



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

## **Agraïments**

Arribar fins aquí ha estat un camí ple d'esforç, aprenentatges i creixement personal. Aquest treball no hauria estat possible sense el suport i l'acompanyament de moltes persones que vull agrair de tot cor.

En primer lloc, al professor Xavier Cufí, tutor del treball, per la seva implicació constant, la seva paciència i les orientacions encertades, sempre transmesses amb sinceritat, assertivitat i una claredat que s'agraeix. Gràcies per confiar en mi des del primer moment.

Als meus pares, gràcies per haver-me transmès des de petit la importància de l'educació com a eina de canvi i de creixement personal. El vostre suport constant ha estat fonamental per arribar fins aquí. La universitat m'ha enriquit a nivells que ni jo podia imaginar, i res d'això hauria estat possible sense vosaltres.

A la meua germana, gràcies per ser un pilar ferm i constant. Fins i tot en els moments en què jo dubtava de mi mateix i no veia clar si seria capaç d'assolir aquest repte, tu vas ser-hi, per recordar-me que sí, que ho aconseguiria.

També vull agrair als meus amics de la carrera, el meu fidel suport durant aquests quatre anys. Hem compartit molt més que assignatures: hem construït una família a Girona, una xarxa de complicitats i amistat que m'emporto per sempre.

I finalment, gràcies a totes aquelles persones que, d'una manera o altra, en més o menys mesura, m'han ajudat a arribar fins aquí. Aquest èxit també és vostre.

## índex

1.	Introducció .....	0
1.1.	Context i Antecedents .....	0
1.2.	Objectiu del projecte.....	0
1.2.1.	Objectius específics .....	1
1.3.	Especificacions i abast del TFG .....	1
1.3.1.	Especificacions àmbit tècnic .....	1
1.3.2.	Especificacions experimentals .....	2
1.3.3.	Abast del TFG.....	2
1.4.	Objectiu didàctic .....	3
2.	Conceptes Previs.....	3
2.1.	L'ús de robots assistents en la robòtica mèdica .....	3
2.2.	El control per veu en contextos mèdics.....	5
2.3.	Sistemes de visió en l'àmbit mèdic .....	5
2.4.	Robots col·laboratius (Co-Bots) i treball conjunt amb humans .....	6
2.5.	Importància d'aquests conceptes pel projecte.....	7
3.	Estudi de Mercat.....	7
3.1.	Definició del mercat.....	7
3.2.	Anàlisi de l'oferta i la demanda .....	9
3.3.	Preus i canals de distribució.....	10
3.4.	Anàlisi de la competència .....	10
3.5.	Necessitats dels usuaris .....	13
3.6.	Components i materials necessaris .....	14
3.7.	Braç robòtic.....	14
3.8.	Sistema de captació i reconeixement de veu.....	15
3.8.1.	Xarxes Neuronals Convolucionals.....	16
3.9.	Interfície de control i comunicació.....	17
3.10.	Estructura i entorn de simulació .....	18
4.	Normativa i aspectes legals .....	20
4.1.	Normativa sobre dispositius mèdics .....	20
4.2.	Tractament de dades.....	22
4.3.	Comité d'ètica .....	22
5.	Concepte d'enginyeria.....	22
5.1.	Descripció de l'entorn de Treball .....	22
5.1.1.	Entorn de Treball – Co-Bot.....	23
5.1.1.1.	Eina de subjecció: Ventosa .....	23

5.1.2.	Entorn de Treball – Sistema de visió .....	24
5.1.3.	Entorn de Treball – Sistema de control per veu.....	26
5.1.4.	Entorn de Treball – Servidor de Modbus.....	26
5.2.	Calibració del sistema de visió .....	28
5.2.1.	Calibració del Tool Centre point (TCP).....	29
5.3.	Obtenció de la posició dels instruments quirúrgics .....	30
5.4.	Sistema de reconeixement de veu.....	33
5.5.	Enviament de dades.....	34
5.6.	Programa del co-bot.....	35
6.	Implementació i resultats .....	36
6.1.	Calibració càmera.....	36
6.2.	Reconeixement dels objectes.....	37
6.2.1.	Reconeixement dels objecte a través de bounding boxes .....	37
6.2.2.	Reconeixement dels objectes amb transmissió de coordenades .....	38
6.3.	Modbus.....	39
6.3.1.	Prova comunicació entre ordinador i cobot.....	39
6.3.2.	Enviament d'ordres estructurades per consola .....	39
6.3.3.	Enviament de dades estructurades a partir de comandes per veu.....	40
6.4.	Control per veu .....	40
6.4.1.	Prova inicial de control per veu .....	40
6.4.2.	Prova d'identificació objectes mitjançant control per veu .....	41
6.5.	Prova final .....	42
7.	Conclusions.....	43
7.1.	Acompliment dels objectius .....	43
7.2.	Contribució als objectius de Desenvolupament Sostenible (ODS) .....	44
7.3.	Linies de millora i treball futur.....	44
	Referencias.....	45
	ANNEXOS .....	48
A.	Planificacion .....	48
A.1.	Taula de planificació detallada .....	48
A.2.	Recursos i equipament utilitzat .....	49
A.3.	Diagrama de Gantt.....	49
A.4.	Desviacions respecte a la planificació inicial .....	50
B.	Codi i manual .....	51
C.	Pressupost.....	51
C.1.	Preus unitaris.....	52

C.2. Pressupostos parcials.....	52
C.3. Pressupostos totals.....	53
C.4. Comentari final .....	54
D. Càlcul i justificacions .....	54
D.1. Introducció .....	54
D.2. Patro de calibratge.....	54
D.3. Bounding Boxes .....	55
D.4. Conclusió i aplicació final .....	56

## ÍNDIX FIGURES

Figura 1: Robot assistent subministrant instruments quirúrgics.....	4
Figura 2: UR3 e-series d'Universal Robots .....	6
Figura 3: previsió de creixement del mercat de Robòtica mèdica.....	8
Figura 4: Creixement dels papers publicats sobre robòtica mèdica en els darrers anys.....	9
Figura 5: Robot Assistent quirúrgic Penelope.....	12
Figura 6: Flux operatiu del sistema de reconeixement i lliurament d'instruments .....	19
Figura 7: Eina per calcular el TCP .....	30
Figura 8:Dataset visualitzat des de Ultralytics HUB.....	31
Figura 9: Evolució del loss durant l'entrenament.....	32
Figura 10: Exemple de detecció amb coordenades i etiquetatge de l'instrument sol·licitat.....	33
Figura 11: Inicialització del servidor .....	34
Figura 12: Punts homografia calculada .....	37
Figura 13: Detecció de bounding boxes .....	38
Figura 14: Exemple de detecció amb coordenades convertides i instrument destacat en vermell. ....	39
Figura 15: Proba reconeixement d'audio.....	41
Figura 16: Exemple d'identificació correcta del objecte "knife" mitjançant comanda de veu "cuchillo".....	41
Figura 17: Diagrama de Gantt del projecte.....	50

## ÍNDIX TAULES

Taula 1: Sistemes robòtics actuals .....	11
Taula 2: Relació entre les coordenades del mon real i coordenades de la càmera .....	29
Taula 3: Planificació del projecte i tasques principals realitzades. ....	49
Taula 4: Estimació de preus unitaris.....	52
Taula 5: Pressupost final estimat .....	53

## 1. Introducció

### 1.1. Context i Antecedents

S'han dut a terme grans avenços tecnològics amb un gran impacte en l'àmbit de la medicina, i un dels camps amb més projecció és el de la robòtica mèdica. Aquesta, permet automatitzar processos, augmentar la precisió de les intervencions i reduir la càrrega de treball del personal sanitari. Dins dels quiròfans, la robòtica s'està consolidant com una eina clau, especialment en tasques repetitives i d'alta coordinació, com el lliurament d'instruments quirúrgics.

En aquest context, la manca creixent de personal qualificat —especialment, personal d'infermeria amb experiència en aquests procediments— representa un repte per a molts hospitals. Aquesta situació pot comprometre la fluïdesa del procediment quirúrgic i afectar la seguretat del pacient. La incorporació de robots col·laboratius com a assistents quirúrgics es presenta com una solució eficient per complementar el personal humà i garantir un nivell òptim de qualitat assistencial.

### 1.2. Objectiu del projecte

L'objectiu principal d'aquest treball és el desenvolupament d'un sistema robòtic col·laboratiu capaç d'assistir en el subministrament d'instruments quirúrgics durant una intervenció, a partir d'ordres de veu emeses pel personal mèdic. Aquest sistema actuarà com un assistent quirúrgic intel·ligent, combinant tècniques de visió per computador i reconeixement de veu per identificar i localitzar els estris requerits, i transmetre les seves coordenades a un robot UR3 perquè els manipuli de forma autònoma, segura i eficient.

Es desenvoluparà en un entorn simulat de quiròfan i servirà com a prova de concepte per explorar la viabilitat d'integrar aquest tipus de solucions robòtiques en l'àmbit sanitari, especialment en situacions amb escassetat de personal qualificat o recursos limitats.

### 1.2.1. Objectius específics

A fi d'aconseguir l'objectiu general plantejat es defineixen les següents fites:

1. Dissenyar i implementar un algorisme de visió per computador capaç de reconèixer automàticament diferents estris predefinits, representatius dels instruments quirúrgics habituals.
2. Desenvolupar un sistema per detectar les coordenades espacials dels estris en el pla d'imatge, transformar-les a coordenades reals i transmetre-les al sistema del robot UR3.
3. Crear un mòdul de reconeixement de veu que permeti identificar ordres verbals associades a cada estri quirúrgic, garantint una resposta àgil i fiable.
4. Integrar el sistema de reconeixement de veu amb el sistema de visió artificial, de manera que la petició verbal d'un instrument activi tot el procés de detecció, localització i transmissió de dades al robot.
5. Definir i provar diferents estratègies de moviment del braç robòtic per garantir que la manipulació dels estris sigui estable, precisa i adaptada a l'entorn quirúrgic.
6. Realitzar proves funcionals en un entorn controlat (in vitro) per validar l'eficàcia, precisió i robustesa del sistema desenvolupat.
7. Analitzar els resultats, a fi de valorar les possibles millores que pot aportar l'ús de sistemes robòtics assistents en aquest tipus de procediments.

### 1.3. Especificacions i abast del TFG

#### 1.3.1. Especificacionsn àmbit tècnic

Per garantir el correcte desenvolupament del sistema, s'han establert els següents requisits tècnics:

- Ha de ser capaç de reconèixer visualment, mitjançant visió per computador, un conjunt predefinit d'estrís quirúrgics col·locats sobre una superfície plana, independentment de la seva orientació.
- Les coordenades dels objectes detectats s'han de transformar del sistema d'imatge (píxels) a coordenades físiques reals per ser utilitzades pel robot UR3.
- El sistema de reconeixement de veu ha de permetre identificar comandes verbals associades a cada estri, amb una resposta ràpida i precisa.



- La comunicació entre els diferents mòduls (visió, veu, robot) s'ha de realitzar en temps real i amb una latència mínima.
- El protocol de comunicació entre el sistema informàtic i el braç robòtic ha de ser fiable i segur. En aquest projecte es farà servir el protocol MODBUS TCP
- El sistema ha de funcionar correctament en un entorn simulat de laboratori amb condicions d'il·luminació estables.

### 1.3.2. Especificacions experimentals

El sistema ha de garantir una experiència d'ús coherent amb el comportament esperat en un entorn quirúrgic simulat. Concretament, ha de complir amb les següents funcionalitats:

- Ha de permetre al personal sanitari seleccionar un instrument quirúrgic mitjançant una comanda verbal.
- El sistema ha de reconèixer l'estri sol·licitat, localitzar-lo dins el camp de visió i calcular-ne la posició relativa.
- Un cop identificat, el robot ha de ser capaç de desplaçar-se fins a la posició de l'estri, recollir-lo i deixar-lo en una ubicació determinada (zona d'entrega).
- El sistema ha de ser tolerant a petits errors de pronunciació mitjançant tècniques de comparació de cadenes (fuzzy matching).
- La manipulació dels estris ha de ser estable i suau, evitant moviments bruscos o col·lisions.
- Ha de permetre una fàcil reconfiguració del conjunt d'estris detectables, mitjançant la substitució o ampliació del dataset d'imatges.

### 1.3.3. Abast del TFG

Tenint en compte els recursos disponibles i els objectius plantejats, aquest TFG es limitarà als aspectes següents:

- El desenvolupament d'un prototip funcional que integri els mòduls de visió artificial, reconeixement de veu i control robòtic.
- La implementació del sistema en un entorn simulat (*in vitro*), utilitzant el robot UR3 i instrumental quirúrgic de prova.
- La realització de proves de funcionament per validar la precisió del reconeixement, la resposta del sistema i la viabilitat de la solució.

- No s'inclouran assajos en entorns clínics reals . El projecte s'enfoca exclusivament com una prova de concepte dins d'un marc acadèmic i experimental.

#### 1.4. Objectiu didàctic

A més del valor tecnològic, aquest projecte també té un objectiu didàctic: servir com a eina d'aprenentatge i experimentació per a estudiants i professionals de l'àmbit biomèdic. Mitjançant la simulació d'un entorn quirúrgic i la integració de diferents tecnologies (visió, veu i robòtica), es pretén oferir una plataforma accessible per a la formació en robòtica mèdica.

## 2. Conceptes Previs

Aquest apartat pretén introduir les bases necessàries per a poder comprendre el desenvolupament d'un sistema robòtic assistent en l'àmbit quirúrgic. Aquests conceptes seran claus per entendre la presa de decisions duta en el disseny d'aquest projecte.

### 2.1. L'ús de robots assistents en la robòtica mèdica

La robòtica mèdica ha evolucionat significativament en les darreres dècades, donant lloc a sistemes que poden assistir en procediments quirúrgics complexos, afavorint una major precisió i eficiència. Els robots assistents quirúrgics no pretenen substituir els professionals sanitaris, sinó complementar-ne la tasca, alleugerint la seva càrrega de treball i millorant la seguretat dels pacients. Aquests sistemes poden encarregar-se de funcions com el subministrament d'instruments, el manteniment d'espais estèrils o el seguiment de protocols operatius.



*Figura 1: Robot assistent subministrant instruments quirúrgics*

L'impacte d'aquestes tecnologies és especialment rellevant en contextos amb escassetat de personal especialitzat, on poden contribuir a mantenir un alt nivell d'exigència quirúrgica mitjançant la col·laboració directa amb el personal humà [1] [2].

En comparació amb els mètodes quirúrgics convencionals, la incorporació de sistemes robòtics aporta beneficis importants. Aquests inclouen una major precisió en els moviments, la possibilitat d'automatitzar tasques repetitives, la reducció de la fatiga del personal i una millora en la gestió del temps quirúrgic.

Tanmateix, l'objectiu d'aquesta tecnologia no és substituir cap rol professional, sinó actuar com a suport a les tasques que poden ser mecàniques o logístiques, alliberant el personal per concentrar-se en decisions crítiques i accions que requereixen judici clínic i expertesa. La rotació freqüent de personal d'infermeria i tècnic deguda a retallades pressupostàries fa que sovint els cirurgians hagin de treballar amb personal que no coneix les seves preferències o no les recorda, afectant la fluïdesa del procediment. Solucions com aquesta, tenen l'intenció de resoldre aquest problema actuant com un assistent que apren les preferències i idiosincràsies del cirurgià, i no les oblida mai, la qual cosa proporciona un nivell de personalització que rarament s'aconsegueix en entorns amb alta rotació de personal.

