



UNIVERSITAT
POLITÈCNICA
DE VALÈNCIA

Escola Tècnica
Superior d'Enginyeria
Informàtica



etsinf

Escola Tècnica Superior d'Enginyeria Informàtica

Universitat Politècnica de València

11 de junio de 2025

MEMORIA DEL PROYECTO

PICK BY COUNT

VISIÓN 3D | GRUPO PL1 – 5

Grado en Informática Industrial y Robótica

Autores/Autoras:

Marcos Belda Martínez | mbelmar@etsinf.upv.es

Ángela Espert Cornejo | aespcor@etsinf.upv.es

Lourdes Francés Llimerá | lfralli@epsa.upv.es

Tutor:

Antonio José Sánchez Salmerón | asanchez@isa.upv.es

1. PROYECTO REALIZADO

1.1 DESCRIPCIÓN DEL PROYECTO

¿EN QUÉ CONSISTE?

El proyecto propuesto consiste en la creación de un sistema avanzado de visión 3D que, mediante el uso de visión estéreo, permite detectar objetos y calcular sus posiciones mediante triangulación. La propuesta se centra en el reconocimiento y análisis de una mano, de modo que a través de gestos se pueda controlar el sistema. Este, a su vez, se integra con un gemelo digital del entorno, simulado en *RoboDK*, donde se representa un brazo robótico realizando tareas de *picking*. La cantidad de objetos a recoger será determinada por el gesto realizado con la mano y comunicada a la simulación vía MQTT.

1.2 ELEMENTOS PRINCIPALES

El sistema se estructura en varias fases funcionales que trabajan de forma modular y secuencial.

ADQUISICIÓN Y CALIBRACIÓN DE IMÁGENES ESTÉREO

Se emplean dos cámaras calibradas para capturar imágenes simultáneas de la escena. La calibración se realiza mediante la detección de las esquinas de un tablero de ajedrez, lo que permite calcular los parámetros intrínsecos y extrínsecos de las cámaras. Los resultados obtenidos se almacenan en archivos *.npz* para su reutilización en ejecuciones posteriores.

INTERFAZ, DETECCIÓN DE OBJETO Y TRIANGULACIÓN

Una vez obtenidas las imágenes estéreo, se procede a ejecutar el programa principal, el cual gestiona la interacción con el usuario, detección de objeto, obtención del centroide del objeto y triangulación del centroide para la obtención de la posición 3D.

CONTROL POR GESTOS CON MEDIAPIPE

Se emplea la librería *MediaPipe* para el reconocimiento de gestos de la mano, de modo que cada dedo levantado se interpreta como un comando numérico, siendo el resultado a enviar la suma de todos ellos. Por ejemplo, si se levantan 2 dedos, el sistema interpreta que se debe realizar un *picking* de dos objetos.

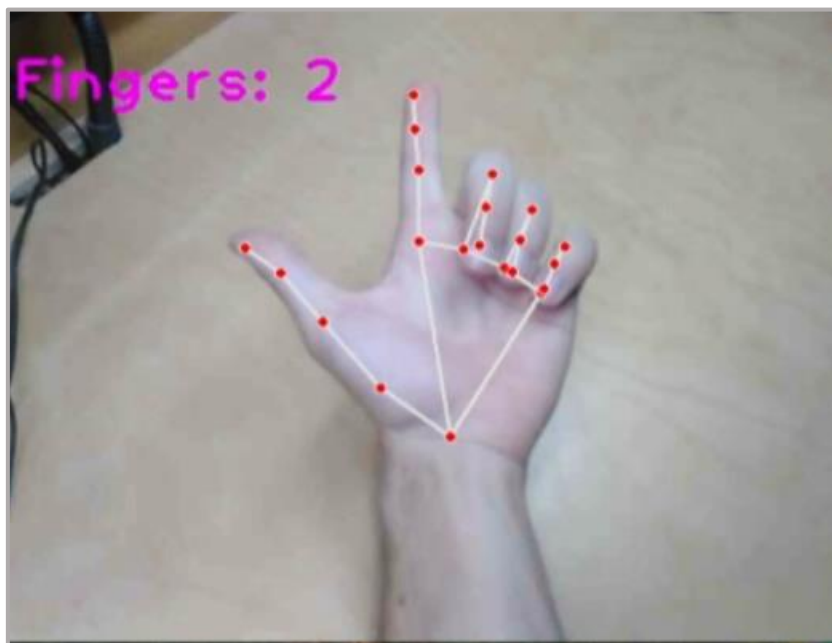
COMUNICACIÓN CON ROBODK VÍA MQTT

Se establece un canal de comunicación bidireccional entre el sistema de visión 3D y el entorno de simulación *RoboDK* mediante el protocolo *MQTT*. El sistema envía el número indicado con la mano a la simulación para que simule el *picking* de la cantidad de objetos indicados. Asimismo, el sistema recibe un booleano confirmando si la acción ha sido completada para finalizar el programa.

1.3 RESULTADO FINAL

Para demostrar la funcionalidad de la propuesta, se muestran a continuación ciertas imágenes que representan tanto el entorno real como el virtual.

ENTORNO REAL DETECTANDO LA MANO Y EL OBJETO



Código 1: Detección de la mano.



Código 2: Detección del objeto de interés.

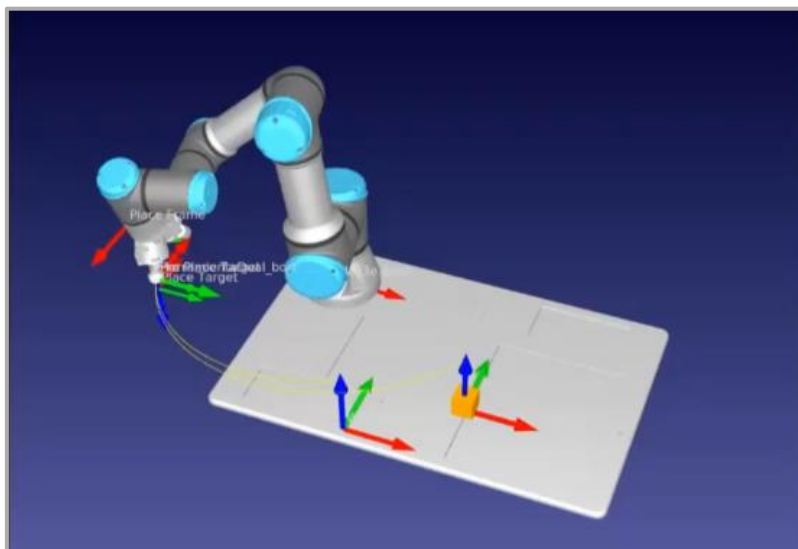


Figura 3: Entorno Virtual en RoboDK.

2. REPARTICIÓN DE TAREAS

BÚSQUEDA DE MATERIAL	Ángela y Lourdes
CALIBRACIÓN DE LAS CÁMARAS	Lourdes
CÁLCULO DEL CENTROIDE	Lourdes
TRIANGULACIÓN	Lourdes
DETECCIÓN Y CONTROL DE GESTOS	Ángela
CREACIÓN SIMULACIÓN ROBODK	Marcos
COMUNICACIÓN MQTT	Ángela y Lourdes
PRUEBAS DE FUNCIONAMIENTO	Marcos
UNIFICACIÓN DEL CÓDIGO	Ángela Marcos y Lourdes
MEMORIA (WORD)	Ángela
PRESENTACIÓN (POWERPOINT)	Ángela

3. BIBLIOGRAFÍA

- Chat GPT y Copilot.
- Trabajo Miguel Amador Alfaro: para la calibración.