



TREBALL FINAL DE GRAU

GRAU EN ENGINYERIA BIOMÈDICA

**Desenvolupament d'un sistema de control per veu per a un braç robòtic en entorns quirúrgics: Aplicació en la simulació de tasques assistencials**

AUTOR: NIL LESLIE BOKESA SALOMÓN

TUTOR: XAVIER CUFÍ SOLÉ

CO-TUTOR: STEVE HERNÁNDEZ UPTEGROVE

CONVOCATÒRIA: SEPTEMBRE 2025

DEPARTAMENT: Arquitectura i tecnologia de computadors

## 1. INTRODUCCIÓ

La robòtica mèdica és una disciplina emergent amb una gran capacitat de transformació dins l'àmbit sanitari. Aquest treball es centra en el desenvolupament d'un sistema robotitzat d'assistència quirúrgica, especialment orientat a la tasca d'infermeria instrumentalista. L'objectiu és explorar com un robot pot complementar aquestes funcions, aportant eficiència, precisió i repetibilitat en entorns quirúrgics.

## 2. OBJECTE DEL PROJECTE

L'objectiu principal d'aquest Treball de Fi de Grau és desenvolupar un sistema robòtic col·laboratiu capaç d'assistir en el subministrament d'instruments quirúrgics durant una intervenció simulada, mitjançant comandes verbals emeses pel personal mèdic. El sistema combina reconeixement d'objectes, reconeixement de veu i manipulació robòtica per identificar, localitzar i lliurar els estris sol·licitats de manera autònoma i eficient.

### 2.1. OBJECTIUS ESPECÍFICS

- Implementar un sistema de visió per computador que reconegui correctament els estris quirúrgics col·locats sobre una superfície plana.
- Desenvolupar un mòdul de reconeixement de veu capaç d'identificar ordres verbals associades als diferents instruments.
- Coordinar el moviment d'un braç robòtic UR3 perquè localitzi i manipuli l'estri sol·licitat, lliurant-lo a una zona d'entrega específica.
- Integrar totes les funcionalitats en un entorn de prova estable per validar el correcte funcionament del conjunt.

Aquest projecte també té una clara voluntat didàctica, ja que ofereix una plataforma útil per a la formació i experimentació en robòtica mèdica en entorns simulats.

## 3. DESENVOLUPAMENT TÈCNIC

Aquest projecte es basa en la integració de tres tecnologies fonamentals: la visió per computador, el reconeixement de veu i el control robòtic. L'objectiu és construir un sistema funcional que permeti a un robot UR3e identificar i manipular estris quirúrgics a partir d'ordres verbals, en un entorn simulat de quiròfan.

La solució s'ha estructurat en diversos mòduls independents però interconnectats que treballen de manera coordinada en temps real. L'arquitectura general del sistema inclou els següents components principals:

- **Mòdul de visió per computador:** encarregat d'identificar els estris sobre la taula mitjançant càmeres RGB.
- **Mòdul de reconeixement de veu:** processa les comandes verbals i les tradueix en instruccions interpretables pel sistema.
- **Conversió de coordenades:** transforma les ubicacions dels objectes del sistema de visió (píxels) a coordenades físiques reals.
- **Sistema de comunicació (MODBUS TCP):** permet la transmissió de dades entre el sistema de processament i el braç robòtic.
- **Control del robot UR3e:** mitjançant l'entorn Polyscope, que rep les coordenades i executa els moviments necessaris.

