

# Asynchronous Collaborative Autoscaning with Mode Switching for Multi-Robot Scene Reconstruction

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

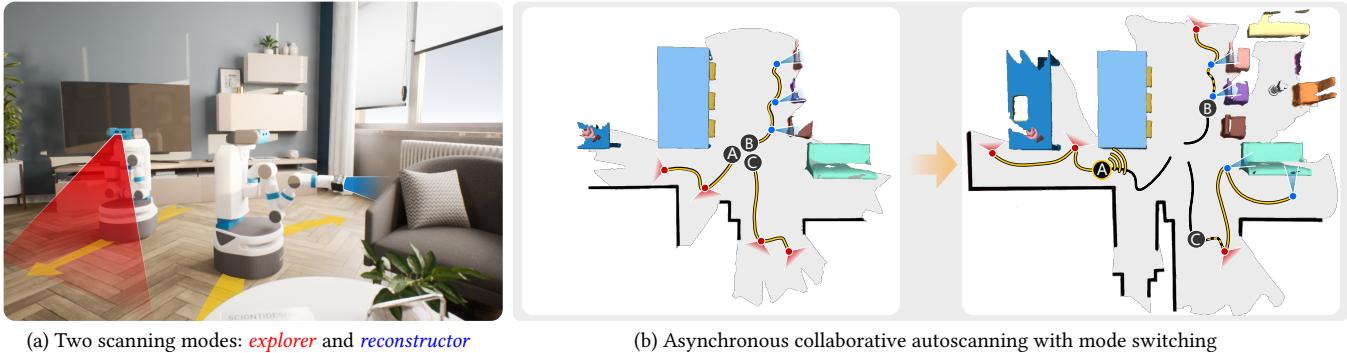


Fig. 1. (a) Two scanning modes: *explorer* mode with rapid moving speed and far vision (shown in red) for exploration task and *reconstructor* mode with low moving speed and narrow vision (shown in blue) for object reconstruction task; (b) Our asynchronous collaborative autoscaning method: given the initially reconstructed scene by turning the three robots around their initial locations, our method first generates a set of tasks with the yellow path for each robot (left). Once one robot has completed all its assigned tasks, robot A in this case, new tasks will be generated and appended to all the robots immediately (right). The completed paths are shown in black, and the paths assigned in the previous round but haven't been completed are shown with dashed yellow lines on top of the black.

When conducting autonomous scanning for the online reconstruction of unknown indoor environments, robots have to be competent at exploring scene structure and reconstructing objects with high quality. Our key observation is that different tasks demand specialized scanning properties of robots: rapid moving speed and far vision for global exploration and slow moving speed and narrow vision for local object reconstruction, which are referred as two different scanning modes: *explorer* and *reconstructor*, respectively. When requiring multiple robots to collaborate for efficient exploration and fine-grained reconstruction, the questions on when to generate and how to assign those tasks should be carefully answered. Therefore, we propose a novel asynchronous collaborative autoscaning method with mode switching, which generates two kinds of scanning tasks with associated scanning modes, i.e., exploration task with *explorer* mode and reconstruction task

with *reconstructor* mode, and assign them to the robots to execute in an asynchronous collaborative manner to highly boost the scanning efficiency and reconstruction quality. The task assignment is optimized by solving a modified Multi-Depot Multiple Traveling Salesman Problem (MDMTSP). Moreover, to further enhance the collaboration and increase the efficiency, we propose a *task-flow* model that activates the task generation and assignment process immediately when any of the robots finish all its tasks with no need to wait for all other robots to complete the tasks assigned in the previous iteration. Extensive experiments have been conducted to show the importance of each key component of our method and the superiority over previous methods in scanning efficiency and reconstruction quality.

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 Autoscaning with Mode Switching for Multi-Robot Scene Reconstruction. *ACM Trans. Graph.* 41, 6, Article 198 (December 2022), 13 pages. <https://doi.org/10.1145/3550454.3555483>

\*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.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

© 2022 Association for Computing Machinery.

0730-0301/2022/12-ART198 \$15.00

<https://doi.org/10.1145/3550454.3555483>

## 1 INTRODUCTION

With an increasing demand for applications such as augmented and virtual reality, gaming and robotics, the research community has studied extensively toward generating digitized 3D indoor scenes with RGB-D sensors using robots [Charrow et al. 2015; Huang et al.

2020; Wu et al. 2014; Xu et al. 2015]. To improve the scanning efficiency, there is a growing trend for adopting multiple robots to reconstruct the unknown area [Duhautbout et al. 2019; Guan 2006; Tian et al. 2021].

The work of [Dong et al. 2019] shows an impressive result in the field of multi-robot dense reconstruction of an unknown environment. However, both scanning efficiency and reconstruction quality are sub-optimal as robots could remain *inactivated* waiting for others before receiving new tasks synchronously, and objects may not be reconstructed with high quality due to rapid scanning without special focus. As noticed in some single-robot autoscanning systems [Huang et al. 2020; Liu et al. 2018; Xu et al. 2015, 2016], objects in the indoor scenes demand more careful scanning due to their fine-grained surface and self-occlusion. Therefore, how to deploy multiple robots to explore the unknown indoor scenes rapidly while obtaining reconstruction results with high completeness and accuracy is still an open question.

To realize high-quality reconstruction of the unknown indoor scenes with high efficiency, our key observation is that other than subdividing the mission into two tasks: exploration tasks and reconstruction tasks, specialized scanning properties of robots should be associated to different tasks to serve their goals better. The exploration task is designed to explore unknown regions and locate the targets for further reconstruction, which requires rapid moving speed and far vision, while the reconstruction task focuses on more detailed scanning of objects to obtain fine-grained geometries, which requires slow moving speed and narrow vision. We denote the above two different scanning modes as *explorer* and *reconstructor* modes, which are associated with the exploration and reconstruction tasks, respectively. To further require multiple robots to collaborate for high exploration efficiency and reconstruction, we propose a novel asynchronous collaborative autoscanning method with *mode switching*, which generates two kinds of scanning tasks with associated scanning modes, and then assigns to the robots to execute in an asynchronous collaborative manner to highly boost the scanning efficiency and reconstruction quality, as shown in Figure 1.

With the exploration tasks generated based on the current frontiers and reconstruction tasks aiming at incomplete objects, assigning these tasks to the robots according to the current situation is a complex problem. To achieve an optimal assignment with the least total traveling distance and processing time, we formulate the problem as a modified *Multi-Depot Multiple Traveling Salesman Problem* (MDMTSP). However, reaching the exact solution of this problem consumes plenty of time and resources [Cheikhrouhou and Khoufi 2021], so it is impossible to search for the optimal solution when the assignment should process frequently. Therefore, we modify the multi-robot goal assignment solution [Faigl et al. 2012] to be suitable for our new problem formulation by approximating the optimal solution using an iterative optimization within a short time.

To prevent some robots from being inactive and waiting for incoming tasks, we propose a task-scheduling model named *task-flow* to allow robots to receive new tasks asynchronously to further enhance the collaboration efficiency. Specifically, the *task-flow* model activates the control center to generate and assign new tasks when any of the robots finishes its current task sequence. As a result,

all robots can be occupied with executing various tasks during the entire scanning procedure without waiting for others to finish their tasks. Both the decision-making and task-processing carry out asynchronously until the scene is fully explored with all objects carefully reconstructed.

Our method has been implemented on top of Robot Operating System (ROS) [Quigley et al. 2009] and tested using both multiple Fetch robots [Wise et al. 2016] in simulation, and four turtlebot3 robots [Amsters and Slaets 2019] in reality. Extensive experiments have been conducted to show the importance of each key component of our method and the superiority of our method over previous methods in scanning efficiency and reconstruction quality.

To summarize, we propose a novel asynchronous collaborative autoscanning method with mode switching, which efficiently utilizes multiple robots for exploring, reconstructing, and understanding an unknown scene. Our technical contributions include:

- Two scanning modes tailored for the execution of two different scanning tasks, i.e., *explorer* for exploration task and *reconstructor* for reconstruction task.
- A new modified MDMTSP and corresponding approximate solver to optimize each robot's task assignment and execution order for high efficiency and load balance.
- A task scheduling model named *task-flow* to minimize the waiting time of each robot and enable asynchronous collaboration among multiple robots.

## 2 RELATED WORK

### 2.1 Single-robot autonomous reconstruction

To date, literature has dug deep into the field of autonomous reconstruction. Studies devote to the scanning of a single object at first [Krainin et al. 2011; Vasquez-Gomez et al. 2014; Wu et al. 2014], then gradually move to the scanning of an entire scene [Charrow et al. 2015; Ramanagopal and Le Ny 2016; Schmid et al. 2020; Xu et al. 2015, 2016]. Unlike the works that focus on the quickly covering of the unknown environment, there are also works trying to enhance the reconstruction quality simultaneously. For example, Xu et al. [2017] harness a time-varying tensor field to plan the robot's movement and solve a 3D camera trajectory under the path constraint. The work of [Liu et al. 2018] is the first one to utilize the semantic information of objects in path planning: they propose a *next best object* (NBO) algorithm and a model-driven *next best view* (NBV) algorithm to scan an indoor scene efficiently with high object reconstruction quality. A similar idea has also emerged in the field of reconstructing scenes with a single drone. Several methods [Heng et al. 2015; Liu et al. 2021; Roberts et al. 2017] are proposed to allow the (micro aerial vehicle) MAV to fulfill exploration and reconstruction objectives simultaneously by adopting an explore-then-exploit or a reconstruct-while-explore strategy. Deep learning-based methods [Fang et al. 2019; Niroui et al. 2019; Sorokin et al. 2021] also show promising results when building 2D maps of unknown scenes. Nonetheless, how to extend these approaches to scene-level 3D representation remains unsolved.

