Asynchronous Collaborative Autoscanning with Mode Switching for Multi-Robot Scene Reconstruction Supplemental Material

JUNFU GUO, University of Science and Technology of China, China
CHANGHAO LI, University of Science and Technology of China, China
XI XIA, University of Science and Technology of China, China
RUIZHEN HU[*], Shenzhen University, China
LIGANG LIU, University of Science and Technology of China, China

CCS Concepts: • **Computing methodologies** → Shape analysis.

Additional Key Words and Phrases: Indoor scene reconstruction, autonomous reconstruction, multiple robots cooperation, asynchronous task assignment

ACM Reference Format:
Junfu Guo, Changhao Li, Xi Xia, Ruizhen Hu, and Ligang Liu. 2022. Asynchronous Collaborative Autoscanning with Mode Switching for Multi-Robot Scene Reconstruction Supplemental Material. *ACM Trans. Graph.* 41, 6, Article 198 (December 2022), 4 pages. https://doi.org/10.1145/3550454.3555483

1 METHOD DETAILS

Details in reconstruction task generation. Suppose there are two different viewpoint $\mathcal{V}_1 = \{\boldsymbol{p}_1, \boldsymbol{d}_1\}$ and $\mathcal{V}_2 = \{\boldsymbol{p}_2, \boldsymbol{d}_2\}$, where $\boldsymbol{p}_1$, $\boldsymbol{p}_2$ are the positions of viewpoints, and $\boldsymbol{d}_1$, $\boldsymbol{d}_2$ are the directions of viewpoints. Then we compute the distance bewteen $\mathcal{V}_1$ and $\mathcal{V}_2$:

$$D(\mathcal{V}_1, \mathcal{V}_2) = ||\boldsymbol{p}_1 - \boldsymbol{p}_2||_2 + ||\boldsymbol{d}_1 - \boldsymbol{d}_2||_2. \tag{1}$$

indicating the differences between two viewpoints both on position and direction. We define the distance threshold $D_0 = 1.414$ for viewpoints, We also define that two viewpoints $\mathcal{V}_1$ and $\mathcal{V}_2$ are neighboring viewpoints, if $D(\mathcal{V}_1, \mathcal{V}_2) < D_0$.

Details in MDMTSP solver. In our modified MDMTSP problem, we attempt to search for the optimal assignment of the tasks to the robots aiming at minimizing the distance term and capacity term. These two terms are mentioned in section 5.1. To solve this complex problem which includes both the ownership and the sequences of the tasks, we adopt the clustering-while-assigning method to approximate the optimal solution iteratively.

Since each robot processes only one category of tasks in one assignment, we need to decide the scanning mode of the robots in the assignment first. To achieve this, we add another iteration process via the simulated annealing method to optimize the scanning modes. Algorithm 1 shows the optimization details.

ALGORITHM 1: Simulated annealing algorithm

```
Input: Current weighted graph G;
Information of robots: $\mathcal{R} = \{\mathcal{R}_1, ..., \mathcal{R}_R\}$;
Initial duality tuple for robots: $C = \left\{C_r | C_r = \left(\mathcal{R}_r, \{\mathcal{T}_{r_1}, ...\}\right)\right\}$;
Time threshold $T$;
Initial time $t_0$;
Annealing rate $r$;
Output: Final duality tuple for each robot: $C = \{C_1, ..., C_R\}$
$t \leftarrow t_0$;
$c \leftarrow$ ComputeAssignmentCost($C$);
while $t \leq T$ do
    $t \leftarrow t \cdot r$;
    $\mathcal{R}^* \leftarrow$ RandomTurbulant($\mathcal{R}$);
    $C^* \leftarrow$ AssignmentOptimizer($\mathcal{R}^*$, G);
    $c^* \leftarrow$ ComputeAssignmentCost($C^*$, G);
    if $e^{(c-c^*)/t} >$ Random(0.0,1.0) then
        $\mathcal{R} \leftarrow \mathcal{R}^*$;
        $C \leftarrow C^*$;
        $c \leftarrow c^*$;
    end
end
```

In the AssignmentOptimizer of Algorithm 1, we perform a soft clustering method with modified *Gauss Mixture Model* (GMM), where the capacity term is multiplied by each robot's likelihood to consider the unprocessed tasks. After each optimization step, we calculate the assignment cost of the new cluster and decide whether to update the states via the comparisons of the candidate assignment cost.

The situation when a robot finishes all possible tasks nearby and is ready to move to the new area at a distance is also considered in the constrained parameters. Once a robot has processed more than two tasks and has longer than 15m to travel, the rest of the tasks will be ignored.

