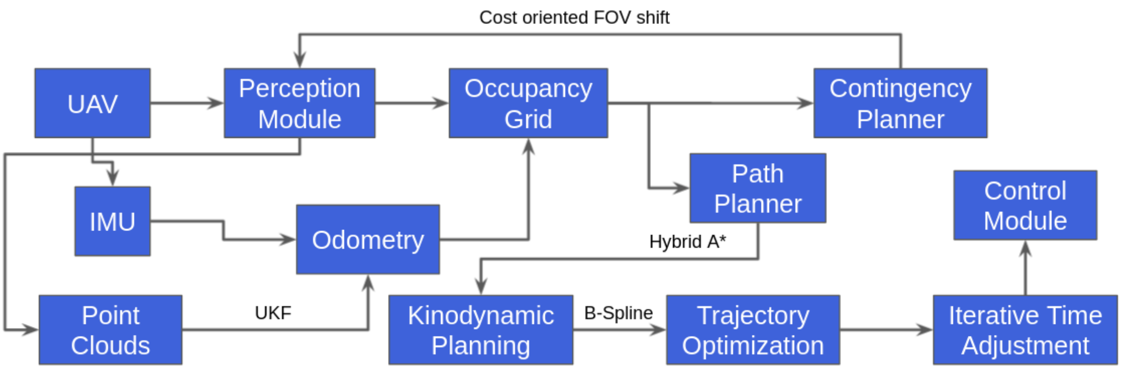

Figure 4: Planning Module

Kinodynamic path planning approach is adapted to navigate from the information perceived from the perception model for Agent-II. B-spline optimization is done to improve the smoothness and clearance of the estimated trajectory. Hybrid State A* algorithm is adapted to search a feasible path between the current pose and desired goal pose as a graph based problem over the generated configuration space. Cost is assigned to generated trajectory and heuristics enforce faster search.

Deriving the mathematical model of Agent-II includes understanding the reference frames, deriving kinematic and dynamic equations of Agent-II. The inertial frame, body frame and the vehicle frame upon transformation can be shifted to a common reference frame. Here formulation of equation of motion is done on the body frame considering inertial measurements are in this

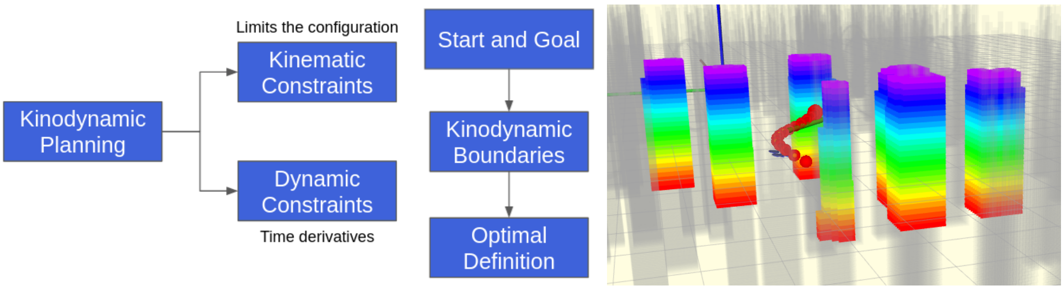

Figure 5: Kinodynamic Path Planning

frame (symmetry of Agent-II tends to simplify mathematical equations). Euler angles are derived in the vehicular frames and angular rate from body frame.

The control module for Agent-II majorly consist of Model Predictive Controller for navigating effectually from the information and commands understood from the perception and the planning modules. Altitude and Position are studied as outputs by the PID controller. Genetic algorithm (optimization and search technique) based tuning is done for the PID controller coefficients.

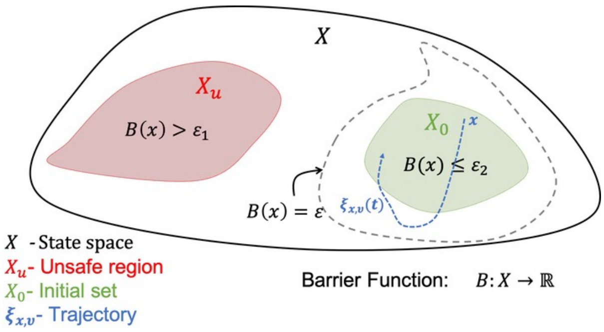

Figure 6: Barrier Function

The control module of Agent-I is based on control barrier functions which uses MPC and reduces the cost thereby establishing a very desirable control point navigation. It splits the free space (C*) from the occupancy grid (C) and uses it as constraints and cost functions (J) is aimed to be as less as possible.

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 depth accuracy for reconstruction is maintained. Testing was done using Logitech C930e 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. Kinodynamic planning for IRIS drone was experimented with pixhawk flight controller and good path planning was established. Currently working on MPC-CBF approach for ground robots.

Conclusion

In this proposal, a novel architecture is proposed for collaborative dense mapping and reconstruction in continually changing environments (infrastructure) with non-dynamic scene constraints. With very less sensor usage, monocular vision sensor, Event vision sensor and inertial sensor, the proposed system can work effectually in a robust manner while decreasing the system latency and increasing the overall performance of the system. The proposed system was for ubiquitous sites having 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: 1.) Feature matching and registration in low overlap conditions, inspired from the works of [4]. 2.) Unsupervised or Semi-Supervised semantic segmentation approach by coupling in optical flow information like Lucas Kanade Technique from the monocular vision sensor. 3.) Bring in precise constraints to limit updating the map thereby significantly reducing the computational cost of the architecture. 4.) 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] 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.

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

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

[4] 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.

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

[6] 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.

[7] Aniketh Manjunath and Quan Nguyen. Safe and robust motion planning for dynamic robotics via control barrier functions. *CoRR*, abs/2011.06748, 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] Esteban Restrepo, Ioannis Sarras, Antonio Loria, and Julien Marzat. 3d uav navigation with moving-obstacle avoidance using barrier lyapunov functions. *IFAC-PapersOnLine*, 52(12):49–54, 2019. 21st IFAC Symposium on Automatic Control in Aerospace ACA 2019.

[11] Patrik Schmuck and Margarita Chli. CCM-SLAM: Robust and efficient centralized collaborative monocular simultaneous localization and mapping for robotic teams. In *JFR*.

[12] 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.

[13] Boyu Zhou, Chuhao, and Shaojie Shen. Robust and efficient quadrotor trajectory generation for fast autonomous flight. *IEEE Robotics and Automation Letters*, 4(4):3529–3536, 2019.

[14] 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.