Our work tries to achieve efficient exploration and high-quality reconstruction with a multi-robot system via using both geometry information of the environment and the semantic information of

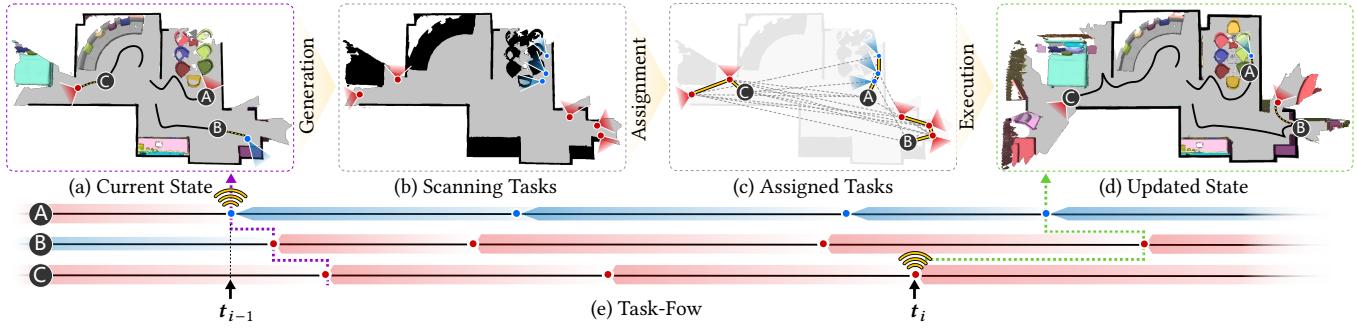


Fig. 2. Overview of our asynchronous collaborative autoscanning method. Once a robot finishes its current tasks, robot A in this case (a), new exploration tasks (red viewpoints) and reconstruction tasks (blue viewpoints) will be generated (b) and assigned to the robots by solving a modified Multi-Depot Multiple Traveling Salesman Problem (MDMTSP) (c). The control center will be activated again if any of the robots finishes the new tasks again, robot C in this case (d).

the objects. Compared to a single robot, the multi-robot system has the advantage that it can process the exploration and reconstruction tasks in parallel instead of finding the balance of choosing between these two tasks and processing them sequentially. However, the multi-robot system has to overcome the obstacle of adjusting the task assignment to save traveling energy and time.

## 2.2 Multi-robot collaborative reconstruction

Multi-robot systems can adopt either centralization or decentralization as their organization strategy. Although the decentralized systems [Atanasov et al. 2015; Sartoretti et al. 2019; Schneider-Fontan and Mataric 1998] have an unparalleled advantage in robustness, the centralized system can improve cooperative efficiency by sharing global information [Li et al. 2019; Matoui et al. 2020] and unifying the decision-making process. Since the communication and localization between robots and the control center can be solved by using auxiliary tools like motion capture cameras in indoor environments, we utilize the centralized scheme in our work for a higher scanning efficiency.

Multi-robot active mapping has been an active research area for decades [Bhattacharya et al. 2014; Burgard et al. 2005; Juliá et al. 2012]. Both rule-based [Albina and Lee 2019; Faigl et al. 2012] and learning-based strategies [Hu et al. 2020; Ye et al. 2022] have been proposed to distribute the assignments among robots to maximize the exploration coverage, but these methods still fail to produce a dense and high-quality reconstruction due to the lack of quality-driven view planning. The most relevant work to ours is [Dong et al. 2019]. They separate the scanning procedure into multiple intervals. Each interval comprises a series of robot scanning, task extraction, task assignment, and path planning processes. Despite the promising results, this method generates new reconstruction tasks only based on the frontiers without distinguishing the objects in the interior scenes. As a result, some holes appear on the surface of the reconstruction results, leading to the declination of the reconstruction quality. And this algorithm fails to take full advantage of each robot. Since assigning the workload evenly is an NP-hard problem, the working time of each robot varies in each interval, leading to the robots being inactivated and waiting for others. In contrast, our method contains not only the exploration tasks to

broaden the known area but also reconstruction tasks to scan the objects in detail. We also add a *task-flow* synchronization model to drive the robots and generate new tasks simultaneously, no robots can be idle until the scene is fully reconstructed.

## 2.3 Multi-Depot Multi-Travelling Salesman Problem

Even though researchers have studied the optimal task assignment problem with constraints in engineering and economics from a variety of extents [Albina and Lee 2019; Trigui et al. 2017], it is still a challenging research axis in the field of robotics. The task assignment problem in a centralized multi-robot system is to drive multiple robots from various depots to multiple task positions with minimum travel distances. The problem can be naturally formulated as a *Multi-Depot Multi-Travelling Salesman Problem* (MDMTSP). The work of [Sundar and Rathinam 2017] introduces the integer linear programming formulation and proposes a branch-and-cut algorithm to reach the optimal solution. This approach, along with other methods [Albina and Lee 2019; Vali and Salimifard 2017], seek to solve the problem directly. However, these solutions are too time-consuming [Cheikhrouhou and Khoufi 2021] to be adopted in the task assignment procedure. There are also other approaches using the meta-heuristic algorithms [Chen et al. 2018; Khoufi et al. 2015; Yuan et al. 2013] to solve the problem from different aspects such as genetic algorithm or ant colony optimization. Still, how to extend these methods to the new situation when there are two interchangeable identities in the formulation remains obscure. In our approach, we extend the work in [Faigl et al. 2012] and modify it to approximate the exact solution via a divide-and-conquer strategy to reduce the computational complexity.

## 3 OVERVIEW

In this paper, we tackle the problem of automatic scene reconstruction with a centralized multi-robot system as in [Dong et al. 2019]: given an unknown indoor scene and the initial positions of several robots, the goal is to explore and scan the scene such that the scanning coverage and reconstruction quality are maximized while the scanning effort is minimized. Figure 2 shows an overview of one iteration of our asynchronous collaborative autoscanning method.

At the very beginning, all the robots turn around to scan their surroundings so that an initial reconstruction is established to start the iteration. Each time when a robot finishes its current task sequence, the control center is activated to generate new scanning tasks based on the current reconstruction and assigns them to all the robots. Then, each robot executes the assigned tasks with associated scanning modes to get the reconstruction updated. To execute the new tasks assigned, each robot adapts its moving speed and camera view according to the task type. The control center will be activated again if any of the robots finish the new tasks again, and the iteration ends if no more new tasks can be generated.

**Task generation.** In our approach, the grand scanning task is divided into two sub-tasks: exploration tasks for high coverage and reconstruction tasks for high quality, which are represented using red viewpoints and blue viewpoints in Figure 2 (b), respectively. The exploration task is responsible for the rapid expansion of the known areas, and the corresponding viewpoints are generated based on the frontiers of the 2D occupancy grid updated with the 3D reconstruction results. The reconstruction task is responsible for covering the objects’ surfaces with high scanning accuracy, and the corresponding viewpoints are generated based on the comparison between the predicted completion and the current scan. More details about the task generation are provided in Section 4.

**Task assignment.** With the task viewpoints generated, we start optimizing the scanning paths for all robots. We first construct a weighted graph by setting the task viewpoints and robot positions as nodes and their traveling paths as edges. To assign the generated tasks to the robots with minimum execution effort and time consumption, we then formulate the problem as a modified *Multi-Depot Multiple Traveling Salesman Problem* (MDMTSP) and search for the approximate optimal solution by an iterative optimization method. We determine the task distribution to the robots first. Next, the execution sequence of tasks for each robot is determined by solving a traditional traveling salesman problem (TSP). Examples of the constructed graph and the final assigned task sequences for all robots are shown in Figure 2(c). More details about the formulation and solution of the task assignment can be found in Section 5.

**Task execution.** Each robot maintains and moves along a path that connects a sequence of assigned task viewpoints to complete the tasks one by one, as shown in Figure 2(d). Meanwhile, the path keeps growing with new tasks assigned in each iteration. Note that there are two kinds of tasks associated with two scanning modes. When the robot is processing one specific type of task, its moving speed and scanning range should be automatically shifted to the corresponding values. In particular, the robot with *explorer* mode processes the exploration task by tilting the head up and down at a fixed height to scan forward when arriving at the task position. With the known areas expanded, objects are detected by *Mask R-CNN* [He et al. 2017] and the semantic information is stored in the reconstruction result. To enlarge the known area, robots keep scanning on the route. The intermediate scans along the task path are also utilized. For reconstruction tasks, the robot with *reconstructor* mode uses the camera holding in hand with a 7-DOF arm to scan the objects carefully and stably with a particular angle and distance.