Per al personal d'infermeria, aquest tipus de sistema pot suposar un gran avantatge, ja que alleujaria part de les seves funcions rutinàries, permetent-los concentrar-se en tasques més crítiques i oferint una millor atenció al pacient. A més, la reducció de la càrrega de treball i la possibilitat de delegar certes operacions contribueix a minimitzar el risc d'errors provocats per fatiga o sobrecàrrega de tasques [1] [3].

## 2.2. El control per veu en contextos mèdics

El reconeixement de veu representa una via d'interacció home-màquina especialment útil en entorns quirúrgics, on la necessitat de mantenir la màxima esterilitat i concentració és fonamental. Aquest factor fa que les solucions sense contacte, com el control per veu, adquireixin un valor pràctic molt elevat.

Aquest tipus de tecnologia permet accedir a funcionalitats complexes de sistemes informàtics o robòtics sense necessitat d'interrompre les tasques manuals del professional sanitari. En un entorn com el quiròfan, on sovint es requereix una atenció constant i una actuació directa, aquesta capacitat de comunicació verbal resulta més natural, eficient i intuïtiva que l'ús de pantalles o interfícies físiques [4].

A més, aquests sistemes poden ser personalitzats per reconèixer exclusivament les veus de persones autoritzades. Això afegeix una capa addicional de seguretat, evitant activacions accidentals i assegurant que només el personal acreditat pugui emetre comandes rellevants durant el procediment.

## 2.3. Sistemes de visió en l'àmbit mèdic

La visió per computador és una tecnologia fonamental per al funcionament dels robots assistents quirúrgics, ja que és el que els permet "veure", identificar i entendre el seu entorn de treball. Aquesta capacitat visual fa possible la localització i el reconeixement automàtic d'instruments, pacients i superfícies, i constitueix el primer pas per a qualsevol acció robòtica guiada per percepció.

Hi ha diverses tècniques per implementar la visió per computador segons les necessitats de cada aplicació. Es poden utilitzar sistemes de visió en dues dimensions (2D) per reconèixer estris quirúrgics mitjançant trets visuals com contorns, formes o colors. Aquest enfocament és suficient en molts casos, especialment quan els instruments són simètrics i els trets clau es poden identificar des d'una sola perspectiva [5]. Per a instruments més complexos, es pot dividir l'objecte en diverses parts o etiquetes visuals que el sistema aprèn a relacionar.

En situacions que requereixen una percepció de profunditat o comprensió espacial més acurada, es pot optar per sistemes de visió estereoscòpica. Aquesta tècnica simula la visió binocular humana mitjançant dues càmeres col·locades a certa distància i angle. Les imatges obtingudes són comparades per calcular la profunditat i generar una

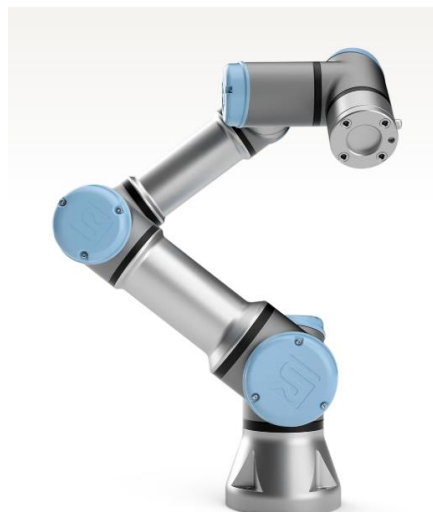
reconstrucció tridimensional de l'escena. Aquest tipus de visió és especialment útil per a tasques de manipulació precises o quan cal distingir diferents altures o capes dins del camp de treball quirúrgic.

Gràcies a aquestes tècniques, la visió per computador facilita una comprensió digital de l'espai que permet als robots operar amb precisió i autonomia dins d'un entorn tan exigent com el quiròfan.

#### 2.4. Robots col·laboratius (Co-Bots) i treball conjunt amb humans

Els robots col·laboratius (cobots) són dissenyats per compartir l'espai i les tasques amb els humans. A diferència dels robots industrials tradicionals, que requereixen espais segregats per seguretat, els cobots incorporen sensors i sistemes de protecció que els permeten detectar la presència humana i adaptar el seu comportament per evitar col·lisions [6] [7].

Aquestes característiques fan dels cobots una eina molt útil en el camp mèdic, on les condicions de treball poden ser variables i les interaccions amb personal humà són constants. El model UR3 de Universal Robots és un exemple de cobot versàtil, compacte i programable, apte per ser integrat en entorns quirúrgics i realitzar tasques com la manipulació d'instruments o la col·locació d'eines de manera segura [8] [9]. A continuació es mostra una fotografia del robot esmentat.



*Figura 2: UR3 e-series d'Universal Robots*

## 2.5. Importància d'aquests conceptes pel projecte

Conèixer aquestes tecnologies i el seu paper en l'entorn mèdic és essencial per comprendre el plantejament del projecte que es desenvolupa en aquest treball. La integració de sistemes de visió per computador, reconeixement de veu i robòtica col·laborativa respon a una realitat tecnològica emergent, que ja s'està començant a aplicar en contextos reals.

Aquest marc conceptual permet entendre millor les motivacions, les decisions de disseny i les implicacions del sistema proposat, tant des del punt de vista funcional com ètic i logístic.

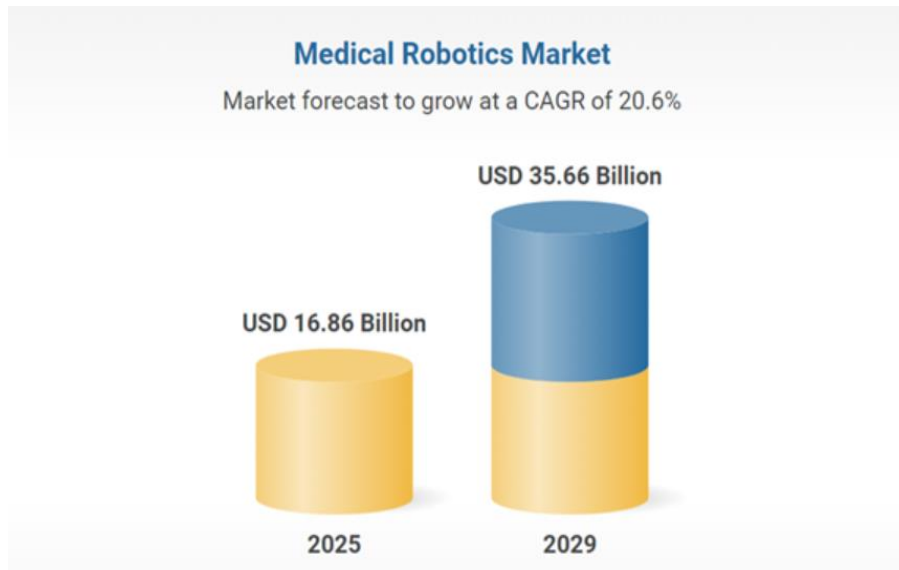
## 3. Estudi de Mercat

### 3.1. Definició del mercat

L'ús de robots mèdics ha experimentat un creixement molt significatiu durant la darrera dècada, impulsat per la seva alta precisió, fiabilitat i capacitat per millorar els resultats. Aquestes qualitats han convertit la robòtica mèdica en una eina clau per augmentar la qualitat assistencial, reduir la variabilitat dels procediments i oferir un millor servei als pacients.

Aquest creixement està estretament vinculat a la recerca constant d'eficiència, productivitat i seguretat en l'àmbit sanitari [10] [11]. Cada cop són més les solucions robòtiques disponibles al mercat, capaces d'assumir tasques diverses com ara la manipulació d'instruments, el suport al diagnòstic o la navegació quirúrgica.

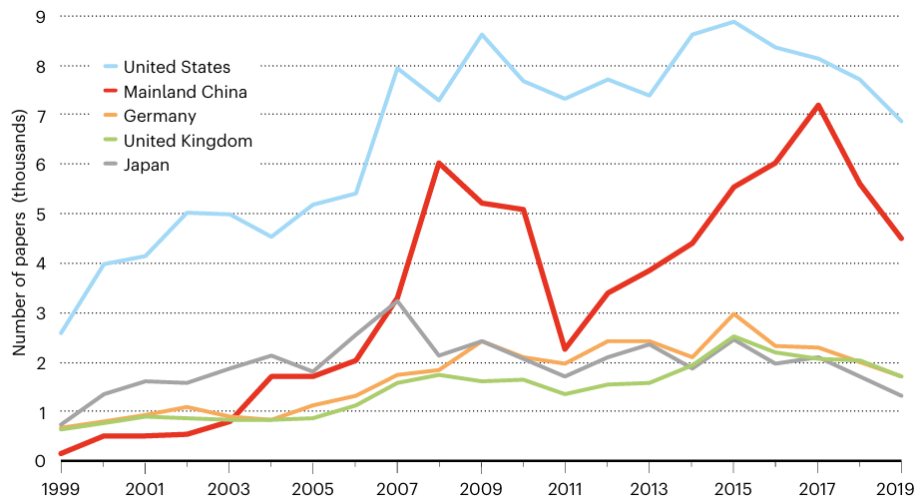
Segons dades recents, el mercat global de la robòtica mèdica va assolir un valor de 14,59 mil milions de dòlars el 2014, amb una previsió de creixement fins als 16,86 mil milions el 25 i una taxa de creixement anual composta (CAGR) propera al 16 %. Es preveu que aquesta xifra superi els 20,6 mil milions el 2029, tal com es pot observar en la gràfica següent.



*Figura 3: previsió de creixement del mercat de Robòtica mèdica*

Dins d'aquesta tendència de creixement, tecnologies com la visió per computador i el reconeixement de veu han guanyat un pes destacat. La seva integració en entorns quirúrgics obre la porta a sistemes més automatitzats, eficients i versàtils, capaços d'adaptar-se a múltiples escenaris i necessitats clíniques [12].

A més del creixement comercial, el desenvolupament de la robòtica mèdica es veu reflectit també en l'àmbit de la recerca científica. Tal com es pot observar en la gràfica següent, el nombre de publicacions sobre sistemes robòtics mèdics ha augmentat notablement durant les últimes dues dècades, especialment als Estats Units i la Xina. Aquesta tendència indica una inversió sostinguda en innovació i validació científica, requisits imprescindibles per consolidar aquestes tecnologies en el mercat sanitari. L'augment de publicacions evidencia l'interès global creixent per sistemes com els



*Figura 4: Creixement dels papers publicats sobre robòtica mèdica en els darrers anys*

assistents quirúrgics, i reforça la rellevància d'aquest projecte en un context d'evolució constant [13].

És en aquest context que el present projecte vol posar en relleu la necessitat d'explorar la confluència de camps com la robòtica, el reconeixement de veu i la visió artificial, amb aplicacions directes en l'àmbit mèdic. Aquesta visió està plenament alineada amb les tendències actuals tant a de recerca, com a nivell de mercat, i que apunten cap a una transformació digital progressiva de les pràctiques clíniques.

### 3.2. Anàlisi de l'oferta i la demanda

L'escassetat de personal sanitari qualificat és un factor que continua impulsant la demanda de sistemes automatitzats dins del quiròfan. Tal com recull l'estudi de Max Schäfer [14], la introducció de robots del tipus *scrub nurse*, dissenyats per subministrar instruments quirúrgics, pot ajudar a alleujar la càrrega de treball dels professionals d'infermeria, especialment en contextos de tensió laboral i alta rotació de personal. Aquest tipus de robots poden complementar la tasca humana i aportar estabilitat i eficiència a tasques altament repetitives i crítiques per a la fluïdesa del procediment.

Així doncs, la demanda actual no només es basa en una voluntat de modernització, sinó també en una necessitat real d'adaptar-se a una nova realitat assistencial: més pacients, més complexitat quirúrgica, i menys recursos humans disponibles. Alhora, l'oferta



tecnològica està responent amb sistemes cada vegada més versàtils, assequibles i orientats a entorns col·laboratius.

### 3.3. Preus i canals de distribució

El mercat de la robòtica mèdica presenta una gran varietat de preus, trobem sistemes com serien el Penelope Surgical Instrument Server, un robot semiautònom centrat en el subministrament d'instruments quirúrgics. Penelope utilitza visió per computador i reconeixement de veu per facilitar les tasques logístiques dins del quiròfan, amb un enfocament clarament orientat a complementar la feina del personal d'infermeria instrumentista [1]. El seu cost estimat era d'uns 100 000 \$, amb una vida útil prevista de cinc anys i un cost operatiu d'uns 7 \$ per hora [15]. Aquest preu, tot i ser significativament més baix que el de sistemes com el *da Vinci Surgical System* (que pot superar el milió d'euros), continua sent una inversió considerable per a molts centres sanitaris.

Els sistemes com el desenvolupat en aquest projecte (basat en el cobot UR3, càmera, micròfon i software fet amb llenguatge Python) es poden construir a partir de components molt més econòmics i disponibles comercialment, mantenint una funcionalitat útil per a la docència o, fins i tot, quiròfans experimentals que vulguin fer passos en aquest tipus d'investigacions o pràctiques en l'àmbit de la robòtica mèdica.

Pel que fa als canals de distribució, aquests solen incloure:

- Distribuïdors especialitzats en tecnologia mèdica.
- Contractació directa amb fabricants per a grans sistemes.
- Transferència tecnològica i spin-offs per a prototips universitaris o modulars.
- Compra pública centralitzada, en el cas de sistemes sanitaris públics.

Aquesta diversitat de canals permet l'aparició de solucions intermitjes com el sistema proposat, que pot actuar com una alternativa viable en termes de cost i funcionalitat, adaptant-se a les necessitats concretes d'un hospital mitjà o petit.

### 3.4. Anàlisi de la competència

El mercat de la robòtica mèdica està dominat per un reduït grup de grans empreses tecnològiques que concentren la major part de les vendes globals. Aquestes companyies han desenvolupat sistemes quirúrgics sofisticats, generalment centrats en la cirurgia mínimament invasiva (CMI), que combinen precisió mecànica, visió avançada i suport digital. Les tres empreses amb més impacte actualment són:

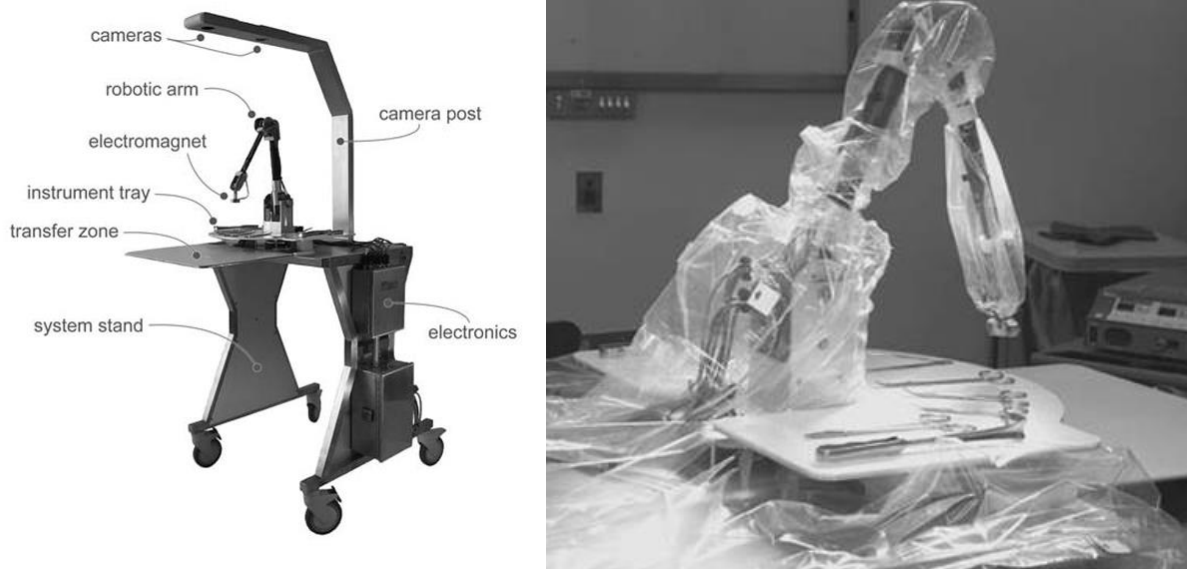
- **Intuitive Surgical:** Líder indiscutible del sector amb el seu sistema da Vinci Surgical System, present a més de 70 països. És un sistema integral que permet al cirurgià operar mitjançant una consola, controlant braços robòtics amb visió 3D d'alta definició. Més de 12.000 equips instal·lats arreu del món avalen la seva consolidació comercial.
- **CMR Surgical:** Empresa britànica creadora del sistema Versius, una alternativa modular i més assequible al da Vinci. Cada braç robòtic és independent, fet que facilita la seva instal·lació en quiròfans petits. El sistema ha tingut una ràpida adopció en centres públics i privats a Europa i Llatinoamèrica.
- **Medtronic:** Multinacional que ha entrat amb força al sector amb el sistema Hugo, dissenyat per competir directament amb da Vinci. Destaca per la seva arquitectura oberta, capacitat d'integració amb sistemes d'imatge i la connexió remota per monitoritzar intervencions.