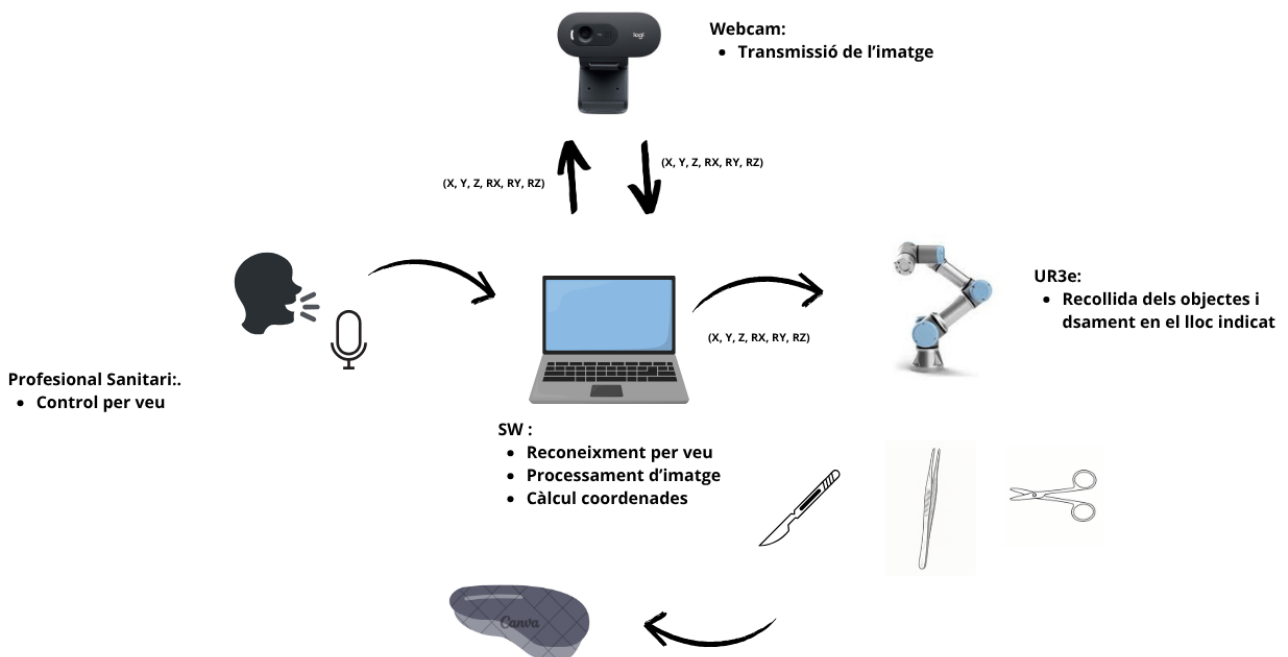


Figura 1: Flux operatiu del sistema

### 3.2. VISIÓ PER COMPUTADOR I RECONeixEMENT D'OBJECTES

Per tal de detectar i reconèixer els instruments quirúrgics, s'ha implementat un sistema de visió per computador mitjançant OpenCV. La càmera col·locada en vista zenital capta imatges en temps real de la superfície de treball, on es col·loquen els estris.

- **Captura i pre-processament d'imatge**, mitjançant una càmera RGB situada en posició zenital. Les imatges s'adapten en mida i format per ser compatibles amb el model d'inferència.
- **Inferència amb YOLOv8**, entrenat prèviament amb un dataset d'imatges d'estrís quirúrgics. El model retorna les etiquetes dels objectes detectats, juntament amb les seves coordenades en el pla d'imatge (bounding boxes).
- **Conversió de coordenades**, transformant les posicions de la imatge (en píxels) a coordenades físiques reals per

Aquesta metodologia proporciona una detecció robusta, tolerant a canvis d'il·luminació, i orientació.

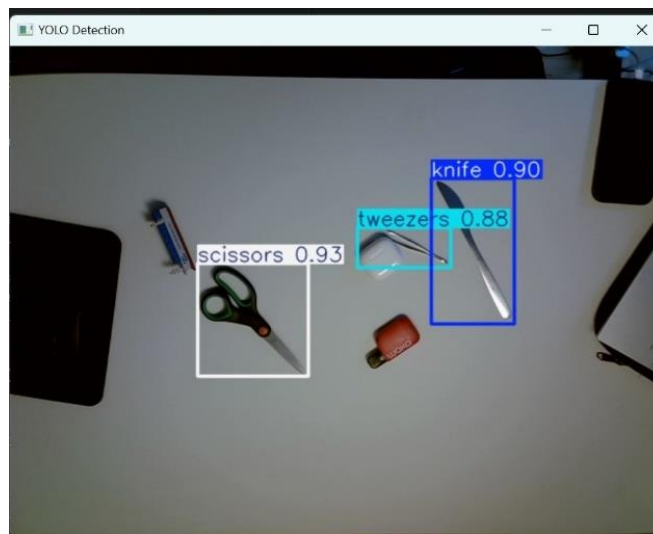


Figura 2: Bounding boxes de YOLOv8

### 3.3. Reconeixement de veu

El reconeixement de veu es basa en l'ús de biblioteques com SpeechRecognition i motors com Google *Speech API* per captar i transcriure comandes verbals. El sistema escolta contínuament fins que detecta una paraula clau associada a un estri quirúrgic.