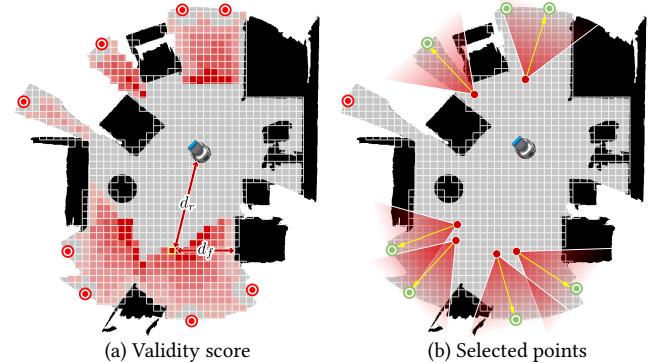


Fig. 3. Exploration task generation with viewpoint selection on 2D occupancy grid. (a) A set of frontiers (red circles) are first selected to determine the corresponding candidate viewpoints, and the validity score of each candidate viewpoint is defined by measuring how close it is to the nearest robot ( $d_r$ ) and how far from the nearest obstacle ( $d_f$ ). (b) A set of viewpoints (red dots) with high validity scores are selected with the constraint that each frontier can only be covered once. Then frontiers covered by the field of vision of the selected viewpoints are turned into green.

To achieve a high-quality reconstruction, the robot moves continuously and meticulously to ensure complete capture and sufficient overlap between adjacent scanning frames to reduce the error in object reconstruction. Figure 1 gives an illustration of these two different scanning modes when executing the corresponding tasks.

**Task scheduling.** Due to various reasons, including different task difficulties, it is often irrational to expect all robots to finish their tasks simultaneously. Therefore, a less time-consuming way is to let the robot automatically carry out new tasks after completing its assigned tasks in the last round. Under this strategy, we propose the *task-flow* model to accelerate the task scheduling process, as shown in Figure 2(e). Once a robot is about to finish its current tasks, it signals the control center to trigger the decision-making stage. After analyzing the geometry and semantics of the current scene, the control center starts a new iteration of task generation and assignment to append new tasks to the end of each robot’s task path. As a result, while all robots are processing the stored tasks, all the task paths keep updated and no robot needs to wait for other robots, which highly enhances the system’s efficiency. While the scanning procedure runs, tasks are generated from the control center and processed by robots like a flow. This procedure runs continuously until no latest task is updated, which indicates that the entire scanning task has been completed.

## 4 TASK GENERATION

### 4.1 Exploration task generation

Given  $R$  robots  $\{\mathcal{R}_r\}_{r=1}^R$  in an unknown environment, we represent each robot  $\mathcal{R}_r$  with a tuple  $\mathcal{R}_r = \{\mathbf{r}_r, \theta_r, I_r\}$ . The tuple contains the elements of the robot’s position and orientation on the 2D occupancy grid as well as the scanning mode. The exploration task is also defined on the 2D map and denoted as  $T_i^{exp} = (x_i, y_i, \theta_i)$ , encoding its position and orientation.

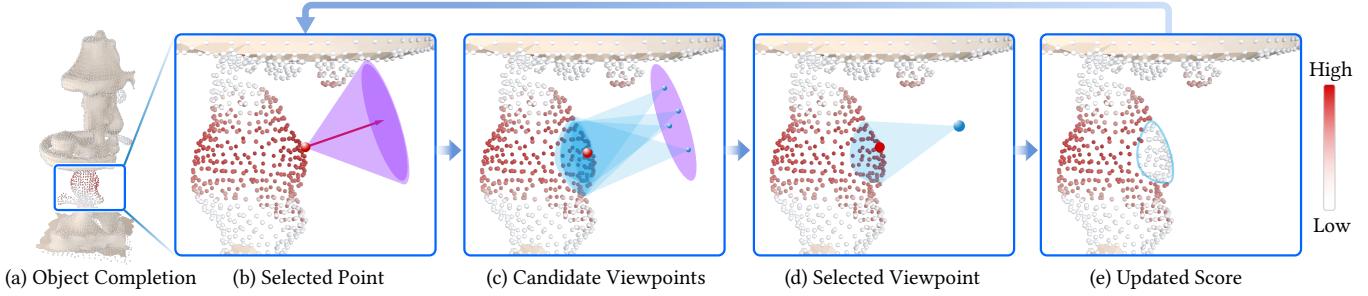


Fig. 4. Reconstruction task generation based on the completeness analysis on the 3D point cloud. (a) Object scan with the completed point cloud, and the color on the point indicates the corresponding incompleteness score. (b) The point with the highest incompleteness score is selected, and an optic cone is created for candidate viewpoints selection. (c) An optic cone is further created for each candidate viewpoint to compute its vision coverage. (d) The viewpoint with the highest vision coverage is selected. (e) The incompleteness score is updated by removing the region covered by the selected viewpoint.

To better explore the unknown area, we select the viewpoints for exploration tasks based on the frontiers of the current 2D occupancy grid, as shown in Figure 3. In detail, we first extract all frontiers via the Canny algorithm [Canny 1986] and then perform farthest sampling to get the set of uniformly distributed frontiers for further exploration, shown as the red points in Figure 3 (a). Then, to have a better vision for each frontier, we restrict the scanning distance to be within a predefined range and further exclude all the grids inside the region where the frontier point is not visible. As a result, only a subset of grids on the 2D map become candidate viewpoints. As later the robots need to move to the selected viewpoint to conduct the scan, we expect the viewpoint to be close to the nearest robot to minimize the moving effort as well as far from the obstacles for safety. So we further define a validity score on each candidate viewpoint as  $s = d_r - d_f$ , where  $d_r$  is the distance to the nearest robot and  $d_f$  is the distance to the nearest unfree grid, which are visualized using the colormap shown on the grids in Figure 3 (a).

With the candidate viewpoints with computed validity scores, we select the final viewpoints iteratively. Specifically, we first select the viewpoint with the highest validity score and then find the frontiers that can be covered by this viewpoint when setting the vision angle to be in the *explorer* mode. Then, the validity scores will be updated by setting the score of any candidate viewpoint associated with the covered frontiers to be zero to avoid further selection. The iteration continues by selecting the next viewpoint with the highest score until a pre-defined number of viewpoints have been selected. Figure 3 (b) shows the selected viewpoints and the frontiers covered by their visions turn green.

#### 4.2 Reconstruction task generation

When reconstructing objects, the viewpoint of the sensor need to be placed in 3D space to scan the objects from various views, as the reconstruction quality can be quite limited if the camera height is fixed. Therefore, the reconstruction task is denoted as  $T_i^{rec} = (x_i, y_i, z_i, \theta_i, \phi_i)$ , where  $\theta_i \in [0, 2\pi]$  is the orientation, and  $\phi_i \in [-\frac{\pi}{2}, \frac{\pi}{2}]$  is the elevator angle of the view point.

As the goal of the reconstruction task is to get high-quality reconstruction of the 3D objects, more scans should be conducted around the incomplete objects, especially toward regions with big

holes. Thus, we keep track of the object instances during the whole scanning process as in the work of *Voxblox++* [Grinvald et al. 2019] and analyze the completeness of each object for viewpoint selection. Note that by using a combined geometric-semantic segmentation scheme, the mapping framework *Voxblox++* is able to detect recognized elements from a set of known categories and simultaneously discover previously unseen objects in the scene. Thus, we can detect almost all the objects in the scene which have been scanned.

Given a reconstructed object with a unique instance label, we first uniformly sample  $N = 2048$  points  $\mathcal{P}$  on the object surface and use the GR-Net [Xie et al. 2020] to predict the corresponding complete point cloud  $C$ . Then, the incompleteness score of each point  $q \in C$  is defined as its nearest distance to the original point cloud  $\mathcal{P}$ :

$$S_c(q) = \min_{p \in \mathcal{P}} \|q - p\|_2, \quad (1)$$

which is then further normalized by the maximal distance among the set:  $\tilde{S}_c(q) = S_c(q) / \max_{q \in C} S_c(q)$ . It is worth noting that, the completion process is only used to search for the incomplete parts of the objects and will not be used directly as the reconstruction results, so the prediction doesn't have to be very precise.

To filter out the objects that have already been completely reconstructed, we compute the average incompleteness score  $\bar{S}_c(C)$  of all points belonging to the complete point cloud  $C$ , and set  $\tau = 0.2$  as a heuristic threshold that if  $\bar{S}_c(C)$  is less than  $\tau$ , we consider the object to be complete and no further scanning is needed. For the remaining incomplete objects, we rank them based on the  $\bar{S}_c(C)$  score and select the viewpoints one by one. For the selected incomplete object  $C$ , we iteratively select a set of viewpoints around the shape based on the incompleteness score defined on points to make full coverage of the incomplete region as shown in Figure 4.

In more detail, to search for viewpoints with large coverage of incomplete regions, the point  $q$  with the highest incompleteness score will be selected first to generate a set of candidate viewpoints  $\{v_i\}$  on the bottom surface of an optic cone with  $q$  as the apex and its normal direction  $n$  as the rotation axis, and then another optic cone with each reachable candidate viewpoint  $v_i$  pointing to  $q$  are created to calculate its coverage, as shown in Figure 4(b) and (c).

The view coverage of each viewpoint  $v_i$  is defined by the sum of incompleteness scores of all the points on  $C$  that are visible within

the optic cone of  $v_i$  as well as within the reachable range of the robot's arm. Note that the vision range of the optic cone of each candidate viewpoint is set to be the same as that of the *reconstructor* mode, while the vision range of the optic cone of  $q$  is doubled. Once the viewpoint with the highest view coverage is selected, the incompleteness scores will be updated for next viewpoint selection, as shown in Figure 4(c) and (d).