	Empresa	Model principal	Tipus de sistema	Enfocament quirúrgic	Preu aproximat (USD)	Estat de mercat
1	Intuitive Surgical	da Vinci Surgical System	Sistema integrat (multi-braç, consola 3D)	CMI general (urologia, ginecologia, toràctica, etc.)	> 1.000.000	Consolidat (més de 12.000 unitats venudes)
2	CMR Surgical	Versius	Sistema modular (braços independents)	CMI general, adaptable a quiròfans petits	Entre 600.000 i 800.000	En expansió (múltiples hospitals europeus i llatinoamericans)
3	Medtronic	Hugo	Sistema híbrid amb arquitectura oberta	CMI amb capacitats d'integració i telediagnòstic	Similar a da Vinci (segons configuració)	Recent entrada, en creixement

*Taula 1: Sistemes robòtics actuals*

Tot i que aquests sistemes representen el segment més desenvolupat de la robòtica quirúrgica, la competència directa en l'àmbit dels *surgical instrument servers* (robots assistents encarregats de gestionar i subministrar estris quirúrgics) és molt més reduïda i encara en fase emergent.

Un dels més destacats i que ja s'ha mencionat anteriorment és el projecte Penelope aquest, pot identificar i passar instruments de forma semi-autònoma, utilitzant visió 2D i control per veu.



*Figura 5: Robot Assistant quirúrgic Penelope*

Un altre exemple és Gestonurse, un sistema que combina reconeixement de gestos, veu i visió per computador per respondre a ordres del cirurgià. Malgrat els seus resultats prometedors, no ha estat adoptat àmpliament per la indústria [16]. També trobem recerques recents com la de Schäfer [14], que exploren solucions amb end-effectors flexibles per a la manipulació d'instruments mitjançant cobots, però es troben encara en fase de laboratori.

Aquesta manca de competència directa estable deixa un espai de mercat obert per a solucions com la proposada en aquest projecte, enfocades a tasques molt concretes, però rellevants dins del quiròfan.

D'altra banda, cal destacar que els sistemes robòtics estan evolucionant cap a una major modularitat i diversificació. Fet que fa que puguin construir-se a partir de mòduls independents: braç robòtic, sistema de visió, control per veu, etc. Aquesta tendència permet adaptar els sistemes a les necessitats específiques de cada hospital, reduint costos i facilitant la integració progressiva.

Els últims avenços del sector apunten cap a sistemes cada cop més col·laboratius, més assequibles i més intel·ligents, amb capacitat per adaptar-se a diferents especialitats, procediments i entorns. La incorporació d'intel·ligència artificial, reconeixement de veu, aprenentatge per reforç i simulació 3D en temps real obre la porta a una nova generació

de sistemes quirúrgics flexibles, que no només actuen, sinó que aprendran i optimitzaran les seves accions en funció de l'ús clínic real.

### 3.5. Necessitats dels usuaris

L'entorn quirúrgic és altament exigent tant des del punt de vista tècnic com humà. El personal d'infermeria instrumentista, en particular, té la responsabilitat de gestionar múltiples estris quirúrgics, assegurant la seva disponibilitat i localització en tot moment, sovint mentre aquests objectes canvien constantment de lloc durant la intervenció. Aquesta tasca requereix una gran capacitat d'atenció, rapidesa i coordinació amb el cirurgià.

Sovint, els professionals es veuen obligats a treballar amb col·laboradors que no coneixen les preferències del cirurgià, la qual cosa pot afectar la fluïdesa i eficiència del procediment. Aquesta situació genera estrès, incrementa el risc d'errors i limita la capacitat dels professionals per centrar-se en tasques realment crítiques.

Tots aquests sistemes comparteixen una mateixa missió: reduir la càrrega de treball del personal sanitari, automatitzant aquelles funcions repetitives que, tot i ser essencials, no requereixen una valoració clínica directa. Això permet als professionals enfocar-se en les tasques de més valor afegit, millora la coordinació dins de l'equip i redueix el risc d'errors derivats de la fatiga o el desgast.

És cert que la implementació d'aquests robots pot presentar una corba d'aprenentatge inicial, especialment en entorns clínics amb poca experiència en tecnologia. No obstant això, els estudis mostren que, un cop integrats, aquests sistemes aporten una millora tangible en l'eficiència operativa i en la qualitat de l'atenció al pacient. A més, la capacitat de memòria, precisió i consistència dels robots els converteix en una eina extremadament valuosa en un entorn on la fiabilitat pot ser crítica [17].

En definitiva, els robots assistents quirúrgics com Penelope, Gestonurse o els sistemes experimentals més recents, responen a una necessitat concreta i creixent del sistema sanitari: alleujar la càrrega del personal i garantir una execució precisa i estable de tasques logístiques en el quiròfan.

### 3.6. Components i materials necessaris

Per al desenvolupament del projecte s'ha optat per una combinació de components de laboratori i materials de fàcil adquisició, amb l'objectiu de recrear un entorn quirúrgic funcional i accessible per a la fase de prototipat.

Els components principals utilitzats són els següents:

- **Braç robòtic UR3-e** de Universal Robots: utilitzat per a la manipulació física dels objectes, com a sistema de simulació del moviment dins del quiròfan.
- **Càmera webcam Logitech 720p**: col·locada en una posició zenital per capturar l'escena i permetre el processament d'imatges per visió per computador.
- **Micròfon integrat en auriculars**: emprat per captar les ordres verbals emeses per l'usuari durant l'execució de les proves.
- **Estris simulats**: s'han utilitzat objectes domèstics com unes tisores, unes pinces d'higiene personal i un ganivet de cuina, actuant com a equivalents simbòlics d'instruments quirúrgics reals.

Amb aquests elements s'ha pogut recrear un escenari assequible i funcional que simula les condicions bàsiques d'un entorn quirúrgic, suficient per validar les capacitats del sistema en termes de detecció, reconeixement i manipulació d'objectes.

### 3.7. Braç robòtic

Els robots col·laboratius (o *cobots*) representen un nou paradigma dins la robòtica industrial i mèdica. A diferència dels robots convencionals, que operen en espais aïllats per garantir la seguretat, els cobots han estat dissenyats per treballar mà a mà amb els humans en entorns compartits. Aquesta capacitat de col·laboració es deu a la integració de sensors de força, visió i detecció de presència, així com algorismes de control adaptatiu, que permeten detectar persones properes i modificar la seva trajectòria o acció per evitar col·lisions [8] [7].

Aquesta filosofia de disseny fa dels cobots una solució especialment útil en entorns dinàmics i delicats, com els quiròfans. A diferència dels sistemes robòtics industrials, que solen requerir personal tècnic altament especialitzat per a la seva programació [18], els cobots estan pensats per facilitar-ne el seu ús.

En aquest projecte s'ha optat per utilitzar el UR3e de Universal Robots, un dels cobots més compactes i versàtils del mercat. Aquest robot actua com a element d'execució dins

del sistema desenvolupat, encarregant-se de recollir els instruments quirúrgics identificats pel sistema de visió per computador i sol·licitats mitjançant ordres de veu. El seu paper és essencial per materialitzar la interacció entre la detecció digital i l'acció física, permetent simular de manera realista l'assistència robòtica dins d'un quiròfan.

El UR3e, és fàcilment programable mitjançant entorns gràfics i compatible amb entorns de simulació o control extern (com ROS, MODBUS o Python APIs) [19]. Aquestes característiques fan del UR3e una plataforma idònia per al desenvolupament de prototips mèdics col·laboratius, i en concret, per reproduir de forma efectiva un sistema de lliurament d'instruments quirúrgics controlat mitjançant veu i visó artificial.

### 3.8. Sistema de captació i reconeixement de veu

En aquest projecte, el reconeixement de veu juga un paper fonamental, ja que mitjançant ordres verbals, l'usuari pot sol·licitar l'instrument quirúrgic desitjat, el qual és posteriorment localitzat i manipulat pel robot.

Per al desenvolupament funcional del sistema, s'ha implementat la llibreria *SpeechRecognition* de Python [20], coneguda per la seva facilitat d'ús, compatibilitat amb diverses APIs (*Google, Sphinx, IBM, etc.*) i integració directa amb micròfons i arxius d'àudio. Aquesta eina captura el senyal de veu, l'envia a un servei de reconeixement i en rep com a resposta una transcripció en text de la comanda pronunciada.

Tot i l'efectivitat d'aquests sistemes, és habitual que es produeixin errors de transcripció, especialment quan les paraules tenen sons similars, hi ha soroll ambiental o l'usuari vocalitza de forma poc clara. Per tal de fer el sistema més robust, s'ha integrat una tècnica anomenada *Fuzzy String Matching* [2], que permet comparar cadenes de text de manera tolerant a errors.

Un dels algorismes més emprats per implementar aquesta tècnica és la distància de Levenshtein [21], que calcula el nombre mínim d'operacions (insercions, eliminacions o substitucions de caràcters) necessàries per transformar una cadena en una altra. Com menor sigui aquesta distància, més similars es consideren les cadenes. Aquest mètode permet al sistema seleccionar l'instrument quirúrgic més semblant a la paraula detectada per veu, fins i tot si la coincidència no és exacta.

A continuació es mostra la formulació recursiva de la distància de Levenshtein:

$$\text{lev}(a, b) = \begin{cases} |a| & \text{if } |b| = 0, \\ |b| & \text{if } |a| = 0, \\ \text{lev}(\text{tail}(a), \text{tail}(b)) & \text{if } \text{head}(a) = \text{head}(b), \\ 1 + \min \begin{cases} \text{lev}(\text{tail}(a), b) \\ \text{lev}(a, \text{tail}(b)) \\ \text{lev}(\text{tail}(a), \text{tail}(b)) \end{cases} & \text{otherwise} \end{cases}$$

*Equació 1: Formula distancia de Levenshtein*

Explicació de la fórmula:

- Si una de les cadenes té longitud zero, la distància és igual a la longitud de l'altra (cal inserir o eliminar tots els caràcters).
- Si els primers caràcters de totes dues cadenes coincideixen, la distància és la mateixa que entre les cadenes sense aquest primer caràcter (cas recursiu).
- En cas contrari, es calcula el cost mínim entre tres possibles operacions: inserció, eliminació o substitució, afegint 1 a cada cas.

Aquest sistema híbrid entre reconeixement de veu i comparació tolerant a errors és clau per garantir la fiabilitat i la usabilitat del sistema, fent-lo més proper al llenguatge natural i menys dependent de la perfecció de la comanda verbal.

### 3.8.1. Xarxes Neuronals Convolucionals

Dins d'aquest projecte, l'aprenentatge automàtic juga un paper fonamental a l'hora de detectar i identificar visualment els estris quirúrgics. Per aconseguir-ho, s'ha utilitzat una Xarxa Neuronal Convolutional (CNN), concretament el model YOLOv8 [22], conegut per la seva alta velocitat i precisió en tasques de detecció d'objectes en temps real.

Aquest sistema de visió per computador permet localitzar els instruments quirúrgics dins del camp visual de la càmera mitjançant la generació de requadres delimitadors al voltant de cada objecte reconegut, associant-los a una classe predefinida. El model ha estat entrenat amb un dataset personalitzat d'imatges dels instruments a reconèixer, capturades en diverses condicions d'il·luminació i posició.

A diferència de mètodes tradicionals basats en regles o patrons manuals, les CNN tenen la capacitat d'aprendre automàticament les característiques visuals rellevants (contorns, formes, textures) a partir de les dades. Aquest aprenentatge es duu a terme mitjançant una cadena de capes convolucionals, pooling i capes completament connectades, que processen la imatge d'entrada fins a generar una predicció amb una probabilitat associada per a cada objecte detectat [23].

La integració d'aquest model dins del sistema permet, un cop reconegut l'instrument sol·licitat per veu, localitzar-lo automàticament dins del pla de treball i enviar-ne les coordenades al braç robòtic UR3e per a la seva manipulació física. Aquesta combinació de detecció visual i comandes verbals constitueix el nucli funcional del prototip, i demostra el potencial de les tecnologies d'intel·ligència artificial en aplicacions reals dins l'àmbit quirúrgic.

### 3.9. Interfície de control i comunicació

Un dels elements clau per garantir la coordinació entre els diferents mòduls del sistema és la interfície de comunicació entre el reconeixement de veu, el processament visual i el braç robòtic UR3e. Aquest sistema segueix un flux seqüencial: l'usuari emet una comanda verbal, aquesta és transcrita a text mitjançant la llibreria *SpeechRecognition*, i posteriorment es compara amb les ordres permeses utilitzant una tècnica de *Fuzzy Matching*.

Un cop l'instrument és identificat, el sistema de visió per computador localitza la seva posició dins del pla de treball i calcula les coordenades físiques. Aquestes dades s'envien al braç robòtic mitjançant el protocol de comunicació *MODBUS*, un estàndard industrial molt utilitzat per establir comunicació entre dispositius electrònics en entorns de control.

*MODBUS* segueix una arquitectura client-servidor (també anomenada mestre-esclau). En aquest projecte, l'ordinador actua com a client (o mestre), que envia les instruccions, mentre que el UR3e funciona com a servidor (o esclau), que escolta i executa les ordres rebudes [24].

A nivell de programació, aquest sistema s'ha implementat mitjançant la llibreria *pymodbus* en Python [25], que facilita l'escriptura i lectura dels registres necessaris per transmetre les coordenades calculades.

Pel que fa al software del UR3e, es basa en el sistema operatiu *Polyscope*, que proporciona una interfície intuïtiva per a la programació directa, i també ofereix suport



per a scripts en URScript, el seu propi llenguatge. A més, el robot permet la comunicació externa mitjançant protocols com *MODBUS*, *ROS*, *socket TCP* o *API REST*, cosa que facilita la seva integració en entorns personalitzats i sistemes de recerca com el desenvolupat en aquest treball.

Aquesta arquitectura modular i la compatibilitat amb protocols estàndard, permeten construir una cadena de comunicació robusta i flexible que integra veu, visió i acció robòtica en temps real, com aquesta.