Un cop transcrita la comanda, es compara amb les etiquetes disponibles mitjançant tècniques de *fuzzy matching*, per tolerar variacions en la pronunciació o errors menors.

Aquest sistema permet una interacció natural i sense contacte amb el robot, afavorint la seva aplicació en entorns quirúrgics simulats.

### 3.4. Comunicació amb el robot a través de *MODBUS TCP*

Per transmetre les coordenades i ordres entre el sistema de processament (ordinador central) i el robot UR3e, s'ha fet servir el protocol *MODBUS TCP*, que garanteix una comunicació robusta i en temps real.

El sistema envia dades en forma de registres (valors numèrics) que representen les coordenades X, Y i Z, així com comandes de control per iniciar o aturar el moviment.

Aquest enfocament ofereix una solució estàndard, fàcilment integrable i segura, que evita la necessitat de sistemes propietaris o protocols complexos.

### 3.5. Control del UR3e mitjançant Polyscope

El control del braç UR3e es realitza mitjançant un programa desenvolupat a l'entorn gràfic *Polyscope*, propi d'Universal Robots.

Un cop rebudes les dades, el robot:

1. Inicialitza i assigna variables a partir dels registres MODBUS (coordenades X, Y, Z i signes de direcció).
2. Reescale i aplica el signe corresponent a cada coordenada per reconstruir la posició espacial real de l'estri sol·licitat.
3. Calcula el punt de destí (p\_tool) i es desplaça cap a ell mitjançant moviments de tipus MoveL i MoveJ.
4. Simula l'acció de recollir l'instrument (amb pinça o buit), el trasllada fins a la zona d'entrega i torna a la posició inicial (home).

Aquest codi funciona de manera autònoma dins del controlador del robot, i es manté en espera de noves instruccions mentre el sistema central envia dades noves via MODBUS.

## 4. RESULTATS

Per validar el funcionament del sistema desenvolupat, es van dur a terme diverses proves en un entorn controlat que simulava un quiròfan.

Els resultats més rellevants obtinguts són els següents:

- **Precisió del reconeixement d'objectes:** El sistema basat en YOLOv8 va mostrar una alta fiabilitat en la detecció i classificació d'estrís, en condicions d'il·luminació estàndard.
- **Resposta a comandes verbals:** El sistema de reconeixement de veu va ser capaç d'interpretar correctament la majoria de comandes, incloent variants lèxiques simples. Es van detectar algunes dificultats puntuals amb soroll de fons o pronúncies poc clares.
- **Coordinació entre mòduls:** La integració entre els mòduls de veu, visió i control robòtic va funcionar de forma fluida, amb latències acceptables i sense pèrdues de sincronització.

## 5. CONCLUSIONS

Les principals conclusions extretes del treball són:

- És possible reconèixer i localitzar estrís quirúrgics amb precisió mitjançant models CNN com YOLOv8, fins i tot amb orientacions variables.
- La interacció verbal amb el sistema facilita una experiència natural i sense contacte, afavorint entorns clínics segurs.
- El robot UR3e pot manipular de manera fiable els estrís detectats, amb una precisió suficient per a tasques logístiques bàsiques.
- La comunicació mitjançant *MODBUS TCP* resulta eficaç, i la integració amb Polyscope permet desplegar seqüències de moviment adaptades.
- El sistema, tot i estar limitat a un entorn simulat, obre la porta a futurs desenvolupaments en assistència quirúrgica automatitzada.

Aquest TFG també ha assolit el seu objectiu didàctic, servint com a plataforma de formació i experimentació per a l'aprenentatge de tecnologies robòtiques en l'àmbit mèdic.