Note that, in practice, it often happens that the object cannot be completed due to the unreachable holes on the reconstruction surface and scanning only once cannot complete the object to an impressive level. Therefore, we set a historical record for all selected viewpoints of each recognized object. If an object has been scanned three times along neighboring viewpoints without increasing the surface completeness, no more reconstruction tasks will be generated inside the neighborhood of those viewpoints to avoid time-wasting in circling the "impossible missions".

## 5 TASK ASSIGNMENT

### 5.1 Problem formulation

Once the new exploration tasks  $\mathcal{T}^{exp}$  and reconstruction tasks  $\mathcal{T}^{rec}$  are generated, they need to be distributed to the robots for execution. Note that due to our asynchronous *task-flow* model, some of the robots may still have unfinished tasks assigned in the last round, so the newly assigned tasks should be appended to the end of the current paths. For robot  $\mathcal{R}_r$ , we denote its unfinished task sequence as  $\mathcal{T}_r^{rest}$  and the last unfinished task as  $T_r^{end}$ . To find the best task assignment, we first construct a weighted graph  $G = (\mathcal{V}, \mathcal{E})$  to encode the spatial relationship between the tasks and robots, where  $\mathcal{V} = \mathcal{T}^{exp} \cup \mathcal{T}^{rec} \cup \mathcal{T}^{end}$  consists of all new tasks and the set of final tasks of all the robots in last round  $\mathcal{T}^{end} = \{T_r^{end}\}_{r=1}^R$ , and  $\mathcal{E}$  consists of edges connecting each pair of nodes with distance calculated using A\* algorithm [Hart et al. 1968]. As the processing time of each task can be considered as fixed during the execution, we focus on optimization of the traveling cost during the task assignment and formulate it as a modified MDMTSP, where the goal is to find a set of disjoint paths  $\{\mathcal{T}_r\}_{r=1}^R$  that covers the whole set  $G$  such that the sum of all robots' tour costs is minimized:

$$E_d = \sum_{r=1}^R \left( \sum_{T_k \in \mathcal{T}_r} d(T_k, T_{k+1}) \right), \quad (2)$$

However, extra demands and constraints make our problem formulation slightly different. As our method focuses on multi-robot collaboration, other than minimizing the total tour cost, we also attempt to achieve load balance among robots to reduce the total time consumption of the assigned tasks. Thus, we dispatch the robots with uniform workload to decrease this time consumption by including the capacity term:

$$E_c = \sum_{r=1}^R \left( (|\mathcal{T}_r| + |\mathcal{T}_r^{rest}| - C_r)^2 \right), \quad (3)$$

where  $C_r = \sum_{r=1}^R (|\mathcal{T}_r| + |\mathcal{T}_r^{rest}|)/R$  is the total remaining tasks divided by the number of robots, standing for the average task capacity to each robot. As a result, our problem becomes a modified

---

### ALGORITHM 1: Updated MDMTSP problem solver

---

**Input:** Current weighted graph  $G$ ;  
Robot positions  $\mathcal{R}$ ;  
Exploration task positions  $\mathcal{T}^{exp}$ ;  
Reconstruction task positions  $\mathcal{T}^{rec}$ ;

**Output:** Path node sequence for each robot:

```

 $\mathcal{S} = \left\{ \dots, \{\mathcal{R}_r, \mathcal{T}_r^{rest}, T_{r1}, T_{r2}, \dots\}, \dots \right\}$ 
 $\gamma \leftarrow \text{InitialClustering}(\mathcal{R}, \mathcal{T}, G);$ 
 $\gamma \leftarrow \text{SimulatedAnnealing}(\gamma, G);$ 
for  $\gamma_i \in \gamma$  do
    |  $S_i = \text{TSPSolver}(\gamma_i, G)$ 
end
 $\mathcal{S} \leftarrow \{S_1, \dots, S_R\};$ 

```

---

MDMTSP with the goal to minimize the total energy cost:

$$\mathcal{T}^* = \arg \min_{\mathcal{T}=\{\mathcal{T}_r\}_{r=1}^R} E_d + E_c. \quad (4)$$

Moreover, different types of tasks are associated with different scanning modes, which means that the robot needs to switch the mode when starting a different kind of task, leading to extra costs. Besides, since reconstruction tasks aim to obtain fine-grained geometry of 3D objects, it is often preferable to have a sequence of consecutive careful scans. Therefore, to keep the continuity of high-quality scanning toward the object and reduce the times of mode shifting, we add an extra constraint to our objective function so that each task sequence  $\mathcal{T}_r$  only contains one type of task:

$$\prod_{T_k \in \mathcal{T}_r} \mathcal{I}(T_k) + \prod_{T_k \in \mathcal{T}_r} (1 - \mathcal{I}(T_k)) = 1.$$

where  $\mathcal{I}$  indicates the scanning mode corresponding to the task, and is set to be 1 for *explorer* and 0 for *reconstructor*.

### 5.2 Modified MDMTSP solver

It is proved that MDMTSP is NP-hard [Yadlapalli et al. 2009], and with extra constraints and demands in our formulation, it is almost impossible to find an exact optimal solution within a short time. To overcome the problem of computational complexity, we approximate the solution to this complex problem by splitting the model into three parts: assigning scanning modes to robots, clustering tasks for each robot, and planning task execution order inside each cluster. Since these three parts are inseparably interconnected, we propose an iterative approach to this chicken-and-egg situation.

The pseudocode of our solution is provided in Algorithm 1. The method starts with an initial task clustering based on a heuristic method. Specifically, we first compute the proportion of two types of viewpoints and randomly assign the scanning modes to robots according to this proportion with a guarantee that each scanning mode is assigned to at least one robot. Then we cluster the tasks using the k-means clustering algorithm with the robots' end positions  $\mathcal{T}^{end}$  as initial centroids and assign tasks to the robots with the same type and the least traveling cost. Given this initialization, the clustering is iteratively optimized through a simulated annealing algorithm. We use the method of cluster-and-assign similar to [Faigl

et al. 2012] in each iteration of simulated annealing, and in each annealing iteration, two types of disturbance can be chosen randomly: 1) *Exchange*: we separately select two robots from the explorers and the reconstructors and switch their scanning modes; 2) *Reassign*: we randomly change the scanning mode of a robot in the category containing more robots. Once the clusters are optimized, we deploy a traditional TSP solver to further optimize the execution sequence of tasks within each cluster. As a result, each robot receives a new task sequence and appends them to the current workload.

*Traveling distance term.* During the clustering optimization, it is required to calculate the traveling distance of each robot for finishing all the assigned tasks, which is computationally inefficient. Accordingly, we use the method similar to [Dong et al. 2019] to approximate the traveling distance to decrease the computation cost. More specifically, to estimate the distance from the robot to the assigned tasks, we sum up the traveling distance between the tasks and their centroid and the distance from the centroid to the robot's end position  $T_r^{end}$ . Note that different tasks are associated with different scanning modes at different speeds, affecting the assigned tasks' traveling time. So when considering the scanning speed of different tasks  $v(T)$ , we denote the approximate traveling energy as:

$$E_d' = \sum_{r=1}^R \left( \sum_{T_k \in \mathcal{T}_r} \frac{d(T_k, \omega_r)}{v(T_k)} + \frac{d(T_r^{end}, \omega_r)}{v(T_r^{end})} \right), \quad (5)$$

where  $\omega_r$  is the centroid of the tasks cluster assigned to robot  $R_r$ , and  $d(\cdot, \cdot)$  represents the traveling distance.

To combine the capacity term  $E_c$  defined in Equation 3 into the distance of the clustering method, 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. Unlike other clustering methods such as the k-means, it is more convenient for GMM to add terms like capacity uniformity to the optimization process. After each optimization step, we calculate the assignment cost of the new cluster and decide whether to update the states.

*Traveling energy constraint.* As the robot A shown in Figure 5, there are cases when the robot has no choice but to travel a long distance within one cluster to process tasks in the distance. In practice, a better choice is usually to stay and wait for the tasks generated nearby in the next iteration. The appearance of this inefficient situation is likely related to the fact that robots are ordered to process all generated tasks. However, there are also other situations where long-distance traveling is necessary when no possible tasks remain close to the robot. So to achieve a balance, we set a maximum travel distance between two consecutive tasks in a task sequence, and only check the task distance starting from a pre-defined index. Once we find that one task's travel distance is larger than the given threshold, we remove all its subsequent tasks together to facilitate the execution. Thus this robot is more likely to finish all its remaining tasks earlier than other robots to trigger the control center to assign more practical tasks. Experiments on this traveling energy constraint can be found in the supplementary material.

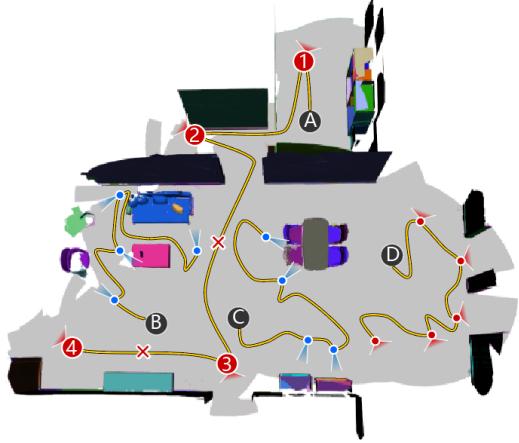


Fig. 5. Illustration of traveling energy constraint. When assigned with a task that has a long traveling distance, e.g., the task 3 for robot A in this case, the scanning efficiency can be significantly reduced, thus all the subsequent tasks of task 3, including task 4, are deleted together for robot A, and it will wait for the task assignment in next iteration.

