FY2025/2025年度

# 3D Environmental Information Understanding Team

Asako Kanezaki

三次元環境情報理解チーム　金崎 朝子

Center for Advanced Intelligence Project

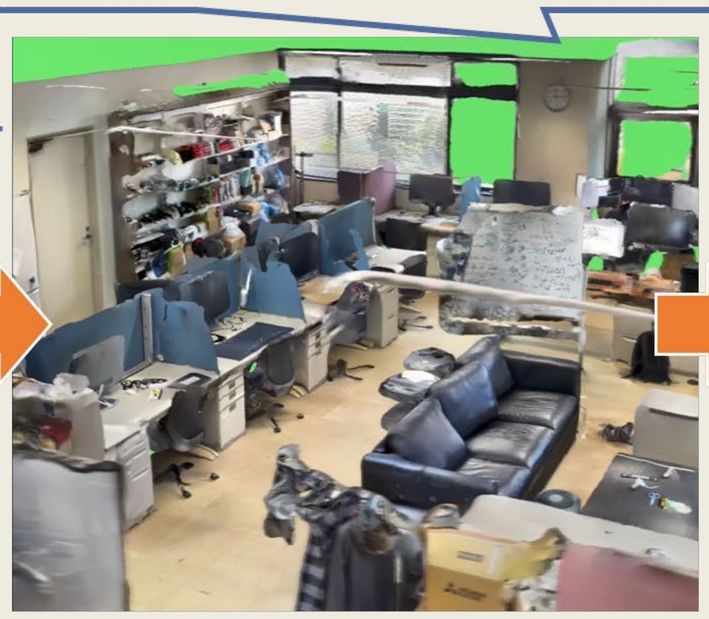
## 研究概要

視覚・聴覚・触覚等のマルチモーダルなセンサを持つ自律移動ロボットを用いて実世界の3D環境地図を作り、豊富な時空間情報を表現・活用する研究を行っている。三次元地図のセマンティクスをどのようにして取得し表現するか、その中からどういった情報に着目すべきか、また、ロボットがどう行動するべきかといった、アクティブセンシングの観点からのロボット研究にも同時に取り組んでいる。

マルチモーダル
セマンティック
3D環境地図作成

- 3Dシーングラフ（ SceneGraphFusion [Wu+, 2021] 等）と画像群を活用し
- ミクロ／マクロな時空間情報を考慮した説明文生成を如何にして生成するか
- 何に着目し、行動するか


Auto-mobile robot



3D Environment Map

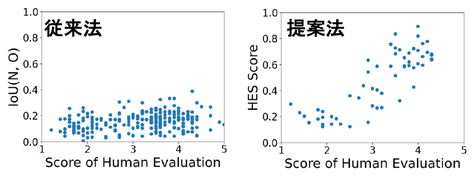
The building's environment is characterized by a mix of office and living spaces. The office areas are furnished with desks, chairs, and computer monitors, suggesting a professional setting. The desks are equipped with various office supplies, and the chairs are arranged in a way that promotes collaboration. The office spaces are well-lit, with natural light coming through the windows. In contrast, the living spaces are more relaxed and comfortable. The living areas feature comfortable couches and chairs, creating a cozy atmosphere. The walls are adorned with artwork and personal items, adding a touch of personality to the space. The living areas also have large windows, allowing for plenty of natural light to fill the room.

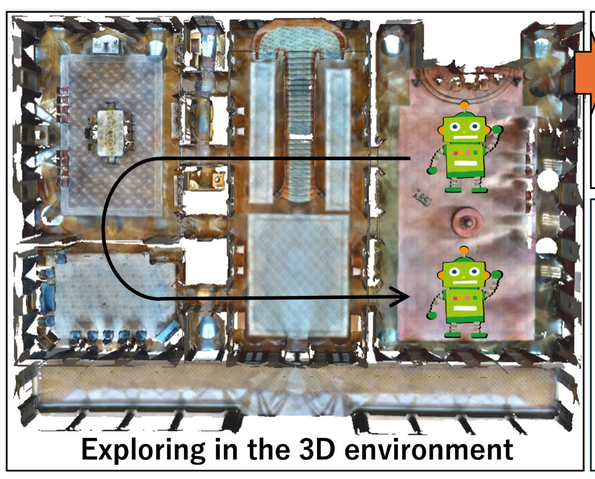
Description

## 2025年度の主な研究成果

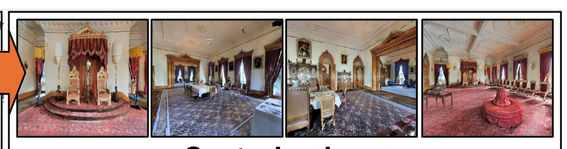
### 環境の説明文を生成するロボット探索タスク [1][5]