### 3.10. Estructura i entorn de simulació

Per tal de validar el funcionament del sistema desenvolupat, s'ha dissenyat un entorn de simulació experimental que reproduceix, de manera simplificada, les condicions d'un quiròfan. L'objectiu ha estat demostrar la viabilitat tècnica de la interacció entre veu, visió artificial i robòtica col·laborativa per automatitzar el lliurament d'instruments quirúrgics.

La disposició general dels elements es pot observar a la figura següent:



Figura 6: Flux operatiu del sistema de reconeixement i lliurament d'instruments

L'escenari es compon dels següents elements principals:

- Una webcam situada en una posició zenital, perpendicular al pla de treball, encarregada de transmetre la imatge dels objectes quirúrgics al sistema informàtic.
- Un conjunt d'objectes simulats (tisores, pinces, bisturí), utilitzats com a representació d'instruments quirúrgics reals.
- Un micròfon, integrat en uns audífons, per captar les ordres verbals de l'usuari (professional sanitari).
- Un ordinador amb el programari desenvolupat per a:

- Reconeixement de veu
  - Processament d'imatges
  - Càlcul de coordenades espacials dels objectes
- El braç robòtic UR3e, encarregat d'executar l'acció física: recollir l'instrument sol·licitat i traslladar-lo a una zona de lliurament.

Aquest entorn experimental ha estat pensat per ser modular i escalable. Tot i que actualment només s'utilitza una càmera i una ventosa com a eina de subjecció, el sistema pot evolucionar fàcilment en funció dels requeriments futurs, tant des del punt de vista tècnic com funcional.

En definitiva, aquest experiment permet validar el concepte, a l'hora que serveix de precedent per aplicacions més avançades en la docència i l'àmbit mèdic. Per exemple, es podria ampliar la infraestructura amb múltiples càmeres, explorar altres mètodes de manipulació, millorar la interacció amb el professional mitjançant lliurament directe dels instruments. Aquest tipus d'adaptacions podrien ser abordades en futurs treballs acadèmics, prototips clínics o iniciatives de recerca aplicada.

#### 4. Normativa i aspectes legals

Aquest treball no ha estat dissenyat amb l'objectiu immediat de ser implementat dins d'un quiròfan real. No obstant això, ha estat fonamental garantir un cert grau de versemblança per tal d'aproximar-se el màxim possible a un entorn clínic. És per aquest motiu que s'han tingut en compte algunes de les normatives i consideracions que caldria complir en un projecte d'aquestes característiques si es volgués aplicar en l'àmbit sanitari professional.

##### 4.1. Normativa sobre dispositius mèdics

Els dispositius mèdics, i en particular els sistemes robòtics aplicats a la medicina, estan subjectes a una regulació estricta tant a nivell nacional com internacional per garantir la seguretat dels pacients i la qualitat dels procediments. Qualsevol dispositiu que s'utilitzi en l'àmbit clínic ha de complir amb una sèrie de normatives, certificacions i controls que assegurin la seva fiabilitat i traçabilitat.

Un dels marcs normatius més importants en aquest context és el Reglament (UE) 2017/745 [26] sobre productes sanitaris, que estableix els requisits essencials per a la

comercialització i ús de dispositius mèdics dins de la Unió Europea. Segons aquest reglament, qualsevol sistema que tingui una aplicació mèdica directa (com la manipulació d'instruments durant una intervenció quirúrgica) es considera un dispositiu mèdic i ha de ser sotmès a:

- Classificació de risc (Classe I, IIa, IIb o III) segons el seu grau d'invasivitat, duració d'ús i interacció amb el cos humà.
- Certificació per part d'un organisme notificat que validi el compliment dels requisits de seguretat i rendiment clínic.
- Avaluació clínica i validació experimental, on es demostrï l'eficàcia del sistema mitjançant estudis controlats o literatura científica.
- Marcatge CE, que acredita la conformitat amb la legislació europea i permet la seva distribució al mercat.

A més, per als dispositius que incorporen programari intel·ligent, com els sistemes basats en intel·ligència artificial o visió per computador, s'han de considerar també els requisits de la norma IEC 62304 (programari de dispositius mèdics), així com les guies sobre IA publicades per l'Agència Europea del Medicament (EMA) i la Medical Device Coordination Group (MDCG).

En l'àmbit nacional, l'Agència Espanyola de Medicaments i Productes Sanitaris (AEMPS) és l'encarregada de regular i supervisar aquests dispositius. Qualsevol prototip que es vulgui provar en un entorn clínic hauria de comptar amb l'autorització de l'AEMPS i, en molts casos, amb l'aprovació d'un comitè ètic.

Finalment, cal tenir en compte aspectes relacionats amb la protecció de dades (RGPD) si el sistema processa informació personal de pacients, així com la normativa d'higiene, esterilització i seguretat elèctrica recollida a la norma ISO 13485 i a la IEC 60601 (seguretat de dispositius mèdics electrònics).

Tot i això, cal destacar que aquest projecte ha estat desenvolupat en un entorn acadèmic per un estudiant i amb finalitats experimentals. Per tant, no ha estat concebut per a la seva aplicació clínica directa i no pretén, en cap cas, substituir les normatives legals requerides per a dispositius mèdics reals. Tanmateix, s'ha procurat seguir bones pràctiques de disseny i seguretat amb l'objectiu de replicar el màxim possible un entorn mèdic simulat.

## 4.2. Tractament de dades

En cas que aquest sistema fos aplicat en un entorn quirúrgic real, caldria garantir el compliment estricte de la normativa sobre protecció de dades personals, especialment pel fet que el sistema pot captar veu i imatge dins d'un espai clínic. Això implicaria estar subjecte al Reglament General de Protecció de Dades (RGPD) —Reglament (UE) 2016/679— així com a la LOPDGDD (Llei Orgànica 3/2018), que regula el tractament de dades personals a Espanya.

Aquest tipus de dades podrien considerar-se dades especialment sensibles, ja que poden identificar professionals o pacients, i per tant caldria comptar amb consentiment exprés, protocols de xifrat, emmagatzematge segur, i garantir que el sistema no emmagatzemi informació personal innecessària.

Així mateix, la implementació real d'aquesta tecnologia requeriria una avaluació d'impacte en protecció de dades (AIPD) i l'aprovació del comitè ètic corresponent.

Tot i això, cal destacar que aquest projecte ha estat desenvolupat en un entorn simulat i acadèmic, i no realitza cap tipus de tractament de dades personals reals, ni conserva registres d'àudio ni d'imatge amb finalitats d'identificació.

## 4.3. Comité d'ètica

Tot i que aquest projecte no ha implicat la participació directa de persones ni el tractament de dades personals reals, el seu enfocament simula un entorn quirúrgic. Per tant, s'han tingut en compte les consideracions ètiques corresponents, especialment en relació amb la captació d'imatges i ordres de veu, per tal que el sistema pugui evolucionar en el futur dins d'un marc normatiu i ètic adequat.

# 5. Concepte d'enginyeria

## 5.1. Descripció de l'entorn de Treball

En aquest apartat del treball es descriu la solució implementada per a la manipulació d'objectes quirúrgics mitjançant un braç robòtic, garantint les condicions de treball pròpies de l'entorn que es pretén simular: el quiròfan. S'analitza el flux de funcionament del sistema, els diferents algorismes utilitzats, les variables implicades i la lògica darrere de les decisions tècniques adoptades durant el desenvolupament.

L'algorisme ha estat dissenyat per controlar el robot de manera automàtica, sense necessitat de transmetre manualment la informació, utilitzant el protocol de comunicació *MODBUS*.

Tota la informació complementària relacionada amb càlculs, llibreries de codi i altres justificacions tècniques es pot consultar a l'Annex D.

#### 5.1.1. Entorn de Treball – Co-Bot

El sistema es basa en un braç robòtic col·laboratiu UR3e de Universal Robots, el model més compacte de la gamma e-Series. Amb un pes lleuger i una estructura reduïda, l'UR3e presenta un radi d'abast (reach) de 500 mm i una capacitat de càrrega de fins a 3 kg, fet que el fa especialment idoni per a tasques de precisió en espais reduïts, com els quiròfans. La seva mida compacta facilita la integració al costat d'altres equips mèdics i permet realitzar manipulacions de petit format —com subministrar instruments quirúrgics— amb gran precisió.

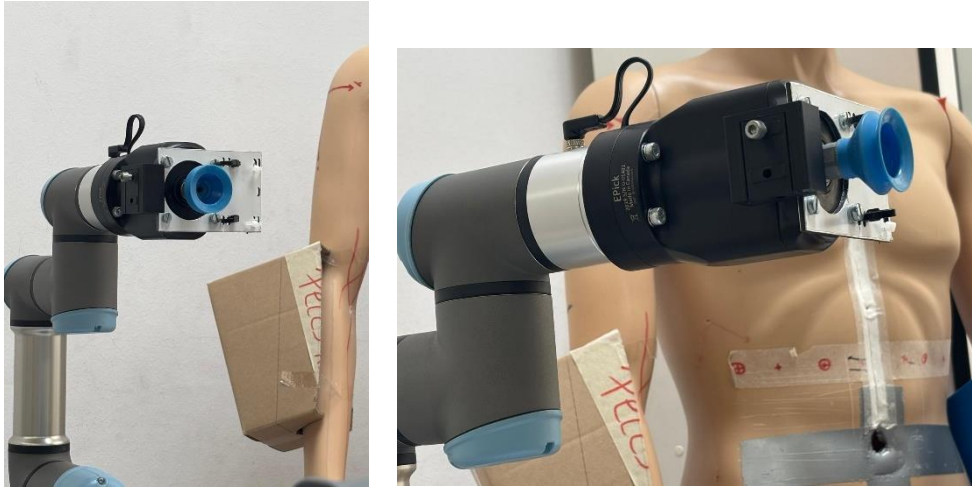
Aquest robot disposa de 6 graus de llibertat, cosa que li proporciona una gran flexibilitat per adaptar la seva orientació i posició segons la necessitat de cada moment. Aquest grau de maniobrabilitat és essencial per interactuar en entorns tridimensionals i complexos com l'espai quirúrgic.

Pel que fa a la seguretat, l'UR3e està dissenyat per operar en entorns col·laboratius, i incorpora múltiples sistemes de detecció de col·lisions i control de força. Això permet que pugui treballar colze a colze amb professionals humans, sense necessitat de barreres físiques, fet que el converteix en una eina versàtil i segura per a la robòtica mèdica.

##### 5.1.1.1. Eina de subjecció: Ventosa

Com a eina de manipulació, s'ha optat per una ventosa d'aspiració, acoblada a l'extrem del braç robòtic. Aquesta solució és senzilla, segura i eficaç per a la recollida d'objectes lleugers i amb una superfície plana, com els simulacres d'instruments quirúrgics emprats en aquest projecte. La ventosa permet agafar els objecte, minimitzant així els riscos de dany o mal posicionament durant la manipulació.

Aquesta eina també facilita una integració ràpida i modular, amb la possibilitat de ser substituïda en futurs desenvolupaments per sistemes de subjecció més avançats, com ara pinces robòtiques o manipuladors amb sensors tàctils.



*Figura 7: Ventosa*

#### 5.1.2. Entorn de Treball – Sistema de visió

Per a la captura d'imatges del pla de treball s'ha utilitzat una webcam Logitech C270, un dispositiu assequible però prou eficaç per a l'objectiu d'aquest projecte. Aquesta càmera ofereix una resolució de 720p (1280x720 píxels) i un rati de 30 fotogrames per segon (fps), la qual cosa garanteix una qualitat d'imatge suficient per detectar els contorns i formes dels instruments quirúrgics amb precisió i fluïdesa. Dins del context simulat en què es desenvolupa aquest projecte, aquestes característiques compleixen els requisits per a un sistema de detecció d'objectes en temps real, sense necessitat d'una càmera industrial o de gamma alta.

A causa de la distorsió que introdueixen les lents de les càmeres comercials, especialment en els marges de la imatge, es va haver de dur a terme un procés de calibratge de la càmera. Aquest calibratge és fonamental per corregir deformacions i obtenir una correspondència precisa entre les coordenades de la imatge (en píxels) i les coordenades físiques del món real (en mil·límetres).

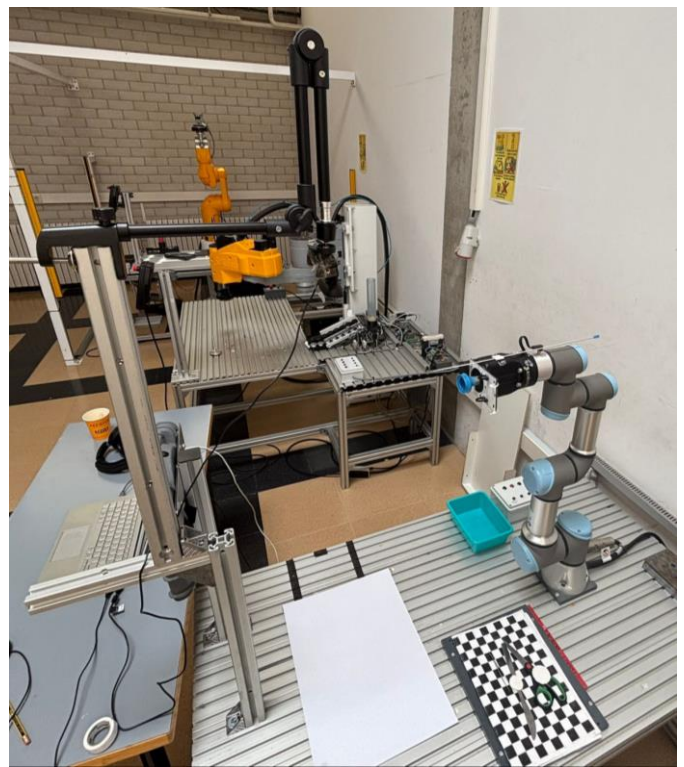
Per a la calibració, es va utilitzar un patró d'escacs, capturant diversos punts coneguts sobre la imatge del taulell (concretament, 6 punts de referència visibles) i comparant-los amb els punts reals corresponents dins el sistema de coordenades del robot. Amb aquesta informació, es va poder calcular la matriu d'homografia, que és una



transformació projectiva que permet passar d'un sistema de coordenades (el de la imatge) a un altre (el del pla físic de treball).

Aquesta homografia proporciona la relació matemàtica que permet transformar qualsevol punt detectat en la imatge a la seva posició real dins l'espai, assumint que els objectes es troben en un mateix pla. És especialment útil en entorns com aquest, on es treballa amb objectes sobre una taula plana i no és necessari conèixer la seva alçada.

Per al muntatge del sistema de visió, es van utilitzar vigues metàl·liques compatibles amb la taula de treball, que es van fixar de manera estable per tal d'obtenir l'alçada adequada per a la càmera. Aquesta es va instal·lar en una posició zenital, és a dir, perpendicular al pla de treball, fet que permet visualitzar tota la zona operativa des d'una perspectiva uniforme i centrada. Aquesta disposició simplifica considerablement el problema de localització, ja que elimina la variable d'alçada i assegura que els objectes quirúrgics siguin visualitzats des de la mateixa perspectiva en tot moment. A més, facilita el càlcul de coordenades bidimensionals i permet utilitzar un model de visió 2D, més senzill i eficient que altres enfocaments tridimensionals. Per tal d'ajustar amb precisió l'angle de visionat i garantir una orientació òptima, es va incorporar un suport articulat per a càmeres web, al qual es va fixar la càmera a l'extrem.