## 6 RESULTS AND EVALUATION

### 6.1 System and dataset

*Robot system.* Our system runs on top of ROS, which supports the robot's standard behaviors, such as navigation and arm actions. For the convenience of development and evaluation, we use the Gazebo [Koenig and Howard 2004] framework as our simulator, which can simulate the interactions with real-world. We further add synthetic noise to the depth maps to simulate the camera noise in reality, as in the work of [Handa et al. 2014]. We use Fetch as our robots in the simulator. Each robot is equipped with a Primesense Carmine 1.09 mounted on the top of it, a RealSense Depth Camera D435i held in hand, as well as a SICK 2D sensor. The SICK is used for robot tracking, which is necessary for exploration tasks and map merging. A desktop PC processes all data with Intel(R) Core(TM) I7-10700K CPU (3.8GHz×8), 64GB RAM, and an Nvidia GeForce RTX3080 GPU. The GPU is used for shape completion (GR-Net) and instance segmentation (Mask R-CNN), which takes about 16GB RAM and 10GB GPU memory in total. The RGB-D data and cameras captured by all robots are utilized in our methods.

For experiments in reality, we use Turtlebots3 instead due to its lower price. Each robot is equipped with a RealSense Depth Camera D435i, as well as an RPLIDAR A2M8 360° LiDAR scanner and Nokov motion capture cameras (<https://www.nokov.com/>). When the robots move beyond the capture range of the Nokov, the LiDAR scanners can track the robots' locations as well.

*Scene dataset.* We select virtual scenes from two datasets: synthetic scenes from *Front3D* [Fu et al. 2021], which are composed of hand-made residential buildings delineated by clean boundaries and separations, and realistic scenes from *Matterport3D* [Chang et al. 2017], which are reconstructed from real scenes and contain more types of indoor scenes such as offices and hotels. The collection contains 40 scenes including residential buildings (30), offices (5),

restaurants (2), and other environments (3). All these indoor scenes can be regarded as *flat-layout* of walls and furniture, without stairs or sunken regions. To avoid wasting time in meaningless exploration beyond the scene, we seal off all windows and doors in the scenes. In our evaluation, we carefully select four representative benchmark scenes in each category of scales: small ( $< 230m^2$ ), medium (230–300  $m^2$ ), and large ( $> 300m^2$ ) in our dataset to frequently be used in most of the quantitative plots in this section.

## 6.2 Evaluation metrics

As our goal is to conduct automatic scene reconstruction with multi-robot collaboration, we evaluate the method in three aspects: reconstruction quality, execution efficiency, and load balance among multiple robots.

**Reconstruction quality.** We evaluate how complete and accurate the objects with detailed geometry are captured.

**Object Completeness (O-Comp)** is measured by the average percentage of the covered surface of all objects in the scene, and the completeness of the reconstructed point cloud  $O$  compared to the ground truth point cloud  $\mathcal{G}$  is defined as:

$$OC(O, \mathcal{G}) = \frac{1}{|\mathcal{G}|} \sum_{q \in \mathcal{G}} \mathbb{1}(d(q, O) \leq \tau), \quad (6)$$

where  $d(q, O)$  is the minimal distance of the point  $q$  to the point cloud  $O$ , and  $\tau$  is the predefined distance threshold.

**Object Accuracy (O-RMS)** is measured by the average distance of the reconstructed objects in the scene, where the distance between each reconstructed point cloud  $O$  and the ground truth point cloud  $\mathcal{G}$  is defined as:

$$OA(O, \mathcal{G}) = \frac{1}{|O|} \sum_{p \in O} \min_{q \in \mathcal{G}} \|p - q\|_2. \quad (7)$$

**Execution efficiency.** To measure scanning efficiency, we compute the **Distance Consumption (D-C)** for the total travel distance of all robots and the **Time Consumption (T-C)** for the total time consumed in the scanning procedure.

**Load balance.** To measure the **Distance Load Balance (D-LB)**, we use the method in the work of [Dong et al. 2019] and compute the ratio of the standard deviation and mean value corresponding to the moving distance of all robots. Moreover, we further compute the **Time Load Balance (T-LB)**, which is the percentage of the waiting time of all robots during the scanning process.

## 6.3 Comparison to baselines

**Baseline methods.** To the best of our knowledge, the work of [Dong et al. 2019] is the only work that shares the same goal as our work, so it's a natural baseline to compare with.

To further evaluate the effect of multi-robot cooperation compared to isolated robots, we also compare our method with the method proposed in [Liu et al. 2018], the state-of-the-art single-robot reconstruction work, used in two different settings. The first is the original single-robot setting, denoted as NBO  $\times 1$ , where evaluation metrics like time consumption and distance consumption of this baseline will be divided by the number of robots, 4 in our

Table 1. Comparison with the work of [Dong et al. 2019] on reconstruction quality of objects, reconstruction efficiency, and load balance.

Quality	Object Completeness			Object Accuracy		
	Small	Medium	Large	Small	Medium	Large
Dong	53.85	41.49	40.27	0.062	0.081	0.097
Ours	<b>66.18</b>	<b>72.49</b>	<b>70.03</b>	<b>0.035</b>	<b>0.039</b>	<b>0.033</b>
Efficiency	Time Consumption			Distance Consumption		
	Small	Medium	Large	Small	Medium	Large
Dong	19.3	23.0	28.8	<b>482.6</b>	<b>525.8</b>	<b>738.6</b>
Ours	<b>14.0</b>	<b>18.9</b>	<b>24.7</b>	536.1	620.7	848.5
Balance	Distance Load Balance			Time Load Balance		
	Small	Medium	Large	Small	Medium	Large
Dong	0.172	0.216	0.159	0.205	0.262	0.273
Ours	<b>0.151</b>	<b>0.200</b>	<b>0.157</b>	<b>0.062</b>	<b>0.082</b>	<b>0.093</b>



Fig. 6. Some of the intermediate results of [Dong et al. 2019] (top) comparing to ours (bottom).

experiment, for a fair comparison. The second setting, denoted as NBO  $\times 4$ , assumes that we have extra prior knowledge of the scene structure and can subdivide it into isolated sub-regions for different robots to explore individually. In this case, each robot maintains its own reconstruction result and occupancy map, and no message will be exchanged during the scanning. Moreover, all baselines are modified by using the Voxblox++ as the underlying reconstruction module as in our method for fair comparisons.

**Results.** Table 1 shows the comparisons of objects' reconstruction quality, reconstruction efficiency, and load balance between Dong and our method. The metrics show that our method surpasses the work of [Dong et al. 2019] in object completeness and accuracy, which justifies the use of reconstruction tasks associated with tailored scanning mode in our method. Figure 6 shows some visual comparisons on the reconstruction results. Due to the nature of their method [Dong et al. 2019], no more task viewpoints will be generated surrounding the incomplete objects, and thus these incomplete results remain until the end of the mission. We also compare the accuracy error between ours and [Dong et al. 2019] by calculating the relative difference per point to the ground truth  $\mathcal{G}$ .

$$AE(\mathbf{p}) = \min_{q \in \mathcal{D}} \|\mathbf{p} - q\|_2 - \min_{r \in \mathcal{O}} \|\mathbf{p} - r\|_2,$$

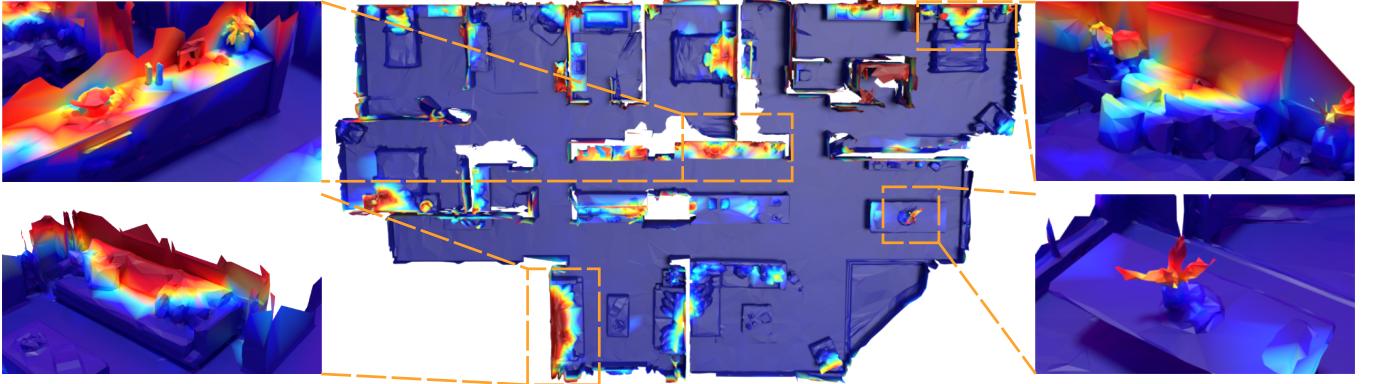


Fig. 7. Reconstruction error difference between the method of [Dong et al. 2019] and our method. Red indicates a higher reconstruction error of [Dong et al. 2019] compared to ours while blue indicates similar reconstruction quality.

where  $\mathcal{D}$  and  $O$  are the reconstruction results of [Dong et al. 2019] and ours. Since we pay more attention to the reconstruction of the objects, the reconstructed objects in our results have lower reconstruction error compared to [Dong et al. 2019]. Figure 7 shows the error map on one representative scene. We can see that the method of [Dong et al. 2019] has larger reconstruction errors on objects with more geometric details.