VLMを用いて環境の説明文生成
- 収集画像をVLMに入力し説明文を生成する。
- AMTでデータセットと評価値を収集・学習。



Exploring in the 3D environment



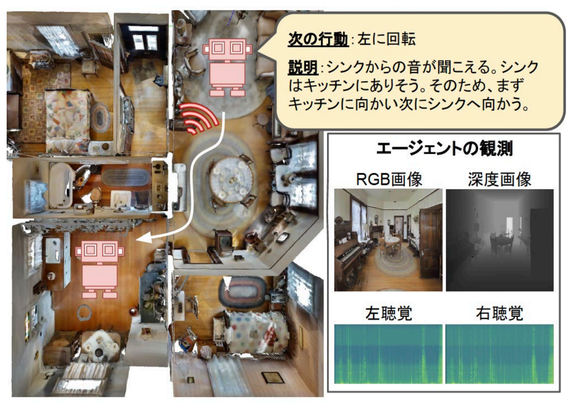
Capturing Images

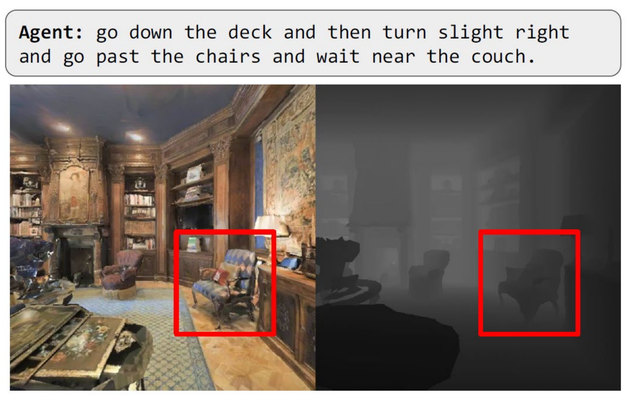
This environment is characterized by a luxurious and elegant living room or drawing room. It is elaborately decorated with large patterned rugs, heavy curtains, and classic furniture such as armchairs and sofas. Paintings and framed pictures are hung on the walls, creating a sophisticated atmosphere. There are also many rooms with other luxurious furnishings, and the overall style gives off a historical or classical atmosphere, perhaps reminiscent of a mansion or large estate. The lighting is soft, and natural light streams in through the large windows, creating a warm and pleasant atmosphere.

Description

### 説明文生成を補助タスクとするロボットナビゲーション [2]

- 強化学習を使った解釈可能なナビゲーション
- 説明文生成学習による常識の獲得によって、ナビゲーションの性能向上を図る　e.g.) シンクはキッチンにありそう




## その他：

- ◆ 動画生成を行うVideo Diffusion Models (VDMs)をロボット動画データセットで学習することで、初期状態の入力画像からロボットマニピュレーションの軌跡を出力する手法を開発し、動きの安定性を向上させるFlow lossを提案した[3]。
- ◆ マルチモーダルなセンサ情報を入力としたロボット行動学習において、各時間ステップにおける行動生成に最も有益なモダリティを特定し選択的に活用するcross-modality attention (CMA)を提案し、ロボット行動の教師無しセグメンテーションを実現した[4]。
- ◆ 自己教師あり学習によるカテゴリレベルでの関節付き3D物体姿勢推定手法[6]の発展研究として、一組のquery-referenceペアのRGBD画像のみを入力とした3D物体姿勢推定の自己教師あり学習手法を開発し、State-of-the-artな性能を達成した。
- ◆ イベントカメラを用いた可視光通信[7]を行うことで高精度な自己位置推定を行うマルチエージェントシステムを開発し、従来システムに比べて高精度な三次元再構成が可能であることを示した。

[1] Kohei Matsumoto and Asako Kanezaki. EED: Embodied Environment Description through Robotic Visual Exploration. *IEEE Robotics and Automation Letters (RA-L)*, 2026.
[2] Haru Kondoh and Asako Kanezaki. Embodied Navigation with Auxiliary Task of Action Description Prediction. *IEEE/CVF International Conference on Computer Vision (ICCV)*, 2025.
[3] Kuanting Wu, Kei Ota, and Asako Kanezaki. FlowLoss: Dynamic Flow-Conditioned Loss Strategy for Video Diffusion Models. *The 19th International Conference on Machine Vision Applications (MVA)*, 2025.
[4] Jiawei Jiang, Kei Ota, Devesh K. Jha, and Asako Kanezaki. Modality Selection and Skill Segmentation via Cross-Modality Attention. *The 19th International Conference on Machine Vision Applications (MVA)*, 2025.
[5] Kohei Matsumoto and Asako Kanezaki. EED: Embodied Environment Description through Robotic Visual Exploration. *The 6th Embodied AI workshop at CVPR*, 2025.
[6] Yuchen Che, Ryo Furukawa, and Asako Kanezaki. OP-Align: Object-level and Part-level Alignment for Self-supervised Category-level Articulated Object Pose Estimation. *The 18th European Conference on Computer Vision (ECCV)*, oral, 2024.
[7] Haruyuki Nakagawa, Yoshitaka Miyatani, and Asako Kanezaki. Linking Vision and Multi-Agent Communication through Visible Light Communication using Event Cameras. *Int. Joint Conf. on Autonomous Agents & Multiagent Systems (AAMAS)*, 2024.