

無人飲料站： 具看、聽、說功能之雙手臂機器人

謝卓熹
李彥平
薛巧妍

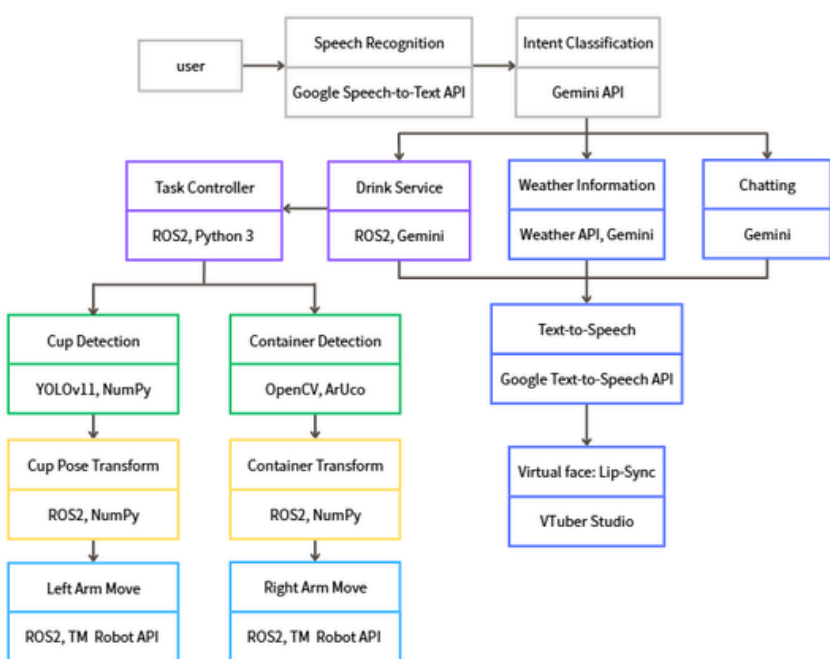
資訊二級

指導教授：連震杰

Introduction

系統核心串接 Gemini API，賦予機器人強大的自然語言處理能力，能與使用者進行流暢對話並精準解析複雜指令。在感測與執行層面，本系統整合視覺偵測技術與空間座標轉換演算法，讓機器人具備精準的空間感知與眼手協調能力；並透過先進的雙手臂控制技术實現高效的雙臂協同作業，打造出具備高互動性、高精準度的全自動化飲品服務系統。

System Architecture



System Setup

感測器: RGB-D深度攝影機 RealSense D415

機器手臂(雙手
臂): 夾取、給予
使用者物品



自走車 (AMR) : 移動至目標地點

Experimental Results & Performance

$$\begin{bmatrix} R_{11} & R_{12} & R_{13} & X \\ R_{21} & R_{22} & R_{23} & Y \\ R_{31} & R_{32} & R_{33} & Z \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

手眼校正

(相機和夾抓的位置關係)

$T_{cam}^{base} = T_{gripper}^{base} \times T_{offset} \times T_{cam}^{gripper}$

$T_{cam_to_gripper}$:
相機相對於夾抓的靜態校正
矩陣 (透過手眼校正工具事
先算好)。

$T_{gripper_to_base}$:
目前夾抓相對於機器人基座
的即時姿態 (由機械手臂的
編碼器/即時反饋取得)。



YOLOv11人像偵測進入指定距離

or
聽到指定喚醒詞(小美)

客人提出問題(Google Speech-to-Text)

Gemini API 思考問題

Google Cloud Text-to-Speech +
Lip-Sync + VTuber Studio 顯示人像(小美)
詢問喝甚麼飲料, 回答今天天氣等問題

針孔相機逆投影公式(2D → 3D)
Cup Pose Transformation

2D 座標

用YOLOv11-OBb得到

$$\begin{bmatrix} X_{base} \\ Y_{base} \\ Z_{base} \\ 1.0 \end{bmatrix} = \begin{bmatrix} R_{11} & R_{12} & R_{13} & T_x \\ R_{21} & R_{22} & R_{23} & T_y \\ R_{31} & R_{32} & R_{33} & T_z \\ 0 & 0 & 0 & 1 \end{bmatrix}_{4 \times 4} \times \begin{bmatrix} x \\ y \\ depth \\ 1.0 \end{bmatrix}_{4 \times 1}$$

$$X_c = \frac{(u - c_x) \cdot depth}{f_x}$$

$$Y_c = \frac{(v - c_y) \cdot depth}{f_y}$$

$$Z_c = depth$$

3D 座標
(x,y,z)

基座 (Base) → 夾爪 (Gripper)
→ 相機 (Camera)



測試並記錄不同水位下,
每一次倒水的笛卡爾空間座標
(x, y, z, rx, ry, rz)

每倒一杯飲料都記錄杯數,
再執行移動到指定杯數的座
標進行倒飲料

Cup(x, y, z)