*Figura 8: Muntatge sistema de visió*



### 5.1.3. Entorn de Treball – Sistema de control per veu

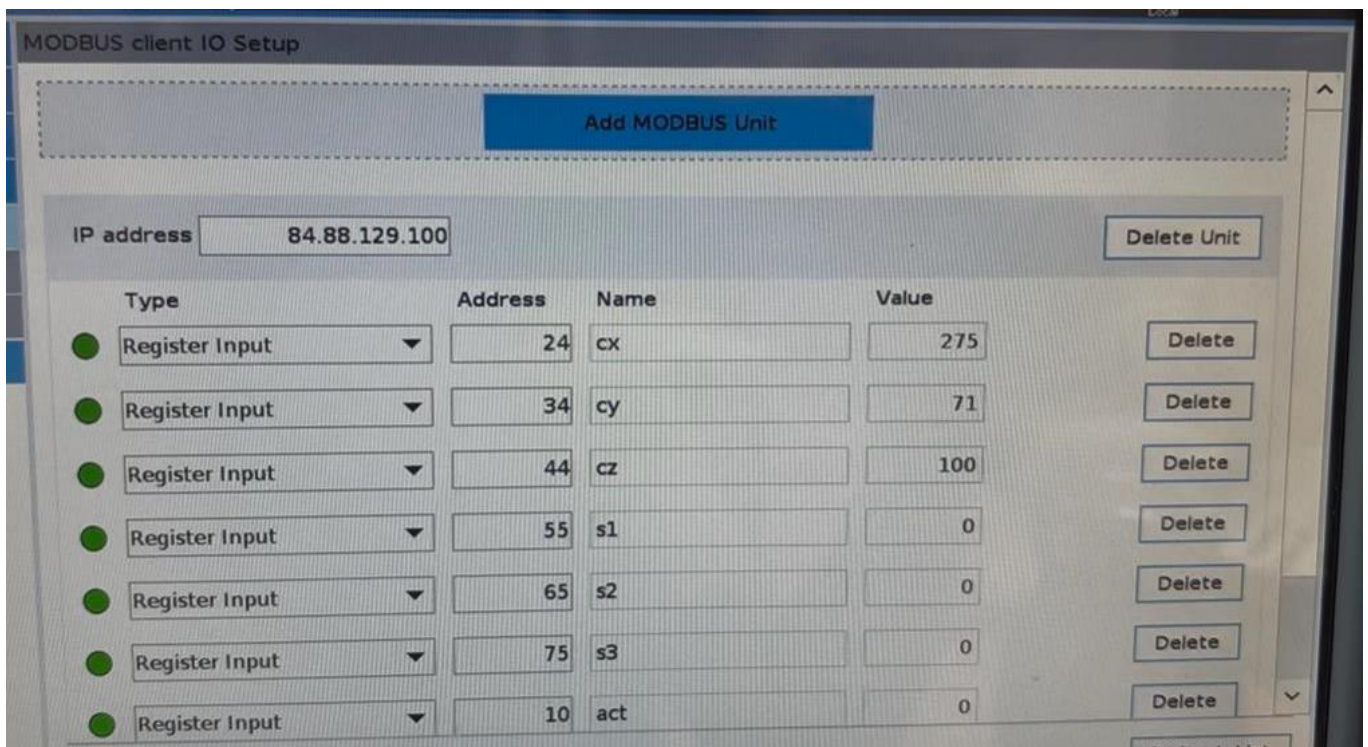
Per a la captació de les ordres verbals s'ha utilitzat el micròfon integrat dels auriculars Sony WH-XB910N. Aquest dispositiu proporciona una qualitat de captació nítida i estable, amb reducció activa de soroll ambiental i orientació cap a la veu de l'usuari, característiques que el fan adequat per reconèixer comandes vocals en entorns amb un cert nivell de soroll de fons. Per a un projecte d'aquestes característiques, és recomanable utilitzar un micròfon que ofereixi almenys una taxa de mostreig una mica superior a la mitja, com per exemple 16 kHz, una relació senyal-soroll (SNR) superior a 60 dB, i connectivitat estable amb l'equip de processament [27]. Els Sony WH-XB910N compleixen amb escreix aquests requisits, i han permès una integració senzilla i eficient amb el sistema de reconeixement de veu desenvolupat en Python.

### 5.1.4. Entorn de Treball – Servidor de Modbus

Per a la comunicació entre el sistema de processament (ordinador) i el braç robòtic UR3e s'ha utilitzat el protocol *MODBUS* TCP, un estàndard industrial àmpliament adoptat per l'intercanvi de dades entre dispositius. Aquest protocol segueix un esquema mestre-esclau: un dispositiu actua com a mestre (client) que envia sol·licituds, i un altre com a esclau (servidor) que respon amb les dades sol·licitades.

Perquè el robot pogués rebre i interpretar correctament aquestes dades, va ser necessari crear una unitat *MODBUS* des de la interfície gràfica de Polyscope. Aquest procés es fa mitjançant el menú *Installation > Fieldbus > MODBUS Client*, on es defineixen els canals de registre d'entrada per a cada variable. Un cop creada la connexió i configurats els canals, aquests apareixen a la interfície amb una circumferència de color gris (desconnectat) o verd (connexió activa), permetent verificar visualment l'estat de la comunicació.

Un cop establerta la connexió, el client calcula i envia la informació necessària per a l'execució d'un moviment: sis coordenades reals (X, Y, Z, RX, RY, RZ) en un rang entre 0 i 65535, ja que Modbus només permet valors en aquest rang, per tant es va decidir afegir sis registres més corresponents als signes de cada coordenada (+1 o -1), que indiquen la direcció de cada component. Aquesta informació es transmet a través de canals d'entrada de registre (*input registers*), que el robot pot llegir des del seu sistema intern.



*Figura 9: Registres coordenades Modbus*

A més de les coordenades i signes, es va afegir un canal extra utilitzat com a indicador d'activitat. Aquest canal rebia un valor de 1 quan les coordenades havien estat enviades correctament i estaven llestes per ser processades pel robot, i tornava a 0 un cop el moviment s'havia completat. Aquesta variable permetia controlar el flux d'execució i assegurar que el robot no iniciés una nova acció abans d'haver completat l'anterior, evitant així condicions de carrera o execucions incompletes.

L'ús de *MODBUS*, juntament amb la configuració gràfica de *Polyscope* i la gestió de registres d'entrada, ha permès una integració estable, ràpida i segura entre el sistema de visió-control i el robot, mantenint una arquitectura clara i escalable.

## 5.2. Calibració del sistema de visió

El procediment va consistir en seleccionar sis punts distribuïts sobre un tauler d'escacs col·locat al pla de treball. Aquests punts van ser clicats a la imatge obtinguda per la càmera, utilitzant un script en Python (vegeu l'Annex B) que permet registrar les coordenades en píxels dels punts seleccionats.

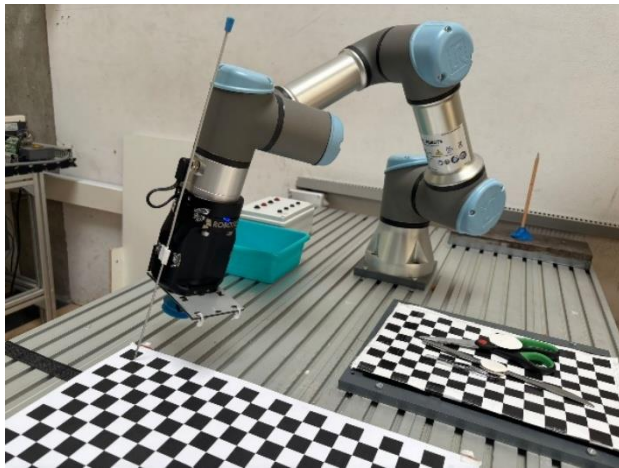


Figura 10: Procés de calibració cobot - càmera

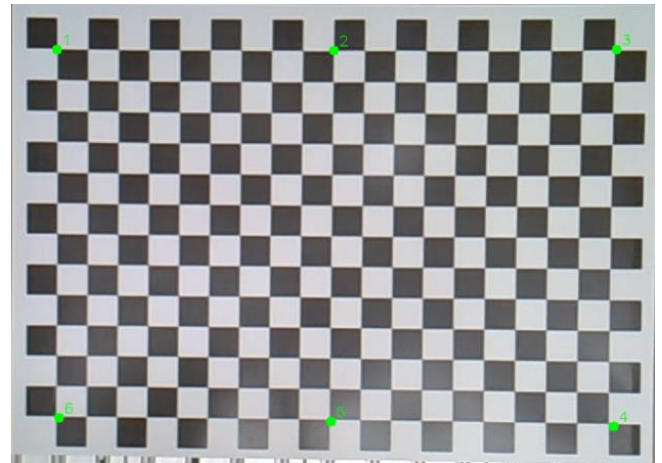


Figura 11: Tauler de calibració cobot – càmera

Posteriorment, per obtenir les coordenades reals corresponents en el sistema del robot, es va fer ús d'un palo de referència fixat al *Tool Center Point (TCP)* del UR3e. El robot es va desplaçar manualment a cada punt sobre el tauler, i les posicions X i Y (en centímetres) es van introduir manualment a través del terminal del script. D'aquesta manera es va obtenir el conjunt de punts reals en el mateix pla.

Punt	Coordenades imatge (x, y)	Coordenades reals (X, Y) [cm]
1	(482, 252)	(15.743, -63.214)
2	(483, 187)	(16.271, -54.844)
3	(197, 186)	(-19.192, -63.548)
4	(196, 249)	(-19.513, -55.321)

*Taula 2: Relació entre les coordenades del món real i coordenades de la càmera*

Amb els dos conjunts de punts (imatge i món real), es va calcular la matriu d'homografia mitjançant la funció `cv2.findHomography()` d'*OpenCV*. Aquesta matriu projectiva permet transformar qualsevol punt detectat en la imatge al seu equivalent en el pla físic (XY), assumint que tots els punts es troben sobre una mateixa superfície plana. La matriu resultant *Homografia.npy* és carregada a l'inici del codi principal i utilitzada durant tot el procés de detecció i actuació robòtica per convertir les coordenades de píxel dels objectes quirúrgics detectats en coordenades reals útils per al robot.

#### 5.2.1. Calibració del Tool Centre point (TCP)

Durant el procés de calibratge del sistema, es va fer necessari definir amb precisió el Tool Center Point (TCP) del robot UR3e. Com que la ventosa prevista com a eina final del sistema no tenia la longitud adequada per arribar còmodament a tots els punts del tauler de calibratge, es va optar per utilitzar un pal com a eina provisional.

Per tal de calcular el TCP associat a aquest pal, es va fer ús de la funció de calibració automàtica del robot, disponible a la seva interfície gràfica. Aquesta eina requereix que l'extrem de l'objecte a calibrar toqui amb precisió un mateix punt fix des de diferents orientacions, per tal que el sistema pugui inferir automàticament la posició del TCP respecte a la brida del robot.

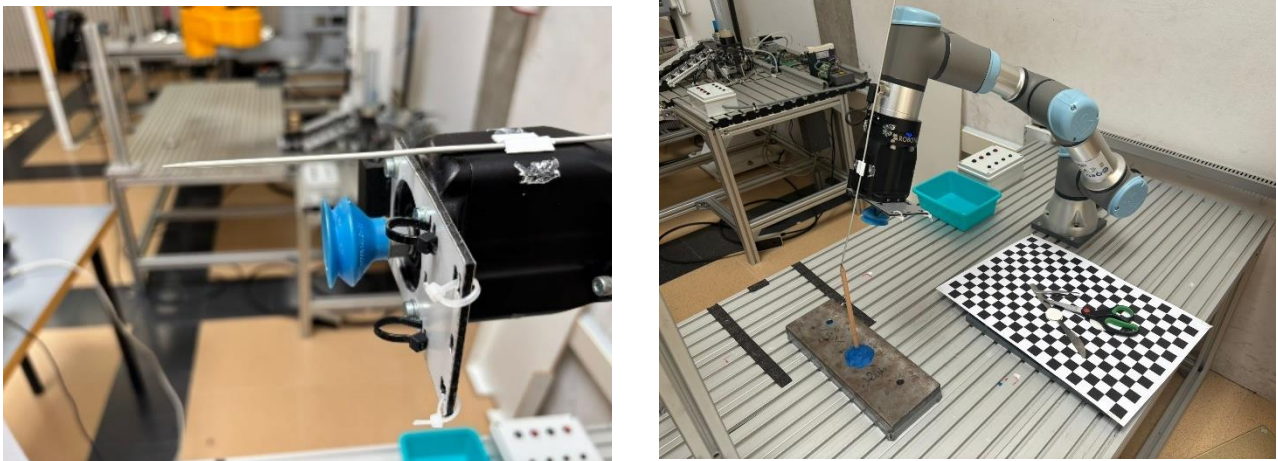
Per facilitar aquesta operació, es va utilitzar un llapis com a referència fixa: el pal fixat al robot es feia coincidir amb la punta del llapis des de diferents angles, seguint les instruccions del sistema. Aquest procediment va permetre definir un TCP associat al pal de manera precisa.

No obstant això, com que el pal no representava l'eina final del sistema, es va calcular posteriorment la diferència física entre aquest TCP provisional i el TCP real corresponent

a la ventosa. Aquesta diferència es va determinar manualment mitjançant mesuraments, obtenint els següents valors relatius:

- **+4 cm** d'alçada (eix Z), pel desfasament vertical entre la punta del pal i el centre de la ventosa.
- **+1 cm** en l'eix X, degut a un lleuger desplaçament lateral.
- **+7,5 cm** en l'eix Y, que correspon a la longitud que sobresortia el pal respecte al centre real.

Aquests valors van ser aplicats manualment al sistema, modificant la definició de la nova eina des de la interfície del robot, per obtenir un TCP equivalent al de la ventosa definitiva, però basat en les mesures preses amb l'eina provisional.



*Figura 7: Eina per calcular el TCP*

### 5.3. Obtenció de la posició dels instruments quirúrgics

Un cop calibrat el sistema de visió i establerta la relació entre els píxels de la imatge i les coordenades reals del pla de treball mitjançant la matriu d'homografia, el següent pas clau és la detecció i localització dels instruments quirúrgics sobre la superfície operativa.

Per tal de reconèixer aquests instruments a través del sistema de visió, es va entrenar un model basat en *YOLOv8* (*You Only Look Once*), una arquitectura d'aprenentatge profund optimitzada per a la detecció d'objectes en temps real amb alta precisió i eficiència. Aquest model va ser seleccionat per la seva capacitat de detectar múltiples



objectes simultàniament, així com per la seva rapidesa i compatibilitat amb equips informàtics limitats.

El primer pas per a l'entrenament del model va ser la creació d'un *dataset* personalitzat, format per 214 imatges que mostraven els tres instruments quirúrgics utilitzats al projecte: tisores, bisturí i pinces. Aquestes imatges es van dividir en tres subconjunts:

- *Train* (entrenament)
- *Validation* (validació)
- *Test* (proves finals)

Cada imatge anava acompanyada del seu fitxer d'anotacions amb l'etiqueta de l'objecte i les coordenades de la seva *bounding box*, seguint el format *YOLO*. La informació estructural del dataset es va definir en un arxiu *.yaml*, on es descrivien les rutes a cada conjunt, així com les classes utilitzades.