2 IMPLEMENTATION DETAILS

Semantic reconstruction. We choose *Voxblox++* [Grinvald et al. 2019] as the basic framework of our reconstruction module. It is a lightweight 3D reconstruction framework, which could provide

[*] Corresponding author: Ruizhen Hu (ruizhen.hu@gmail.com)

Authors' addresses: Junfu Guo, University of Science and Technology of China, China; Changhao Li, University of Science and Technology of China, China; Xi Xia, University of Science and Technology of China, China; Ruizhen Hu, Shenzhen University, China; Ligang Liu, University of Science and Technology of China, China.

0730-0301/2022/12-ART198 $15.00
https://doi.org/10.1145/3550454.3555483

the data needed for our strategy. It presents an approach to incrementally building geometrically accurate volumetric maps of the environment that additionally contain information about the individual object instances observed in the scene. Benefit from this lightweight framework, we could understand objects in the real world at both geometric and semantic levels. The source version of this framework is designed for a single camera, making it unable to receive and process data from multiple cameras simultaneously.

Parameter setting. In the simulation environment, the moving speed is 0.5m/s for the explorer, and 0.2m/s for the reconstructor when moving between task views over each object. We use the dynamic spatial resolution of the occupancy grid. Hence, the resolution changes due to the expansion of the unknown map.

For exploration tasks, we set the number of the exploration tasks as $T = 4R$, where R is the number of the robots. The range of selecting candidate exploration tasks is within 1.5m since the `Primesense Carmine 1.09` has the operation range of 0.3-3m.

To fit the input point cloud size in *GR-Net* in the reconstruction task generation process, we downsample each object into 2048 points and output a 16384-point predicted object. Exploration tasks are set in the fixed height of 1.1m, and reconstruction tasks are limited to the range of 0.3m to 1.8m on Z-axis.

Implementation of Dong. Since [Dong et al. 2019] does not contain the object reconstruction procedures, we replace their *Voxel Hashing* module with voxblox-plus-plus in our method to observe their reconstruction quality towards the objects. In [Dong et al. 2019]'s work, the robots share the same identity similar to the explorer. Moreover, their robots work in the same intervals that only when all robots finish the tasks, the control center can start to generate new tasks. The speed of the robots in their method is set as 0.3m/s, according to the paper.

Treatment of error data. Multi-view inconsistencies in the segmentation and incorrrect prdictions are unavoidable problems in the field of reconstruction. We overcome these shortcomes using the same method as the one in Voxblox++, which first reconstructs the surface of the object based on depth and normal information, and then selects the best semantic label for every identifiable object based on voting.

3 MORE RESULTS AND EVALUATIONS

3.1 Scalability analysis

Study on number of robots. The number of the robots is also evaluated in our experiments. Knowing that the scanning efficiency of various robots is strongly related to the scale of the scene and the complexity of the layout, we deploy the experiments among all the scenes in our dataset to study the scalability of the method. We analyze the performance of our method with different numbers of robots, and Figure 1 shows some meaningful results. We observe that the scene with large areas needs more robots to have a comparable efficiency. Besides, it is clear that the more robots we use, the less time consumption our method costs. Note that the time load balance increases when the scene to be reconstructed is larger, due to the task assignment occurring more times during the process.

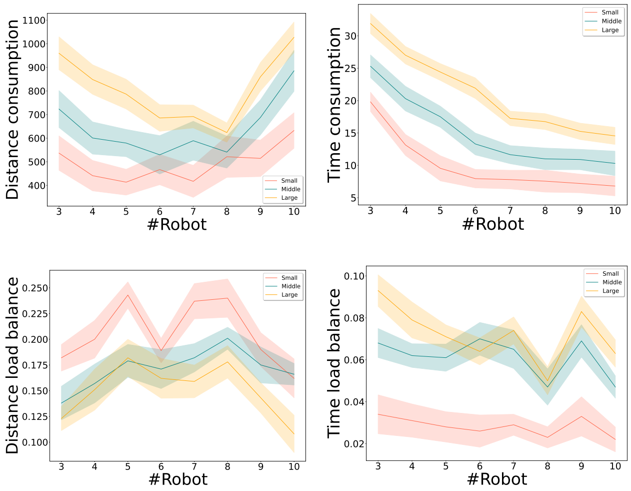

Fig. 1. Studies with different number of robots.

Table 1. Time cost of some important components of our method. All values are measured in milliseconds.

Scene Level	Small	Medium	Large
Frontier extraction	4.67	5.14	6.26
Frontier filtration	0.208	0.212	0.217
Frontier view generation	122.8	123.4	125.6
Object view generation	329.6	479.2	525.3
Task assignment	1728.3	1936.6	2162.5

Despite the various scanning efficiency, the reconstruction qualities of all these cases are at the same level, indicating that our method has the capability of driving multiple robots scanning in an unknown environment with the robots' numbers from 3 to 10. Moreover, when reconstructing in an unknown environment without any prior knowledge, it is difficult to choose a proper number of robots, but our method can drive various numbers of robots to accomplish the scanning mission to a similar reconstruction quality within a reasonable time-consumption and distance-consumption.