On the other hand, we can see that our method is better according to all metrics, except for the distance consumption. This is because that robots with *reconstructor* mode have to make some return trips in discovered areas around the incomplete objects to get better reconstruction results, which leads to an increase in the total traveling distance consumption. However, our method is still more efficient than the work of [Dong et al. 2019] even with longer traveling distance, reflected by our advantages in time consumption in the table. Moreover, the time load balance is significantly lower than that of [Dong et al. 2019]. The superiorities in time efficiency and load balances benefit from our asynchronous *task-flow* model, which makes sure that robots are occupied with tasks throughout the whole autoscanning process with no unemployed time. Though the robots process tasks with different moving speeds, our distance load balance is similar to [Dong et al. 2019], indicating that all robots process a similar workload in the whole scanning approach.

When comparing those two baselines based on the work of [Liu et al. 2018] with isolated robots, our method gets better performance in both time and distance consumption when having comparable reconstruction quality, as shown in Table 2. By using the same underlying method, the local robot paths are quite similar for baselines NBO  $\times 1$  and NBO  $\times 4$ , so we only show some visual comparison of our method to NBO  $\times 4$  in Figure 8. Note that multiple robots in our method can scan in various directions of the scene at a time, while for the NBO  $\times 4$  baseline, each robot can only scan each pre-assigned sub-region without collaboration with others. Moreover, since each robot only collects the local information of the scene, the robots spend a lot of time moving back and forth between different objects, which leads to inefficient performance.

It is also an interesting result that NBO $\times 4$  costs more time and distance than our baseline. The main reason is the underlying NBO

Table 2. Comparison with the work of [Liu et al. 2018] used in two different settings on reconstruction quality of objects and reconstruction efficiency.

Quality	Object Completeness			Object Accuracy		
	Small	Medium	Large	Small	Medium	Large
NBO $\times 1$	61.82	<b>72.91</b>	69.76	0.033	<b>0.037</b>	0.036
NBO $\times 4$	<b>67.52</b>	71.71	69.92	<b>0.031</b>	0.038	0.039
<b>Ours</b>	66.18	72.49	<b>70.03</b>	0.035	0.039	<b>0.033</b>

Efficiency	Time Consumption			Distance Consumption		
	Small	Medium	Large	Small	Medium	Large
NBO $\times 1$	21.8	33.1	47.3	648.8	781.2	979.7
NBO $\times 4$	27.8	35.1	53.8	715.3	827.8	1128.7
<b>Ours</b>	<b>14.0</b>	<b>18.9</b>	<b>24.7</b>	<b>536.1</b>	<b>620.7</b>	<b>848.5</b>

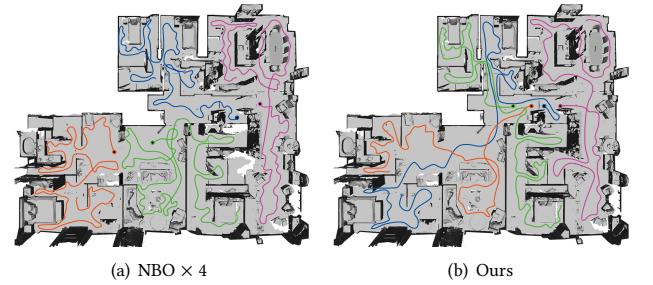


Fig. 8. Visual comparison between the result obtained using baseline NBO  $\times 4$  with multiple isolated robots and ours. The traveling paths of different robots are shown with lines in different colors.

method locates future tasks based on the objects glanced at during scanning, which leads to back-and-forth movement between different objects, while our method is able to jump out of local regions with the guidance of more global information collected from different robots.

Table 3. Ablation studies on different key components of our method.

Method	T-C	D-C	O-Comp	O-RMS	D-LB	T-LB
NoSw(3E1R)	19.6	698.8	58.91	0.0664	0.308	0.293
NoSw(2E2R)	23.0	708.8	60.65	0.0438	0.932	0.326
NoSw(1E3R)	27.6	676.7	59.89	0.0422	1.045	0.368
NoRe	<b>13.7</b>	731.3	59.51	0.0538	0.194	0.085
NoEx	24.0	584.1	58.97	0.0494	0.170	<b>0.073</b>
NoFlow	24.8	<b>536.1</b>	58.54	0.0542	0.189	0.186
<b>Ours</b>	19.3	668.4	<b>69.57</b>	<b>0.0360</b>	<b>0.169</b>	0.079

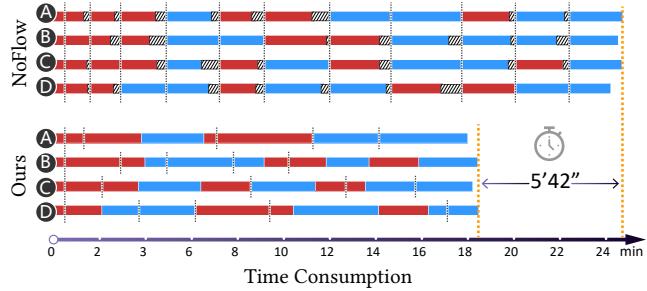


Fig. 9. Comparison of the task scheduling between two settings of our method with (Ours) and without the asynchronous *task-flow* model (NoFlow). All robots switch between the *explorer* mode (red) and *reconstructor* mode (blue), and each time when the control center is activated for new task generation and assignment is indicated by the black dotted line. Comparing to our method, the robots in NoFlow have idle time (shown in black slash region) to wait for others finishing all assigned tasks, which leads to significant time waste during the whole process.

#### 6.4 Ablation studies

*Ablation study on task assignment.* Compared to our dynamic task distribution, we freeze the robots' identities through all scanning periods. We divide four robots into every possible combination, denoted as **NoSw(3E1R)**, **NoSw(2E2R)** and **NoSw(1E3R)**, to mainly check the change in the efficiency of the modified method.

The first three rows in Table 3 show the results of the experiments, and the result of our method is presented in the last row. We can see that our full method beats all those variations with a salient advantage according to all the evaluation metrics.

With the decrease of the number of *explorers*, the time consumption of the modified method rises, although the distance consumption keeps roughly the same, which is mainly due to the speed difference between the *explorer* and *reconstructor* modes. With fewer *explorers* exploring the unknown region, few viewpoints for reconstruction can be generated in the limited discovered space and some of the *reconstructors* will keep waiting for new-discovered objects.

Moreover, the O-RMS decreases as the number of *explorers* decreases. Note that when there are fewer *explorers*, the completeness of the scanned objects is lower, so there are more reconstruction tasks generated for objects, which significantly increase the number of scans on objects and lead to lower O-RMS.

*Ablation study on task execution.* We freeze the scanning mode of each robot to be either *explorer* or *reconstructor* no matter which

type of tasks they are assigned, which results in two settings: 1) all explorer properties (**NoRe**), where all robots are equipped with the properties of high velocity and long scan range and only the camera on the head is utilized, and 2) all reconstructor properties (**NoEx**), where all robots have the properties of low speed and short scan range using the camera in hand to scan. The corresponding results are shown in the fourth and fifth rows in Table 3.

With all *reconstructor* modes in NoEx, the slow moving and scanning speed increase the overlapping of the scanning data, which significantly reduces the number of caves in the reconstruction result. At the same time, it is worth pointing out that the O-RMS doesn't become lower when only using reconstructors. This is because of the accumulation of reconstruction errors when the reconstructors have to travel between objects during the scanning with their short and narrow visions. This error accumulation phenomenon has also been discovered in the previous work [Xu et al. 2017]. However, the distance consumption is lower than our method as our method will have more tasks during the scanning process with more incomplete objects scanned with *explorer* mode. Nevertheless, due to the slow-moving speed, a significant amount of time is consumed during the scanning with all *reconstructor* modes compared to our method.

On the other hand, with all *explorer* modes in NoRe, the time consumption is much lower than our method due to the fast-moving speed of all robots, but this also leads to a considerable distance consumption. To summarize, our method effectively combines these two scanning modes and gets balanced performance in scanning efficiency and reconstruction quality.

*Ablation study on task scheduling.* We remove the *task-flow* model from our method, denoted as **NoFlow**, to test the effectiveness of the asynchronous task-scheduling module, where the calculation of the control center only triggers when all robots finish their task sequences assigned in the previous iteration. The comparison to our method can also be found in the last two rows in Table 3.

We can see that our method is better according to all metrics, except for distance consumption. This is because the new tasks generated overlap with some unfinished tasks due to the asynchronous module, which leads to more distance costs. But this small sacrifice of the distance gains the time-saving to a large level as no waiting is required during the scanning process, which results in generally more satisfactory results.

Figure 9 shows the visual comparison of the task scheduling results of four robots when scanning the same scene. In our setting, more robots act as explorers in the very beginning when few objects are seen in the scene. After finishing several exploring tasks, some incomplete objects are found to trigger more reconstruction tasks. Besides, when almost all exploration tasks were finished, all robots changed their scanning modes into reconstructors to complete the last reconstruction tasks. For the method of NoFlow, the robots can only get new tasks after all robots finish their tasks; thus, there is a lot of waiting time, represented by the black slash interval in the figure. Moreover, this comparison also shows that the mode shift happens to every robot and each robot handles different tasks evenly in our method, which further proves the effectiveness of various scan modes.