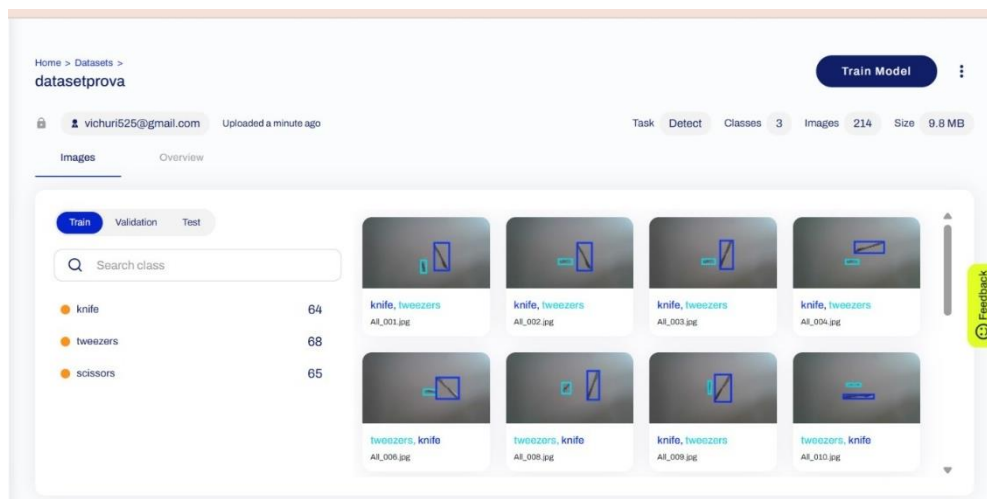


Figura 8:Dataset visualitzat des de Ultralytics HUB

L'entrenament es va realitzar mitjançant la plataforma *Ultralytics HUB*, que facilita la gestió del dataset i la visualització de resultats. Durant aquest procés, es van monitoritzar tres mètriques de pèrdua (loss): *box loss*, *class loss* i *object loss*.

Es pot observar una millora clara en la *box loss* i en la *class loss* per al conjunt d'entrenament, indicant que el model aprèn a ajustar millor les caixes delimitadores i les classes durant el procés. Tot i això, en el conjunt de validació, la *class loss* i la *object loss* es mantenen més estables, mostrant que el model no generalitza tan bé com en l'entrenament. Aquest comportament pot ser indicatiu d'un sobreajustament moderat, però el rendiment global del model es considera adequat per a l'objectiu planteja



*Figura 9: Evolució del loss durant l'entrenament*

Un cop entrenat, el model YOLOv8 va demostrar una capacitat robusta per detectar i identificar amb precisió els instruments quirúrgics dins del camp visual de la càmera. El sistema de detecció, operant en temps real, generava per a cada objecte detectat una *bounding box* i a partir d'aquesta caixa, es calculava el centre de l'objecte (Veure el càlcul en l'Annex D).

Aquest centre, encara en coordenades de píxel, es convertia a coordenades físiques reals (X, Y) mitjançant la matriu d'homografia calculada prèviament. Aquestes coordenades permetien localitzar amb precisió els instruments sobre el pla físic de treball.

Quan l'usuari sol·licitava verbalment un instrument, el sistema activava el reconeixement d'imatges i comparava les deteccions amb la comanda de veu. L'instrument sol·licitat es marcava en color vermell a la pantalla, mentre que els altres apareixien en verd, facilitant-ne el seguiment visual.

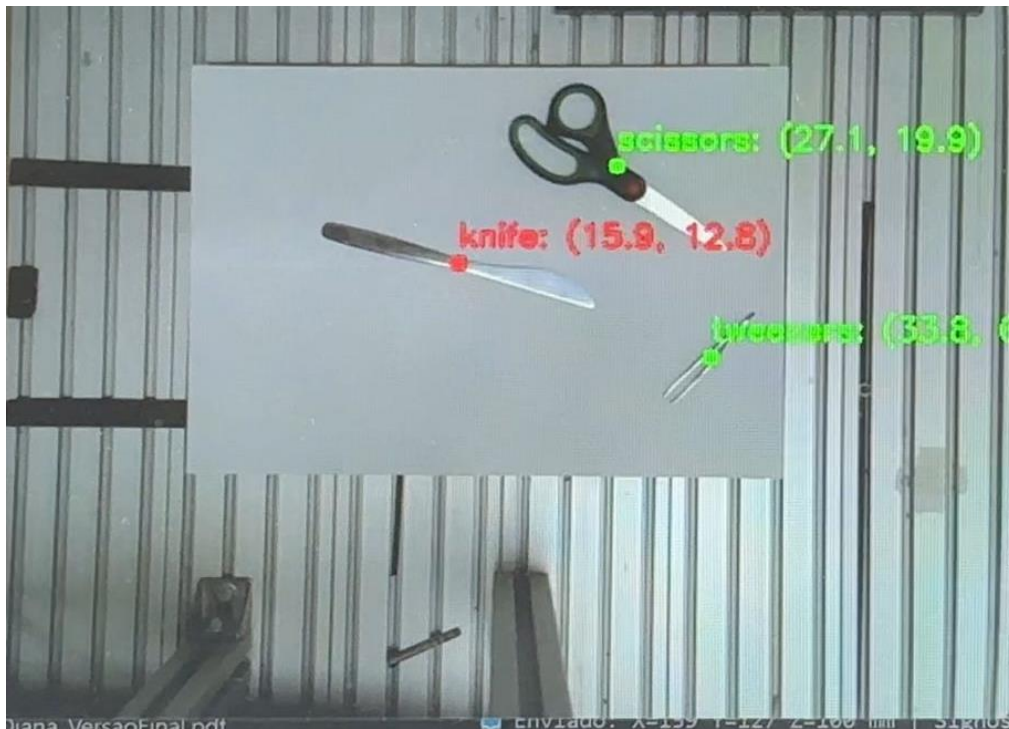


Figura 10: Exemple de detecció amb coordenades i etiquetatge de l'instrument sol·licitat

Finalment, un cop localitzat, el sistema enviava les coordenades de l'instrument al robot a través del protocol *MODBUS TCP*, per tal que aquest pogués recollir-lo i col·locar-lo segons la comanda rebuda.

Aquest model resulta lleuger, ràpid i fàcilment escalable. A més, es podria adaptar a nous entorns clínics o escenaris educatius simplement afegint més classes d'instruments o ampliant el *dataset* d'entrenament.

#### 5.4. Sistema de reconeixement de veu

Per enllaçar la comanda verbal amb l'objecte detectat pel sistema de visió, es va implementar un sistema de traducció automàtica mitjançant un diccionari *translation\_dict* que converteix termes en espanyol com “*tijeras*”, “*cuchillo*” o “*pinzas*” als corresponents en anglès que utilitza el model de detecció visual “*scissors*”, “*knife*”, “*tweezers*”. Això permet identificar coincidències aproximades entre el text reconegut i



els objectes detectats a la imatge, fins i tot en presència de petites errades o variacions en la transcripció de la veu.

Un cop identificat l'instrument sol·licitat, el sistema n'obté les coordenades (X, Y) mitjançant el model de visió i les envia al robot per a la seva manipulació.

Aquest sistema proporciona una interacció intuïtiva i sense contacte físic, i demostra com el reconeixement de veu pot convertir-se en una eina pràctica i eficient en aplicacions mèdiques reals.

### 5.5. Enviament de dades

Per a la implementació d'aquest protocol s'ha fet ús de la llibreria *pymodbus*, que proporciona totes les eines necessàries per actuar com a client i servidor *MODBUS* des d'un script en Python. En aquest projecte, l'ordinador actua com a servidor *MODBUS slave*, i el robot com a client *MODBUS master*, accedint als registres d'entrada de dades de l'ordinador.

La inicialització del servidor s'ha fet amb el següent bloc de codi:

```
server = ModbusServer("84.88.129.100", 502, no_block=True)
print("● Servidor MODBUS activo en 84.88.129.100:502")
server.start()
```

```
PS C:\Users\nbsal\Desktop\TFG> & C:/Python312/python.exe
● Servidor MODBUS activo en 84.88.129.100:502
```

*Figura 11: Inicialització del servidor*

L'adreça IP 84.88.129.100 correspon a la configuració de xarxa manual del port *Ethernet* de l'ordinador. Per tal de garantir la comunicació entre el robot i el sistema de visió, ambdós dispositius es van connectar directament a través d'un cable *Ethernet*, formant part de la mateixa xarxa local. Aquesta connexió directa assegura una latència mínima i evita interferències amb altres dispositius o xarxes externes.

En relació amb els registres utilitzats per transmetre les dades, es va optar per una estructura ordenada que minimitzés possibles conflictes d'accés. En total es van enviar:

- 6 coordenades reals (X, Y, Z, RX, RY, RZ)
- 6 signes associats a cadascuna de les coordenades (+1 o -1)
- 1 registre addicional de control, que indicava si les dades havien estat enviades (1) o no (0)

Els registres es van assignar amb una separació de 10 posicions entre ells. Aquest espai entre registres es va implementar com a mesura preventiva per evitar qualsevol interferència o solapament de dades degut a possibles errors de sincronització, ja que en alguns sistemes *MODBUS* l'accés ràpid o simultani a registres contigus pot donar lloc a lectures incorrectes.

Aquest enfocament modular i espaiat facilita també la llegibilitat del codi i la futura ampliació del sistema, ja que es poden afegir nous camps sense haver de reestructurar el conjunt.

## 5.6. Programa del co-bot

El programa implementat a l'entorn Polyscope del robot UR3e té com a objectiu llegir les coordenades transmeses pel sistema de visió, desplaçar-se cap a l'objecte detectat i dipositar-lo en una posició de destí predefinida.

El robot rep les components X, Y, Z en centímetres a través dels registres configurats. Cada component es descomposa en dos valors: un registre per al valor absolut i un altre per al signe (+1 o -1), donat que el protocol de comunicació només admet valors enters positius.

A l'inici del programa, aquestes dades s'assignen a variables temporals. Després, es reconstrueixen les coordenades originals aplicant la conversió de signe i passant de mil·límetres a metres. Aquest càlcul es fa cada cop que el robot rep noves coordenades.

Un cop reconstruïdes les coordenades, s'assignen a la variable destí amb una orientació constant per a RX, RY i RZ:  $desti := p[C\_x, C\_y, C\_z, 0.841, 3.028, -0.326]$

Després, el robot realitza un moviment lineal (*MoveL*) cap a destí, on es troba l'instrument. Un cop posicionat, s'activa el sistema de succió (*vacuum*) per recollir l'objecte.

A continuació, el robot es mou cap a la posició "tray", que correspon al punt de dipòsit comú dels instruments. En aquest punt, es desactiva el sistema de succió per deixar l'objecte, i es retorna a la posició "home" per preparar la següent iteració.

Tot aquest procés es troba dins d'un bucle principal, que s'activa només quan es detecta que la variable d'activació ha estat posada a 1 per part del sistema central. Aquest mecanisme evita moviments innecessaris i assegura que només es recullin instruments quan la detecció visual ha estat validada.

## 6. Implementació i resultats

### 6.1. Calibració càmera

El procediment va consistir en clicar manualment 12 punts sobre la imatge de la taula de treball, mostrada a la pantalla mitjançant un *script* en Python. Es van seleccionar punts ben distribuïts per tota la superfície (vora, centre, cantonades) per tal de garantir una bona qualitat de la calibració.

Un cop clicats tots els punts sobre la imatge, es van introduir les coordenades equivalents en el món real utilitzant el robot. Per fer-ho, es va col·locar un pal al *Tool Center Point (TCP)* del robot UR3e i es va desplaçar manualment la seva punta cap a cadascun dels punts prèviament clicats, anotant les coordenades X i Y llegides des del sistema i es van introduir manualment a través del terminal del script.

Finalment, el programa calculava la matriu d'homografia utilitzant la funció *cv2.findHomography()* de la llibreria *OpenCV* i la guardava al fitxer *Homografia.npy* per al seu ús posterior.

A continuació es mostra un exemple comparatiu entre punts projectats i les seves coordenades reals corresponents, cosa que va permetre validar visualment la qualitat de la transformació:

```

✅ Homografía calculada y guardada como 'MAT_H_LAB2.npy'

🔪 Comparación de puntos proyectados vs reales:
Punto 1: Proyectado = (-17.88, -38.72) | Real = (-18.01, -38.81)
Punto 2: Proyectado = (-1.30, -39.12) | Real = (-1.45, -39.11)
Punto 3: Proyectado = (15.13, -39.41) | Real = (15.08, -39.37)
Punto 4: Proyectado = (-17.94, -46.11) | Real = (-17.68, -45.89)
Punto 5: Proyectado = (0.35, -46.55) | Real = (0.58, -46.37)
Punto 6: Proyectado = (15.02, -46.74) | Real = (14.97, -46.87)
Punto 7: Proyectado = (-18.28, -55.48) | Real = (-18.24, -55.75)
Punto 8: Proyectado = (-1.61, -55.81) | Real = (-1.66, -55.84)
Punto 9: Proyectado = (14.69, -55.90) | Real = (14.91, -56.17)
Punto 10: Proyectado = (-18.42, -61.00) | Real = (-18.51, -60.87)
Punto 11: Proyectado = (-0.03, -61.26) | Real = (-0.20, -61.42)
Punto 12: Proyectado = (14.57, -61.65) | Real = (14.52, -61.30)

```

Figura 12: Punts homografia calculada

## 6.2. Reconeixement dels objectes

El sistema de reconeixement dels instruments quirúrgics es va implementar de manera modular, dividint el procés en diferents etapes. Això va facilitar tant el desenvolupament com el testatge i depuració del codi. Inicialment es va centrar en la detecció bàsica dels objectes mitjançant *bounding boxes*, i posteriorment es va ampliar amb la transmissió automàtica de les coordenades reals per a la seva manipulació robòtica.

### 6.2.1. Reconeixement dels objecte a través de bounding boxes

Inicialment, es va observar que el model detectava múltiples objectes incorrectament. Elements que no formaven part del conjunt d'estrís quirúrgics (com bolígrafs o auriculars) eren reconeguts com a pinces o bisturís, degut a una confiança de detecció massa baixa establerta per defecte.

Per millorar-ne la precisió, es va anar ajustant progressivament el llindar de confiança mínima *confidence threshold* fins a trobar un valor que filtrés les deteccions falses sense perdre sensibilitat. Finalment, es va fixar una confiança mínima de 0.82, ja que permetia eliminar la majoria de falsos positius mantenint una detecció efectiva dels objectes reals.



*Figura 13: Detecció de bounding boxes*

#### 6.2.2. Reconeixement dels objectes amb transmissió de coordenades

Un cop superada la fase de detecció bàsica mitjançant bounding boxes, es va implementar una extensió funcional que permetia calcular la posició real de cada instrument quirúrgic dins el pla de treball, i transmetre-la al robot per al seu ús posterior.

A partir de cada *bounding box* detectat, es calcula el centre geomètric de l'objecte. Aquest punt central, encara expressat en coordenades de píxels, es transforma automàticament a coordenades físiques (X, Y) mitjançant l'aplicació de la matriu d'homografia obtinguda en la fase de calibratge (vegeu apartat 6.1). Aquest procés permet que el sistema interpreti la ubicació real de l'objecte sobre la superfície de treball.

Per tal d'afavorir la comprensió i la comprovació visual, es va afegir una capa gràfica que pinta el nom de cada objecte i la seva posició real sobre la imatge.



Figura 14: Exemple de detecció amb coordenades convertides i instrument destacat en vermell.

### 6.3. Modbus

#### 6.3.1. Prova comunicació entre ordinador i cobot

La primera prova de comunicació via *MODBUS* va consistir en enviar únicament una coordenada al robot per validar la correcta connexió entre el servidor (ordinador) i el client (UR3e). Mitjançant un script Python senzill, es va escriure de manera repetida un valor concret (per exemple, 10) al registre d'entrada 0. Aquesta dada es podia llegir des de la interfície del robot, confirmant així que el canal estava actiu i operatiu.