Timing. We also counted the time consumption of each component in the decision-making process. Table 1 shows the total time consumption on different level of scenes.

From the results, we can see that with the increase in the scene scales, the time consumption of the frontier tasks' generation increases little, and the consumption of the reconstruction sub-tasks' generation has the computational complexity of $O(n)$, where n is the number of the objects discovered. The former is because we filter the fixed number of the frontier tasks in the scene, which is irrelevant to the scale of the scene. And the latter is because we take the same task-extraction procedure to every object discovered in the scene, which shares a positive correlation to the number of the objects. Since we adopt an approximation method to compute the assignment of the tasks, the number of tasks increases as the number of objects, and the computational time of assignment is also $O(n)$. Moreover, the final TSP is also $O(n^2 log(n))$ since we use the approximation method to search for the optimal solution. As a result, the computational time is roughly proportioned to the objects in the scene.

Table 2. Details of all scenes we used to evaluate. mp3d means the scene comes from Matterport3D dataset, and front3d means the scene comes from Front3D dataset.

Scene	Scene Level	Area/m^2	#Room	#Object
mp3d01	Small	176	19	45
mp3d02	Small	186	29	65
mp3d03	Medium	220	20	24
mp3d04	Medium	288	22	42
mp3d05	Large	328	18	35
mp3d06	Large	350	22	52
front3d01	Small	182	9	40
front3d02	Small	243	6	53
front3d03	Medium	263	11	71
front3d04	Medium	293	12	63
front3d05	Large	318	11	39
front3d06	Large	468	9	61

Table 3. Details of different levels of scenes we used to evaluate. In all of our experiments, if we don't give a further specified explanation, the metrics are the average of the values on all scenes.

Scene Level	Area/m^2	#Room	#Object
Small	196.75	15.75	50.75
Medium	266.00	16.25	50.00
Large	366.00	15.00	46.75
Mean	276.250	15.667	49.167

Table 4. Comparing our method with the work of [Liu et al. 2018] in reconstruction quality of scene.

method	Scene Completeness			Scene Accuracy		
	Small	Medium	Large	Small	Medium	Large
NBO × 1	71.61	73.32	68.17	0.023	**0.022**	0.037
NBO × 4	**73.83**	71.41	**71.38**	**0.020**	0.025	0.036
Ours	72.60	**73.57**	70.25	0.022	0.023	**0.034**

3.2 Scene-Level evaluation

We also implement the scene-level metrics to evaluate the experiment results. The measurements are:

- **Scene Completeness (S-Comp)** measured by the percentage of the covered surface of the entire scene. It's similar to the way we calculate the metric O-Comp.
- **Scene Accuracy (S-RMS)** measured by the average distance error of the reconstructed scene. It's similar to the way we calculate the metric O-RMS.

Table 5. Comparing our method with the work of [Dong et al. 2019] in reconstruction quality of scene.

method	Scene Completeness			Scene Accuracy		
	Small	Medium	Large	Small	Medium	Large
Dong	64.83	63.83	61.17	0.023	0.033	0.039
Ours	**72.60**	**73.57**	**70.25**	**0.022**	**0.023**	**0.034**

Table 6. Ablation studies in reconstruction quality over twelve virtual scenes in our dataset. All evaluations are running with 4 robots.

Method	S-Comp	S-RMS
NoSw[3+1]	70.62	0.0340
NoSw[2+2]	68.92	0.0273
NoSw[1+3]	69.22	0.0213
NoRe	69.01	0.0328
NoEx	63.24	**0.0206**
NoFlow	71.88	0.0218
Ours	**72.14**	0.0260

Table 4 shows that both the completeness and the accuracy of the reconstruction result in [Liu et al. 2018] and our method are similar. On the other hand, Table 5 shows that the frontier-based method can cause holes in the reconstruction results, which significantly reduces the reconstruction quality of the entire scene. These comparisons show that the reconstruction quality can be improved if the objects are specifically considered.

Table 6 shows the ablation studies on the reconstruction quality of the scene.

3.3 Ablation study on energy constrains

No energy constraints (NoCo): We cut off the constraints in Section 5.2 to test the performance of this module. Table 7 shows that if we consider the energy constraints, all the metrics remain no discernible differences except the time and distance consumption. With the energy constraints, the distance of the paths for some robots can be shorter if necessary, leading to more frequent assignment procedures of the tasks. Thus the percentage of the waiting time of robots increases a little, while the distance of the paths between robots becomes more balanced, which leads to lower distance load balance. Moreover, since the tasks are assigned more balanced, both the time and distance consumption are lower than NoCo.