Fig. 10. Some visual results of our method on four virtual scenes (three from the Front3D data and one from the MatterPort3D dataset). The traveling path is represented in yellow lines with black contours, with the exploration tasks (red triangles) and reconstruction tasks (blue triangles) shown on the top of the path.

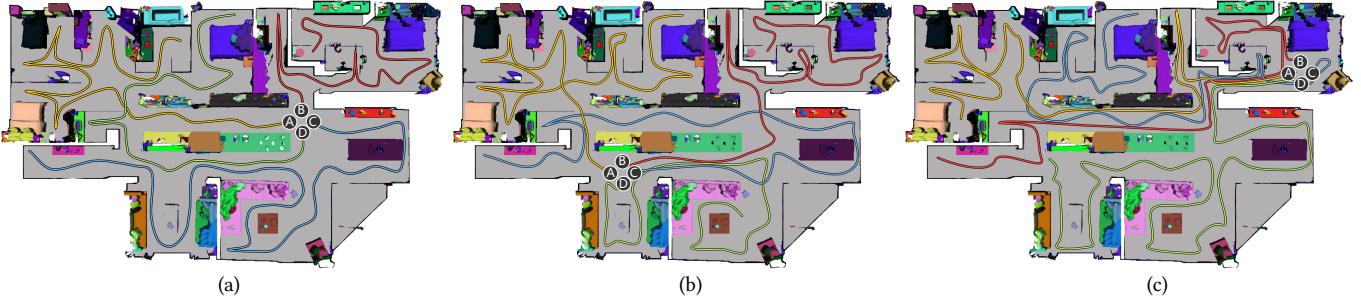


Fig. 11. Visualization of our results obtained with robots located at different initial points. The starting points are depicted with black circles, and the traveling paths of different robots are shown in different colors.

## 6.5 Qualitative results

*Virtual simulations.* Figure 10 shows the scanning paths and the semantic reconstruction results of our method for some exemplar indoor scenes. Different types of tasks are marked as viewpoints in different colors. We can see that our method can work well on indoor scenes with different scales and different structures. Different types of tasks are dynamically assigned to the robots for collaborative scanning during the whole process to maximize the exploration efficiency and reconstruction quality.

Figure 11 further shows the results on the same scene but with different initial robot locations. We observe that our method is quite robust to the initial locations, and the reconstruction quality and efficiency are comparable.

*Real robots tests.* We also test our method by scanning four unknown indoor scenes found around our department, including one rest area, one dorm, one area residence area, and one office. Figure 12 shows one example reconstruction result, and other results can be found in the supplementary material.

By using Turtlebot3 in the real world, the scan result cannot achieve the same quality as the results in the virtual scenes due to the low degree-of-freedom for viewpoint controlling. There are also many inevitable complex environmental factors in the real world, such as lens jitters when moving and the position deviation. Although the final reconstruction quality is relatively lower than the results obtained in the virtual simulations, our method still shows its capability to reconstruct the scenes in the real world.



Fig. 12. Example of the real-world scene reconstruction result with local details of the reconstruction result and corresponding picture taken from a similar viewpoint shown on the right.

## 7 CONCLUSIONS

We present a multi-robot system for scene exploration, understanding, and object reconstruction tasks in unknown indoor scenes. To improve the exploration efficiency and the reconstruction quality of objects, we adopt the following optimization: The robots can change their scan modes dynamically according to the assigned tasks. Hence, the robots can possess suitable properties to accomplish their designated tasks efficiently with high quality. The robots with *explorer* mode explore the unknown areas of the scene, while the robots with *reconstructor* mode process the object reconstruction tasks. Moreover, we formulate the task assignment problem with multiple task varieties as a modified MDMTSP and approximate the optimal solution within a short time. Furthermore, an asynchronous task-scheduling model is introduced to avert the periodic cycles in previous works to enhance the efficiency of the system. Finally, extensive experiments and comparisons are adopted to validate the feasibility and effectiveness of our algorithm.

**Limitations.** Our system for autoscanning suffers from several limitations. First, there are times the proposed optimization method approaches the local optimal solution of the MDMTSP problem, which can decrease the efficiency of the final assignment results. Second, our system can only be used in single-level indoor scenes, and the 2D occupancy grid cannot represent the complex scenes with compound structures. Third, our system does not support the non-distributed solution. When the scene is too large or complex for the control system to operate, robots may lose contact and stop receiving new tasks. In that case, our system may face a significant failure with the robots out of control. Besides, although we have prevented the robot from being scanned and reconstructed by others, the covering space of robots is not considered in the path planning stage. There are also some failure situations when multiple robots are going through a narrow intersection together; the obstacle avoidance algorithm freezes them and leads to the system deadlock.

**Future work.** There are several directions to improve our method in the future. First, task extraction can be updated with weights. Not all viewpoints are equally important; for example, the viewpoints

near a clean wall are less valuable than elaborate furniture. Therefore, it is feasible to significantly improve reconstruction efficiency by gathering more task viewpoints around the complex surfaces. Second, it will be an interesting future direction to generate the tasks using prior knowledge of the indoor scene's structure. The layout of the indoor scene where the furniture is distributed in the different houses generally shares an implicit relationship. Driving the robots towards places with furniture or more implicit unknown areas can improve scanning efficiency. Finally, more types of sub-task can be introduced to our system. For example, the object reconstruction task can be replaced by target object collection or target searching for other purposes.

## ACKNOWLEDGMENTS

We thank the anonymous reviewers for their valuable comments. This work was supported in parts by the National Natural Science Foundation of China (61872250, 62025207), Guangdong Natural Science Foundation (2021B1515020085), Shenzhen Science and Technology Program (RCYX20210609103121030), and Guangdong Laboratory of Artificial Intelligence and Digital Economy (SZ).

## REFERENCES

- Kamalova Albina and Suk Gyu Lee. 2019. Hybrid stochastic exploration using grey wolf optimizer and coordinated multi-robot exploration algorithms. *IEEE Access* 7 (2019), 14246–14255.
- Robin Amsters and Peter Slaets. 2019. Turtlebot 3 as a robotics education platform. In *International Conference on Robotics in Education (RiE)*. Springer, 170–181.
- Nikolay Atanasov, Jerome Le Ny, Kostas Daniilidis, and George J Pappas. 2015. Decentralized active information acquisition: Theory and application to multi-robot SLAM. In *2015 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 4775–4782.
- Subhrajit Bhattacharya, Robert Ghrist, and Vijay Kumar. 2014. Multi-robot coverage and exploration on Riemannian manifolds with boundaries. *The International Journal of Robotics Research* 33, 1 (2014), 113–137.
- Wolfram Burgard, Mark Moors, Cyrill Stachniss, and Frank E Schneider. 2005. Coordinated multi-robot exploration. *IEEE Transactions on robotics* 21, 3 (2005), 376–386.
- John Canny. 1986. A computational approach to edge detection. *IEEE Transactions on pattern analysis and machine intelligence* 6 (1986), 679–698.
- Angel Chang, Angela Dai, Thomas Funkhouser, Maciej Halber, Matthias Niessner, Manolis Savva, Shuran Song, Andy Zeng, and Yinda Zhang. 2017. Matterport3d: Learning from rgb-d data in indoor environments. *arXiv preprint arXiv:1709.06158* (2017).
- Benjamin Charrow, Gregory Kahn, Sachin Patil, Sikang Liu, Ken Goldberg, Pieter Abbeel, Nathan Michael, and Vijay Kumar. 2015. Information-Theoretic Planning with Trajectory Optimization for Dense 3D Mapping.. In *Robotics: Science and Systems*, Vol. 11. Rome, 3–12.
- Omar Cheikhrouhou and Ines Khoufi. 2021. A comprehensive survey on the Multiple Traveling Salesman Problem: Applications, approaches and taxonomy. *Computer Science Review* 40 (2021), 100369.
- Xinye Chen, Ping Zhang, Guanglong Du, and Fang Li. 2018. Ant colony optimization based memetic algorithm to solve bi-objective multiple traveling salesmen problem for multi-robot systems. *IEEE Access* 6 (2018), 21745–21757.
- 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.
- Thibaud Duhautbout, Julien Moras, and Julien Marzat. 2019. Distributed 3D TSDF manifold mapping for multi-robot systems. In *2019 European Conference on Mobile Robots (ECMR)*. IEEE, 1–8.
- Jan Faigl, Miroslav Kulich, and Libor Preucil. 2012. Goal assignment using distance cost in multi-robot exploration. In *2012 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 3741–3746.
- Kuan Fang, Alexander Toshev, Li Fei-Fei, and Silvio Savarese. 2019. Scene memory transformer for embodied agents in long-horizon tasks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 538–547.
- Huan Fu, Rongfei Jia, Lin Gao, Mingming Gong, Binjiang Zhao, Steve Maybank, and Dacheng Tao. 2021. 3d-future: 3d furniture shape with texture. *International Journal of Computer Vision* (2021), 1–25.