Un cop verificada la connexió, es va poder evolucionar cap a l'enviament de dades estructurades més complexes, com ara les sis coordenades (X, Y, Z, RX, RY, RZ), els signes corresponents, requerits per a la manipulació robòtica precisa i el bit d'activació.

#### 6.3.2. Enviament d'ordres estructurades per consola

Inicialment, aquestes dades es van introduir manualment per consola per tal de comprovar que el robot UR3e era capaç de llegir-les i interpretar-les correctament a través dels registres MODBUS definits. Aquest procés va servir per verificar tant la conversió de coordenades com la seva transmissió i recepció, abans d'integrar el sistema complet amb reconeixement de veu i visió per computador.

### 6.3.3. Enviament de dades estructurades a partir de comandes per veu

En aquesta darrera fase, es va integrar el sistema de reconeixement per veu amb el mòdul de visió i el servidor *MODBUS*, permetent un funcionament completament automatitzat. A diferència de les proves anteriors, on les dades s'enviaven de manera manual o per consola, aquí és l'usuari qui dona una comanda verbal.

Un cop la paraula és reconeguda correctament pel sistema de reconeixement de veu i validada mitjançant el procés de *fuzzy matching*, es busca si l'instrument sol·licitat es troba dins el camp visual. En cas afirmatiu, es detecta el centre de l'objecte mitjançant *YOLOv8*, es transforma la seva posició per a poder ser enviat mitjançant *MODBUS*, i aquestes són enviades al robot perquè pugui procedir amb la manipulació de l'instrument.

## 6.4. Control per veu

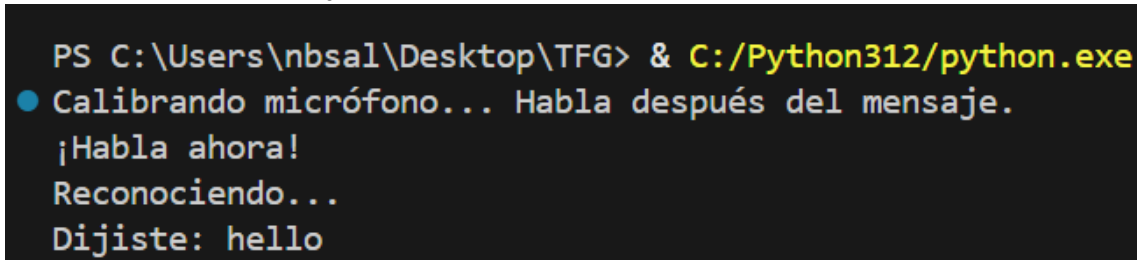
### 6.4.1. Prova inicial de control per veu

En aquesta primera prova es va implementar un sistema senzill de reconeixement de veu. L'objectiu era validar que el micròfon podia captar i interpretar correctament una comanda vocal, convertint-la en text utilitzable per al sistema.

El codi utilitzat inicialitzava el micròfon (concretament el dispositiu amb índex 3) i calibrava el soroll ambiental durant cinc segons per tal de millorar la qualitat de la detecció. A continuació, es capturava una comanda de veu durant un màxim d'un segon i es processava mitjançant l'*API* de Google en anglès (`language="en-US"`), retornant el text transcrit en cas d'èxit.



Aquest primer test va servir per validar la funcionalitat bàsica del reconeixement per veu, detectar possibles errors (com temps d'espera excedit, soroll ambiental elevat o paraules intel·ligibles) i establir una base sòlida per a la seva futura integració amb el sistema de detecció d'objectes i enviament de dades al cobot.



```
PS C:\Users\nbsal\Desktop\TFG> & C:/Python312/python.exe
● Calibrando micrófono... Habla después del mensaje.
¡Habla ahora!
Reconociendo...
Dijiste: hello
```

*Figura 15: Proba reconeixement d'audio*

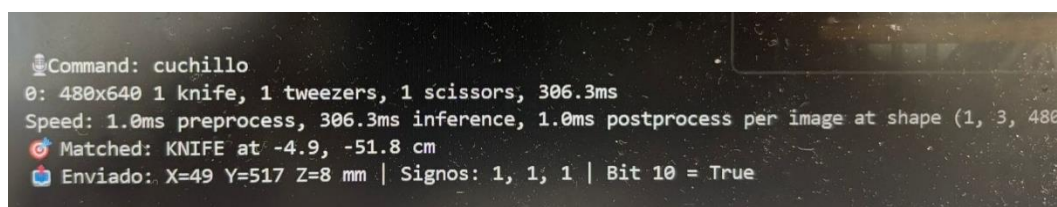
#### 6.4.2. Prova d'identificació objectes mitjançant control per veu

En aquesta fase es va integrar el sistema de reconeixement per veu dins del codi principal, de manera que les ordres verbals pronunciades pel professional poguessin activar directament el sistema de visió i reconeixement d'objectes quirúrgics.

Un cop el sistema escolta i reconeix una comanda, aquesta es compara amb la llista d'instruments disponibles.

A més, es va decidir mantenir les ordres de veu en espanyol, ja que durant les proves es va detectar una millor taxa d'encert del sistema respecte a l'anglès.

Quan el sistema identifica correctament l'objecte sol·licitat, aquest es marca a la imatge amb un punt vermell i el nom de l'objecte en vermell, mentre que la resta d'elements detectats es mantenen en color verd, facilitant així la confirmació visual de la selecció per part de l'usuari.



```
Command: cuchillo
0: 480x640 1 knife, 1 tweezers, 1 scissors, 306.3ms
Speed: 1.0ms preprocess, 306.3ms inference, 1.0ms postprocess per image at shape (1, 3, 480, 640)
Matched: KNIFE at -4.9, -51.8 cm
Enviado: X=49 Y=517 Z=8 mm | Signos: 1, 1, 1 | Bit 10 = True
```

*Figura 16: Exemple d'identificació correcta del objecte "knife" mitjançant comanda de veu "cuchillo".*



### 6.5. Prova final

En aquest darrer apartat es valida el funcionament integrat de tot el sistema, combinant les diferents tecnologies desenvolupades en els apartats anteriors: la detecció d'objectes quirúrgics mitjançant visió per computador, el reconeixement de veu per seleccionar l'objecte desitjat i la transmissió de dades al robot a través del protocol MODBUS.

El funcionament complet del sistema és el següent:

1. El professional sanitari dona una ordre per veu indicant l'instrument que vol (ex. "tissores").
2. El sistema processa l'ordre mitjançant reconeixement de veu, aplica *fuzzy matching* i tradueix l'objecte a l'etiqueta utilitzada pel sistema de visió.
3. La càmera detecta tots els objectes sobre la taula i selecciona aquell que coincideix amb la comanda.
4. Es calcula el centre de l'objecte i es transforma a coordenades físiques mitjançant la matriu d'homografia.
5. Aquestes coordenades es transmeten al robot UR3e via *MODBUS*.
6. El robot es desplaça al punt indicat, recull l'objecte amb la ventosa i el trasllada a un *waypoint* de descàrrega.

Tot i que totes les etapes del procés funcionaven de manera satisfactòria, es va identificar una limitació en el sistema de recollida física dels objectes. Concretament, la ventosa emprada requeria que la seva superfície quedés completament coberta per l'objecte per generar prou succió. Això no es complia en el cas dels estris quirúrgics seleccionats (bisturí, pinces i tissores), ja que tenien una geometria massa prima o estreta.

Com a solució provisional, es van afegir petits cartrons als instruments per augmentar la seva superfície de contacte. Tot i així, aquest canvi va provocar dificultats en el reconeixement visual, ja que els objectes modificats ja no coincidien del tot amb les etiquetes originals del dataset entrenat.

Finalment, el sistema només va aconseguir completar amb èxit tot el cicle amb les tisores, que tenien una forma més ampla i estable, la qual cosa va permetre que fossin tant correctament detectades pel model com recollides per la ventosa sense modificacions.

## 7. Conclusions

Aquest capítol té com a objectiu presentar, de manera clara i sintètica, les principals conclusions derivades del desenvolupament del projecte. Es valora el grau d'acompliment dels objectius definits a la introducció, es reflexiona sobre la seva aportació als Objectius de Desenvolupament Sostenible (ODS) de les Nacions Unides, i es proposen possibles línies de millora i treball futur que podrien ampliar l'aplicabilitat i robustesa del sistema desenvolupat.

L'exposició s'allunya de la descripció detallada del contingut tècnic ja presentat, i posa el focus en els resultats globals, la seva rellevància pràctica i el potencial que pot tenir aquest projecte en entorns clínics simulats, educatius o d'investigació futura.

### 7.1. Acompliment dels objectius

L'objectiu principal del projecte era desenvolupar un sistema capaç de detectar visualment instruments quirúrgics i interactuar amb ells mitjançant un robot col·laboratiu controlat per veu. Aquest objectiu s'ha complert parcialment amb èxit, integrant correctament totes les tecnologies implicades: càmera, reconeixement per veu, detecció amb *YOLO*, comunicació *MODBUS* i programació del robot UR3e.

Tot i això, cal remarcar que alguns aspectes tècnics presenten marge de millora. En concret, la homografia obtinguda entre la càmera i el pla de treball ha resultat ser raonablement precisa però no totalment exacta. La desviació mitjana entre punts reals i projectats ha estat d'aproximadament 3–5 mm, la qual cosa és suficient per a una detecció general, però pot comprometre la precisió de la recollida d'objectes petits o prims com les pinces quirúrgiques.

Malgrat aquesta limitació, el sistema ha estat capaç de:

- Reconèixer amb fiabilitat els instruments presents sobre la taula.

- Interpretar ordres verbals per seleccionar-ne un.
- Convertir les coordenades de píxel a coordenades físiques del robot.
- Transmetre aquestes dades correctament via *MODBUS*.
- Executar moviments cap a l'objecte seleccionat.

Aquesta integració funcional demostra que el sistema proposat és viable com a prototip de prova de concepte, complint la majoria dels objectius inicials.

## 7.2. Contribució als objectius de Desenvolupament Sostenible (ODS)

Aquest projecte contribueix de forma directa als següents Objectius de Desenvolupament Sostenible definits per l'Agenda 2030 de les Nacions Unides:

- ODS 3 – Salut i benestar: Proposa una eina que pot augmentar la seguretat, eficiència i higiene en entorns quirúrgics, mitjançant l'automatització de tasques rutinàries i la reducció de l'exposició directa del personal sanitari.
- ODS 4 – Educació de qualitat: El sistema podria aplicar-se també a entorns formatius per simular cirurgies i practicar tècniques de control robòtic assistit per veu.
- ODS 9 – Indústria, innovació i infraestructura: Es tracta d'un exemple clar de transferència tecnològica entre la robòtica industrial i el sector mèdic, afavorint la innovació aplicada.

## 7.3. Línies de millora i treball futur

Tot i que els resultats obtinguts han estat satisfactoris i han permès validar el concepte plantejat, existeixen diverses línies de millora que podrien augmentar la robustesa i escalabilitat del sistema:

- Optimització de la transformació càmera-plà real: Tot i que la homografia calculada ha resultat prou precisa per a la majoria de casos, una millora en el procés de calibratge —com ara l'ús de més punts de control o patrons de

calibratge més precisos— podria reduir encara més l'error entre les coordenades projectades i les reals.

- Refinament del reconeixement per veu: Tot i la seva efectivitat, el sistema podria beneficiar-se d'un ajust més fi de les comandes reconegudes, així com d'una millor gestió de sinònims o variacions lèxiques mitjançant tècniques de processament del llenguatge natural.
- Millora en el mecanisme de recollida d'objectes: La ventosa emprada ha resultat funcional, però limitada en casos d'estrís molt prims. Explorar altres sistemes com pinces adaptatives o mecanismes híbrids podria millorar la fiabilitat en entorns reals, tot mantenint la compatibilitat amb el sistema de reconeixement.
- Augment de la variabilitat del dataset: Entrenar el model amb una base de dades més extensa, que inclogui variacions d'il·luminació, orientació i modificacions físiques dels objectes, permetria assolir una detecció més robusta i generalitzable.

Aquestes línies de treball podrien reforçar el prototip actual i obrir la porta a futures aplicacions tant en entorns clínics reals com en l'àmbit educatiu o experimental.

## Referencias

- [1] A. Kochan, «Scalpel please, robot: Penelope's debut in the operating room,» *Industrial Robot*, 2005.
- [2] A. K. e. a. Vaclav Snasel, «Approximate String Matching by Fuzzy Automata,» Springer Berlin Heidelberg, 2009.
- [3] a. S. D. S. Ibrahim Abdelkader Salama, «Utility of a Voice-Activated System in Minimally Invasive Surgery,» *Journal of Laparoendoscopic & Advanced Surgical Techniques*, vol. 15, nº 5, 2005.
- [4] A. P. W. Hohenberger, «Integrated operation systems and voice recognition in minimally,» *Surgical Endoscopy*, vol. 27, nº 2, 2012.
- [5] L. J. S. M. L. e. a. Wagner, «Robotic scrub nurse to anticipate surgical instruments based on real-time laparoscopic video analysis,» *Communications medicine*, nº 156, 2024.

- [6] U. K. u. Z. a. P. P. Atal Anil Kumar, Collaborative Robots, 2024.
- [7] G. Michalos, «"Robotic Systems for Human-Robot Cooperation in Industrial and Medical Applications." Journal of Manufacturing Systems.,» 2015.
- [8] M. A. & C. J. E. Peshkin, «"Cobots". Transactions on Robotics and Automation.,» 1999.
- [9] M. R. Pedersen, «"A Survey of Collaborative Robots: Applications and Future Perspectives." ,» *Journal of Robotics* , 2016.
- [10] J. M. V.-G. A. H. & Á.-V. J. C. (. Dávila-Vilchis, «Una visión de la robótica médica y sus contribuciones,» *Revista Sociedad & Tecnología*, vol. 8, nº 1, pp. 39-58, 2025.
- [11] S. E. B. S. M. I. a. A. D. Yulun Wang, «The developing Market for Medical Robotics,» *Invited Paper* , 2025.
- [12] A. A. G. e. al., «The Advances in Computer Vision That Are Enabling More Autonomous Actions in Surgery: A Systematic Review of the Literature,» *Sensors*, vol. 22, nº 13, 2022.
- [13] S. O'Meara, «MEDICAL ROBOTICS,» *Springer*, vol. 582, nº 51, 2020.
- [14] M. S. e. al., «Robotic Scrub Nurse: Surgical Instrument Handling with a Granular Jamming Gripper,» *DE GRUYTER* , vol. 9, nº 1, pp. 174-177, 2023.
- [15] OR MANAGER, «Penelope, the robot, could scrub in,» *OR MANAGER*, vol. 20, nº 3, 2004.
- [16] Y.-T. L. J. P. W. Mithun G. Jacob, «Gestonurse: A Multimodal Robotic Scrub Nurse,» West Lafayette, 2012.
- [17] e. a. Sherif M. Khairallah, «Transferring Surgical Expertise: Analyzing the Learning Curve of Robotic Cardiac Surgery Operative Time Reduction When Surgeon Moves from One Experienced Center to Another,» *Journal of Cardiovascular Development and Disease*, vol. 11, nº 3, 2024.
- [18] e. a. J. krüger, «Cooperation of human and machines in assembly lines,» *Cirp Annals*, vol. 58, nº 2, pp. 628-646, 2009.
- [19] «Universal Robots,» May 2023. [En línea]. Available: [https://www.universal-robots.com/media/1827367/05\\_2023\\_collective\\_data-sheet.pdf](https://www.universal-robots.com/media/1827367/05_2023_collective_data-sheet.pdf). [Último acceso: 20 Juky 2025].
- [20] A. D., «Real Python,» 2017. [En línea]. Available: <https://realpython.com/python-speech-recognition/#how-speech-recognition-works-an-overview>.
- [21] R. L. R., «A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition,» *Proceedings of the IEEE*, 2002.
- [22] e. a. Glenn Jocher, «Ultralytics,» 2024. [En línea]. Available: <https://docs.ultralytics.com/es/#the-evolution-of-object-detection>.
- [23] Z. Keita, «Datacamp: An Introduction to Convolutional Neural Networks (CNNs),» 14 November 2023. [En línea]. Available: <https://www.datacamp.com/tutorial/introduction->

to-convolutional-neural-networks-cnns?utm\_cid=19589720821&utm\_aid=157156374951&utm\_campaign=230119\_1-ps-other%7Edsa%7Etofu\_2-b2c\_3-emea\_4-prc\_5-na\_6-na\_7-le\_8-pdsh-go\_9-nb-e\_10-na\_11-na&utm\_loc=9198192-&utm\_m.