3.4 Ablation study on reconstruction tasks

No reconstruction tasks(NoRT): We remove all reconstruction tasks and only assign exploration tasks to robots during the scanning process. Table 7 shows that if reconstruction tasks are considered, the execution efficiency in our baseline is slightly worse compared to NoRT. This is because there are more tasks to be finished in our method, reconstruction quality of both scenes and objects increases significantly.

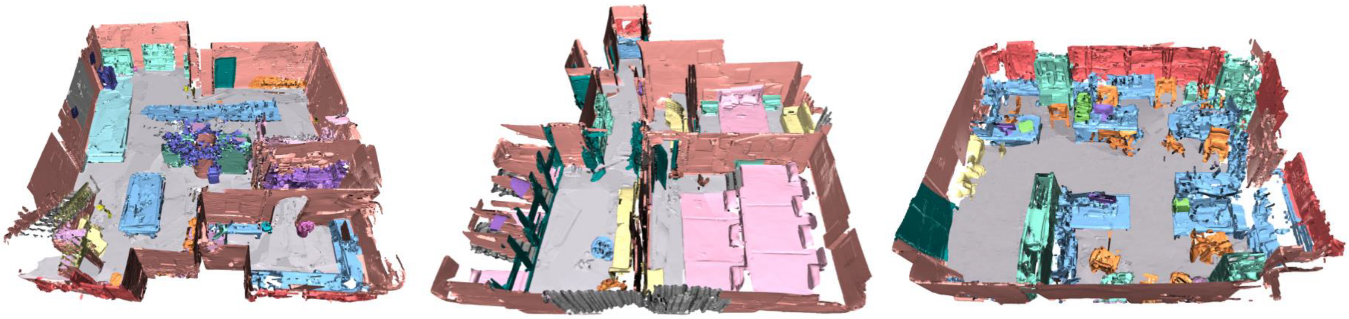

Fig. 2. Three real scenes scanned and reconstructed by our method with objects in different colors.

Table 7. Ablation studies with NoCo and NoRT in reconstruction quality of objects, reconstruction efficiency, and load balance.

Quality	Scene Completeness			Scene Accuracy		
	Small	Medium	Large	Small	Medium	Large
NoCo	**72.62**	**73.87**	70.18	**0.021**	**0.023**	0.035
NoRT	65.29	64.12	61.41	0.022	0.031	0.037
Ours	72.60	73.57	**70.25**	0.022	**0.023**	**0.034**
Quality	Object Completeness			Object Accuracy		
	Small	Medium	Large	Small	Medium	Large
NoCo	**66.19**	**72.52**	70.01	**0.035**	**0.039**	0.034
NoRT	54.28	41.93	40.37	0.059	0.078	0.095
Ours	66.18	72.49	**70.03**	**0.035**	**0.039**	**0.033**
Efficiency	Time Consumption			Distance Consumption		
	Small	Medium	Large	Small	Medium	Large
NoCo	16.3	21.8	29.0	626.7	757.6	1216.8
NoRT	**11.8**	**17.6**	**21.8**	**493.7**	**529.8**	**741.1**
Ours	14.0	18.9	24.7	536.1	620.7	848.5
Balance	Distance Load Balance			Time Load Balance		
	Small	Medium	Large	Small	Medium	Large
NoCo	**0.147**	**0.182**	**0.136**	0.061	0.078	0.091
NoRT	0.169	0.195	0.214	**0.057**	**0.071**	**0.079**
Ours	0.151	0.200	0.157	0.062	0.082	0.093

3.5 More reality results

Figure 2 shows other real-world reconstruction results. Objects with different categories are rendered with different colors.

REFERENCES

Siyan Dong, Kai Xu, Qiang Zhou, Andrea Tagliasacchi, Shiqing Xin, Matthias Nießner, and Baoquan Chen. 2019. Multi-robot collaborative dense scene reconstruction. *ACM Transactions on Graphics (TOG)* 38, 4 (2019), 1–16.

M. Grinvald, F. Furrer, T. Novkovic, J. J. Chung, C. Cadena, R. Siegwart, and J. Nieto. 2019. Volumetric Instance-Aware Semantic Mapping and 3D Object Discovery. *IEEE Robotics and Automation Letters* 4, 3 (July 2019), 3037–3044. https://doi.org/10.1109/LRA.2019.2923960

Ligang Liu, Xi Xia, Han Sun, Qi Shen, Juzhan Xu, Bin Chen, Hui Huang, and Kai Xu. 2018. Object-aware guidance for autonomous scene reconstruction. *ACM Transactions on Graphics (TOG)* 37, 4 (2018), 1–12.