- 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>
- Li Guan. 2006. Sensor-based cooperative multi-robot 3d environment reconstruction. (2006).
- Ankur Handa, Thomas Whelan, John McDonald, and Andrew J Davison. 2014. A benchmark for RGB-D visual odometry, 3D reconstruction and SLAM. In *2014 IEEE international conference on Robotics and automation (ICRA)*. IEEE, 1524–1531.
- Peter E Hart, Nils J Nilsson, and Bertram Raphael. 1968. A formal basis for the heuristic determination of minimum cost paths. *IEEE transactions on Systems Science and Cybernetics* 4, 2 (1968), 100–107.
- Kaiming He, Georgia Gkioxari, Piotr Dollár, and Ross Girshick. 2017. Mask r-cnn. In *Proceedings of the IEEE international conference on computer vision*. 2961–2969.
- Lionel Heng, Alkis Gotovos, Andreas Krause, and Marc Pollefeys. 2015. Efficient visual exploration and coverage with a micro aerial vehicle in unknown environments. In *2015 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 1071–1078.
- Junyan Hu, Hanlin Niu, Joaquin Carrasco, Barry Lennox, and Farshad Arvin. 2020. Voronoi-based multi-robot autonomous exploration in unknown environments via deep reinforcement learning. *IEEE Transactions on Vehicular Technology* 69, 12 (2020), 14413–14423.
- Pengdi Huang, Liqiang Lin, Kai Xu, and Hui Huang. 2020. Autonomous Outdoor Scanning via Online Topological and Geometric Path Optimization. *IEEE Transactions on Intelligent Transportation Systems* (2020).
- Miguel Juliá, Arturo Gil, and Oscar Reinoso. 2012. A comparison of path planning strategies for autonomous exploration and mapping of unknown environments. *Autonomous Robots* 33, 4 (2012), 427–444.
- Ines Khoufi, Mohamed Haddad, Pascale Minet, and Anis Laouiti. 2015. Optimized trajectories of multi-robot deploying wireless sensor nodes. In *2015 IEEE 11th International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob)*. IEEE, 724–731.
- Nathan Koenig and Andrew Howard. 2004. Design and use paradigms for gazebo, an open-source multi-robot simulator. In *2004 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)(IEEE Cat. No. 04CH37566)*, Vol. 3. IEEE, 2149–2154.
- Michael Krainin, Brian Curless, and Dieter Fox. 2011. Autonomous generation of complete 3D object models using next best view manipulation planning. In *2011 IEEE International Conference on Robotics and Automation*. IEEE, 5031–5037.
- Shihui Li, Yi Wu, Xinyu Cui, Honghua Dong, Fei Fang, and Stuart Russell. 2019. Robust multi-agent reinforcement learning via minimax deep deterministic policy gradient. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 33, 4213–4220.
- 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.
- Yilin Liu, Ruqi Cui, Ke Xie, Minglun Gong, and Hui Huang. 2021. Aerial Path Planning for Online Real-Time Exploration and Offline High-Quality Reconstruction of Large-Scale Urban Scenes. *ACM Transactions on Graphics (Proceedings of SIGGRAPH ASIA)* 40, 6 (2021), 226:1–226:16.
- Fethi Matoui, Boumedeny Boussaid, Brahim Metoui, and Mohamed Naceur Abdelkrim. 2020. Contribution to the path planning of a multi-robot system: centralized architecture. *Intelligent Service Robotics* 13, 1 (2020), 147–158.
- Farzad Nirouzi, Kaicheng Zhang, Zendai Kashino, and Goldie Nejat. 2019. Deep reinforcement learning robot for search and rescue applications: Exploration in unknown cluttered environments. *IEEE Robotics and Automation Letters* 4, 2 (2019), 610–617.
- Morgan Quigley, Ken Conley, Brian Gerkey, Josh Faust, Tully Foote, Jeremy Leibs, Rob Wheeler, Andrew Y Ng, et al. 2009. ROS: an open-source Robot Operating System. In *ICRA workshop on open source software*, Vol. 3. Kobe, Japan, 5.
- Manikandasirram Srinivasan Ramanagopal and Jerome Le Ny. 2016. Motion planning strategies for autonomously mapping 3d structures. *arXiv preprint arXiv:1602.06667* (2016).
- Mike Roberts, Debadeeptha Dey, Anh Truong, Sudipta Sinha, Shital Shah, Ashish Kapoor, Pat Hanrahan, and Neel Joshi. 2017. Submodular trajectory optimization for aerial 3d scanning. In *Proceedings of the IEEE International Conference on Computer Vision*. 5324–5333.
- Guillaume Sartoretti, Yue Wu, William Paivine, TK Satish Kumar, Sven Koenig, and Howie Choset. 2019. Distributed reinforcement learning for multi-robot decentralized collective construction. In *Distributed autonomous robotic systems*. Springer, 35–49.
- Lukas Schmid, Michael Pantic, Raghav Khanna, Lionel Ott, Roland Siegwart, and Juan Nieto. 2020. An efficient sampling-based method for online informative path planning in unknown environments. *IEEE Robotics and Automation Letters* 5, 2 (2020), 1500–1507.
- Miguel Schneider-Fontan and Maja J Mataric. 1998. Territorial multi-robot task division. *IEEE Transactions on Robotics and Automation* 14, 5 (1998), 815–822.
- Maks Sorokin, Wenhao Yu, Sehoon Ha, and C Karen Liu. 2021. Learning human search behavior from egocentric visual inputs. In *Computer Graphics Forum*, Vol. 40. Wiley Online Library, 389–398.
- Kaarthik Sundar and Sivakumar Rathinam. 2017. Algorithms for heterogeneous, multiple depot, multiple unmanned vehicle path planning problems. *Journal of Intelligent & Robotic Systems* 88, 2 (2017), 513–526.
- Yulin Tian, Yun Chang, Fernando Herrera Arias, Carlos Nieto-Granda, Jonathan P How, and Luca Carlone. 2021. Kimera-multi: Robust, distributed, dense metric-semantic slam for multi-robot systems. *arXiv preprint arXiv:2106.14386* (2021).
- Sahar Trigui, Omar Cheikhrouhou, Anis Koubaa, Uthman Baroudi, and Habib Youssef. 2017. FL-MTSP: a fuzzy logic approach to solve the multi-objective multiple traveling salesman problem for multi-robot systems. *Soft Computing* 21, 24 (2017), 7351–7362.
- Masoumeh Vali and Khodakarami Salimifard. 2017. A constraint programming approach for solving multiple traveling salesmen problem. In *The Sixteenth International Workshop on Constraint Modelling and Reformulation*. 1–17.
- J Irving Vasquez-Gomez, L Enrique Sucar, Rafael Murrieta-Cid, and Efrain Lopez-Damian. 2014. Volumetric next-best-view planning for 3D object reconstruction with positioning error. *International Journal of Advanced Robotic Systems* 11, 10 (2014), 159.
- Melonee Wise, Michael Ferguson, Derek King, Eric Diehr, and David Dymeshich. 2016. Fetch and freight: Standard platforms for service robot applications. In *Workshop on autonomous mobile service robots*.
- Shihao Wu, Wei Sun, Pinxin Long, Hui Huang, Daniel Cohen-Or, Minglun Gong, Oliver Deussen, and Baoquan Chen. 2014. Quality-driven poisson-guided autoscaning. *ACM Transactions on Graphics* 33, 6 (2014).
- Haozhe Xie, Hongxun Yao, Shangchen Zhou, Jiageng Mao, Shengping Zhang, and Wenxiu Sun. 2020. Grnet: Gridding residual network for dense point cloud completion. In *European Conference on Computer Vision*. Springer, 365–381.
- Kai Xu, Hui Huang, Yifei Shi, Hao Li, Pinxin Long, Jianong Caichen, Wei Sun, and Baoquan Chen. 2015. Autoscaning for coupled scene reconstruction and proactive object analysis. *ACM Transactions on Graphics (TOG)* 34, 6 (2015), 1–14.
- Kai Xu, Yifei Shi, Lintao Zheng, Junyu Zhang, Min Liu, Hui Huang, Hao Su, Daniel Cohen-Or, and Baoquan Chen. 2016. 3D attention-driven depth acquisition for object identification. *ACM Transactions on Graphics (TOG)* 35, 6 (2016), 1–14.
- Kai Xu, Lintao Zheng, Zihao Yan, Guohang Yan, Eugene Zhang, Matthias Niessner, Oliver Deussen, Daniel Cohen-Or, and Hui Huang. 2017. Autonomous reconstruction of unknown indoor scenes guided by time-varying tensor fields. *ACM Transactions on Graphics (TOG)* 36, 6 (2017), 1–15.
- Sai Yadlapalli, Waqar A Malik, Swaroop Darbha, and Meir Pachter. 2009. A Lagrangian-based algorithm for a multiple depot, multiple traveling salesmen problem. *Nonlinear Analysis: Real World Applications* 10, 4 (2009), 1990–1999.
- Kai Ye, Siyan Dong, Qingnan Fan, He Wang, Li Yi, Fei Xia, Jue Wang, and Baoquan Chen. 2022. Multi-Robot Active Mapping via Neural Bipartite Graph Matching. *arXiv preprint arXiv:2203.16319* (2022).
- Shuai Yuan, Bradley Skinner, Shoudong Huang, and Dikai Liu. 2013. A new crossover approach for solving the multiple travelling salesmen problem using genetic algorithms. *European journal of operational research* 228, 1 (2013), 72–82.