A26-063

GAI與AI應用於居家服務型機器人

Applications of Generative AI and Artificial Intelligence in Home Service Robots



隊伍名稱｜早餐好了沒 Is Breakfast Ready ?

隊　　長｜鄭祺東 / 成功大學電機工程研究所

隊　　員｜江庭瑜 / 成功大學電機工程研究所

隊　　員｜陳宣蓉 / 成功大學電機工程研究所

隊　　員｜游翔丞 / 成功大學電機工程研究所

指導教授 ADVISOR



李祖聖 | 成功大學電機工程學系

成功大學電機工程博士，現為成功大學電機工程學系特聘教授。曾任中華民國自動控制學會理事長、臺灣機器人學會理事長、聯合大學電機資訊學院院長。

研究領域

人工智慧與生物智慧研發與應用、人形機器人、AI服務型機器人、模糊系統控制、機電整合系統與四輪獨立轉向與驅動載具。

作品摘要

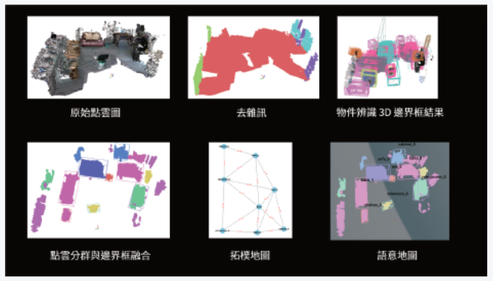
本作品提出一套具備高度語意理解與動態操作能力的智慧服務型機器人系統。該系統整合了大型語言模型(LLM)、視覺語言模型(VLM)、針對已知環境的SLAM建圖與拓撲場景圖，以及針對未知環境的視覺語言導航(VLN)技術，大幅提升了機器人在居家或複雜環境中執行日常任務的自主性與準確度。系統的主要功能涵蓋：居家物品的雙臂精準取放與連續動作協作、移動底盤的精準導航與避障，以及環境的探索與互動。

首先，系統內建的語音模組能準確分辨使用者的語意是一般閒聊，還是實際的任務指令。接收到語音指令後，該模組會處理輸入訊號、給予語音回饋，並根據使用者的需求觸發後續的任務排程。

在雙臂協作方面，系統會將接收到的任務指令結合視覺辨識模組的資訊，進行雙臂的任務排程。這使得決策能真實對應當下的物理狀態，進而規劃出具備環境感知能力的協同動作序列，配合系統設計了一套動作序列邏輯檢查與排序演算法，確保雙臂協作排程在物理空間與邏輯時序上的絕對可行性。當面臨多重任務指令、複雜目標或突發狀況時，系統採用了LG-DABC演算法，能生成兼顧語意邏輯、物理安全與執行效率的全局最佳化動作流。

在導航方面，系統具備雙軌機制：在已知環境中，機器人會透過混合式度量拓撲地圖，自主進行空間關聯推理與導航；若處於未知環境或需尋找動態目標時，則會啟動基於開放詞彙的視覺語言導航(VLN)模組。該模組會先透過LLM理解使用者意圖，再利用VLM即時確認目標狀態，引導機器人自主靠近並完成任務。

總結來說，本系統不僅解決了服務型機器人在自然語言理解以及於已知與未知環境中導航的挑戰，更結合最佳化演算法突破了複雜雙臂協作與狀態感知的技術瓶頸，為自動化服務提供了一個高度智慧化且容錯率極高的解決方案。



圖一 建立語意與拓撲地圖之視覺化。

Abstract

This work proposes an intelligent service robot system equipped with high-level semantic understanding and dynamic manipulation capabilities. The system integrates Large Language Models (LLMs), Vision-Language Models (VLMs), SLAM mapping, and topological scene graphs for known environments, and Vision-and-Language Navigation (VLN) technology for unknown environments, significantly enhancing the robot's autonomy and accuracy when executing daily tasks in domestic or complex environments. The core functionalities of the system include precise bimanual pick-and-place with sequential motion collaboration for household objects, accurate mobile base navigation and obstacle avoidance, and environmental exploration and interaction.

First, the built-in voice module can accurately distinguish whether the user's semantic intent is casual chat or an actual task command. Upon receiving a voice command, the module processes the input signal, provides voice feedback, and triggers subsequent task scheduling based on the user's needs.

Regarding bimanual collaboration, the system schedules dual-arm tasks by combining the received task commands with information from the visual recognition module. This ensures that its decision-making accurately reflects the real-time physical state, thereby enabling planning of a collaborative motion sequence with environmental awareness. Complemented by a system-designed motion sequence logic checking and sorting algorithm, the absolute feasibility of bimanual task scheduling in both physical space and logical timing is ensured. When facing multiple task commands, complex goals, or unexpected situations, the system employs the LG-DABC algorithm to generate a globally optimized workflow that balances semantic logic, physical safety, and execution efficiency.

In terms of navigation, the system features a dual-track mechanism: in known environments, the robot autonomously performs spatial relational reasoning and navigation using a hybrid metric-topological map; in unknown environments or when searching for dynamic targets, it activates an open-vocabulary VLN module. This module first leverages an LLM to understand the user's intent, then uses a VLM to verify the target's status in real time, guiding the robot to approach the target and complete the task autonomously.

In conclusion, this system not only overcomes the challenges faced by service robots in natural language understanding and navigation across both known and unknown environments but also addresses the technical bottlenecks of complex bimanual collaboration and state perception by integrating optimization algorithms. Ultimately, it provides a highly intelligent, fault-tolerant, and automated service solution.

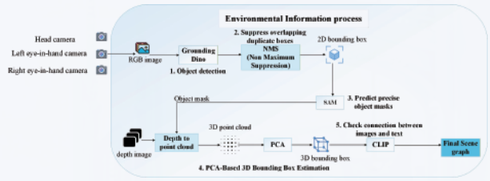


Fig. 2 Visual localization module.

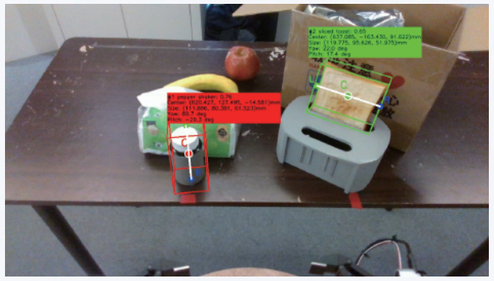


Fig. 3 Final detection result generated by the visual recognition module.