- [24] U. Robots, The URScript Programming Language for e-Series, 2021.
- [25] Pymodbus, «Welcome to PyModbus's documentation!», [En línea]. Available: <https://pymodbus.readthedocs.io/en/latest/>. [Último acceso: 20 July 2025].
- [26] Parlamento Europeo y del Consejo, REQUISITOS RELATIVOS AL DISEÑO Y LA FABRICACIÓN: Reglamento (UE) nº 2017/745, Diario Oficial, 2017.
- [27] IEEE STANDARDS ASSOCIATION, «IEEE STANDARDS ASSOCIATION,» 03 February 2020. [En línea]. Available: <https://standards.ieee.org/ieee/269/5674/>. [Último acceso: 21 July 2025].
- [28] M. M. a. M. M. A. M. P. M. Julie M. Clanahan, «How Does Robotic-Assisted Surgery Change OR Safety Culture?,» August 2023. [En línea]. Available: [https://journalofethics.ama-assn.org/article/how-does-robotic-assisted-surgery-change-or-safety-culture/2023-08?utm\\_source=chatgpt.com](https://journalofethics.ama-assn.org/article/how-does-robotic-assisted-surgery-change-or-safety-culture/2023-08?utm_source=chatgpt.com).
- [29] A. A. e. al., «Progress and Prospects of the Human-Robot Collaboration,» 2018.
- [30] e. a. Tenber Grota, «Current methods of nurse-surgeon training & education,» *International Journal of Nursing Studies Advances*, 2021.
- [31] S. S. H. SILVANA, «ROL DE LA ENFERMERA INSTRUMENTISTA DURANTE EL PROCEDIMIENTO QUIRÚRGICO Y SU IMPORTANCIA: REVISIÓN BIBLIOGRÁFICA,» Ecuador, 2023.
- [32] e. a. M<sup>a</sup> Cristina Chaparro velasco, «Robótica quirúrgica, desde los grandes,» Pereira, 2016.
- [33] I. V. M. Drs. OCTAVIO A. CASTILLO C., «Cirugía robótica,» *Cirugia al Día*, vol. 64, nº 1, 2012.
- [34] P. d. S.-O. I. G.-M. V. F. M. Belén Estebanez, «Interfaz multimodal para un asistente robótico quirúrgico: uso de reconocimiento de maniobras quirúrgicas.,» *Revista Iberoamericana de Automática e Informática Industrial*, vol. 8, nº 2, 2011.
- [35] H. P. H. S. & S. C. Merry D. Constable, «Enhancing surgical performance in cardiothoracic surgery with innovations from computer vision and artificial intelligence: a narrative review,» *Journal of Cardiothoracic Surgery*, vol. 19, nº 94, 2024.
- [36] «Intuitive,» 2023. [En línea]. Available: <https://www.intuitive.com/en-us/products-and-services/da-vinci>. [Último acceso: 2025].
- [37] C. Stokel-Walker, «Symani Surgical System,» *TIME MAGAZINE*, 2024.

## ANNEXOS

### A. Planificacion

Aquest annex mostra la planificació temporal del projecte, des de la seva proposta inicial fins a la redacció d'aquesta memòria. La taula següent reflecteix les diferents fases del desenvolupament i ha servit com a guia per estructurar el treball i assolir els objectius en el termini establert. Aquesta planificació ha permès una millor organització de les tasques, garantint un progrés progressiu i coherent al llarg del temps.

#### A.1. Taula de planificació detallada

Tasca	Inici	Fi	Durada estimada	Recursos emprats
Proposta inicial	26/11/24	11/12/24	16h	Reunions amb tutor i co-tutor, redacció del document de proposta
Estudi previ i definició d'objectius	23/12/24	05/01/25	28h	Ordinador personal, recerca d'articles, estudis de mercat, consulta de TFG/TFM similars, converses amb companys
Recerca de conceptes i calibració de càmeres, visió per objectes i veu	19/02/25	20/05/25	50h	Ordinador personal, càmera Logitech, micròfon, Python amb llibreries, proves de codi a casa, reunions amb tutor, disseny del sistema de reconeixement i comunicació
Desenvolupament i proves funcionals	10/07/25	23/07/25	50h	Robot UR3e, sistema de càmera muntat al laboratori, connexió Modbus, proves amb el cobot, calibració final, feedback del tutor

Integració i redacció de la memòria	30/06/25	23/07/25	70h	Ordinador personal, fotos del sistema i codi, redacció guiada pel tutor, correccions constants, organització de materials i annexos
-------------------------------------	----------	----------	-----	---

*Taula 3: Planificació del projecte i tasques principals realitzades.*

A.2.

Recursos i

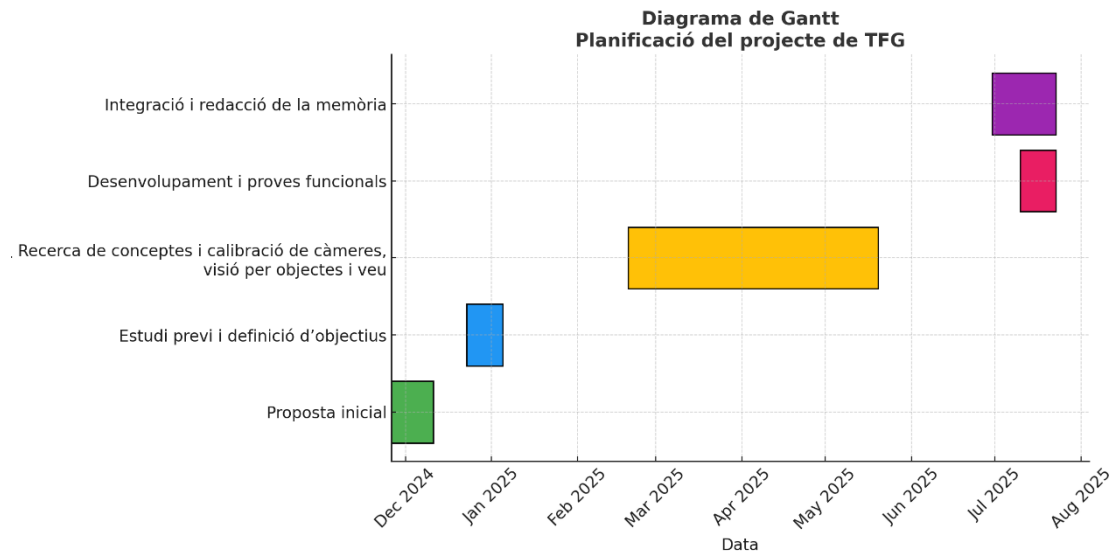
equipament utilitzat

- **Infraestructura:** Ordinador personal amb sistema operatiu Windows i accés al laboratori de robòtica de la Universitat de Girona.
- **Programari:** Python 3.12.5, amb les biblioteques *speech\_recognition*, *pyModbus*, *cv2*, *ultralytics*, *threading*, *difflib*, , així com protocols de comunicació com TCP/IP. Entorn de desenvolupament: Visual Studio Code.
- **Suport:** Supervisió per part del tutor i co-tutor, així com assessorament puntual del professor Arnau Oliver en temes relacionats amb el reconeixement per visió per computador.
- **Material emprat:** Càmera Logitech C270, auriculars SONY WH-XB910N amb micròfon integrat, cobot UR3e, trípode metàl·lic per muntatge de la càmera i taulell d'escacs per a la calibració.

### A.3. Diagrama de Gantt

La figura següent presenta el diagrama de Gantt que reflecteix la planificació temporal del projecte, amb la seva divisió en diferents fases i la durada prevista per a cadascuna de les activitats.





*Figura 17: Diagrama de Gantt del projecte.*

#### A.4. Desviacions respecte a la planificació inicial

En general, l'evolució del projecte s'ha ajustat de manera satisfactòria a la planificació establerta en les primeres fases. Les hores dedicades a cada etapa i el calendari global han estat força coherents amb les previsions inicials, fet que ha permès assolir els objectius dins del termini requerit.

Tanmateix, cal destacar una lleugera desviació temporal entre la fase de recerca de conceptes i desenvolupament del sistema i les etapes posteriors de proves i integració. Aquesta separació es deu principalment a motius personals, ja que durant un període determinat vaig estar fent el programa Erasmus, cosa que va dificultar l'accés a l'equipament del laboratori i va posposar temporalment el treball amb el robot físic.

Tot i aquesta interrupció, el treball de recerca i desenvolupament del codi va continuar de forma remota, cosa que va permetre reprendre el projecte de manera àgil un cop retornat a la Universitat.

## B. Codi i manual

Amb la finalitat de facilitar l'accés obert al codi desenvolupat durant aquest Treball de Fi de Grau, s'ha habilitat un repositori públic a la plataforma GitHub. Aquest espai recull tots els fitxers necessaris per a la implementació, prova i posada en funcionament del sistema de control robòtic mitjançant reconeixement de veu dissenyat durant el projecte.

[https://github.com/nilleslie/TFG\\_-\\_Surgical\\_Instrument\\_Server](https://github.com/nilleslie/TFG_-_Surgical_Instrument_Server)

El repositori incorpora un fitxer README.md a l'arrel, on s'explica breument:

- L'objectiu del projecte i l'organització del codi.
- Els requisits del sistema i les instruccions bàsiques d'instal·lació.
- Referències als manuals d'instal·lació i ús.

A més, s'hi poden trobar:

- Els scripts en Python per al reconeixement de veu i l'enviament de dades via socket.
- El conjunt de fitxers corresponents a la programació del cobot Ur3e.
- El contingut dels manuals, ubicats a la carpeta docs/, on es poden consultar en detall les instruccions per a instal·lar, configurar i utilitzar el sistema.

Tot el material està organitzat i comentat per tal de facilitar-ne la reutilització i possible ampliació en projectes futurs.

## C. Pressupost

Aquest apartat de l'annex ajunta totes les estimacions econòmiques associades al desenvolupament d'un projecte com aquest. S'hi contemplen tant les despeses derivades dels béns materials emprats com les relacionades amb el desenvolupament del codi, així com altres recursos rellevants per al correcte funcionament del sistema. L'objectiu és oferir una visió clara i estructurada del cost total del projecte, tot diferenciant els elements individuals i agrupant-los posteriorment en partides parcials i totals.

## C.1. Preus unitaris

Element	Unitat	Preu unitari (€)	Observacions
<b>Robot UR3e</b>	1 unitat	22.000 €	Equipament cedit del laboratori
<b>Ordinador portàtil</b>	1 unitat	770€	Ordinador utilitzat per desenvolupar i executar el codi
<b>Micròfon</b>	1 unitat	136€	Auriculars multifuncionals
<b>Ventosa <i>Epick</i></b>	1 unitat	5.000€	-
<b>Altres materials</b>	1 unitat	50€	Material auxiliar per a la simulació
<b>Desenvolupament de codi Python</b>	50 hores	35€/h	Visió per computador, client TCP/IP, reconeixement per veu
<b>Desenvolupament del codi interfície gràfica</b>	50 hores	35€/h	Control robòtic des de la interfície Polyscope
<b>Redacció de la memòria i manuals</b>	70 hores	35€/h	Annexos, memòria, manuals d'usuari i d'instal·lació
<b>Despeses generals</b>	-	5% del subtotal	Llum, ús d'equipament, infraestructures
<b>Benefici industrial i imprevistos</b>	-	10% del subtotal	Estimació habitual segons criteris tècnics

Taula 4: Estimació de preus unitaris

## C.2. Pressupostos parcials

A continuació es presenta la valoració econòmica del projecte, desglossada i agrupada per partides principals:

- Materials i equipament:
  - Robot UR3e: 22.000 €
  - Ordinador: 770 €

- Micròfon (estimació): 136 €
- Ventosa (no valorada econòmicament)
- Altres (tissores, pinces, trípode, taulells): 50 €
- Total: 22.956 €
- Desenvolupament i documentació:
  - Desenvolupament de codi Python (50 h x 35 €/h): 1.750 €
  - Desenvolupament interfície gràfica (50 h x 35 €/h): 1.750 €
  - Redacció de la memòria i manuals (70 h x 35 €/h): 2.450 €
  - Total: 5.950 €
- Despeses generals (5%):
 

5% de (22.956 € + 5.950 €) = 1.445,3 €
- Benefici industrial i imprevistos (10%):
 

10% de (22.956 € + 5.950 €) = 2.890,6 €

### C.3. Pressupostos totals

A partir de les partides anteriors, el cost total estimat del projecte és el següent:

CONCEPTE	IMPORT
Total estimat	33.241,90

*Taula 5: Pressupost final estimat*

#### C.4. Comentari final

Tot i que gran part de l'equipament emprat en aquest projecte ha estat cedit pel laboratori de robòtica de la universitat (com el robot UR3e o l'espai físic de treball), s'ha volgut fer una estimació realista del cost que comportaria dur a terme una implementació similar en un entorn professional o industrial.

Així doncs, aquest pressupost reflecteix el valor econòmic aproximat del desenvolupament complet d'un sistema robòtic amb reconeixement de veu i visió artificial, tenint en compte tant el cost del material com les hores dedicades al desenvolupament del programari i la documentació tècnica. La xifra final pot servir de referència per futures propostes, ampliacions o implementacions en entorns reals.

#### D. Càlcul i justificacions

##### D.1. Introducció

Aquest apartat té com a objectiu detallar i justificar els càlculs, valors i paràmetres emprats durant el desenvolupament del projecte. S'hi exposen les raons tècniques per les quals s'han seleccionat determinades magnituds o aproximacions, així com els sistemes de coordenades utilitzats per garantir la coherència i precisió de les operacions robòtiques. També es proporciona informació complementària que permet entendre millor les decisions de disseny adoptades en els apartats tècnics del treball.

##### D.2. Patro de calibratge

Per establir la relació entre les coordenades captades per la càmera i el sistema de referència físic del robot, es va calcular una matriu d'homografia. Aquesta transformació permet convertir punts 2D de la imatge en coordenades reals sobre el pla de treball, sempre que aquest sigui pla i uniforme.

Per dur a terme aquest procés, es va utilitzar un taulell d'escacs amb una graella de 20 x 14 caselles, cada una amb una mida de 20 mm, descarregat del portal de generació de patrons de calibratge:

<https://markhedleyjones.com/projects/calibration-checkerboard-collection>

Tot i que es va considerar inicialment fer servir tècniques completes de calibració de càmera amb els paràmetres interns obtinguts a través del fitxer *calibration1.npz*, finalment es va optar per un enfocament més senzill i directe basat només en homografia, donat que el sistema treballava en un pla bidimensional.

La matriu d'homografia calculada es troba guardada en el fitxer *Homografia.npy*, i conté la transformació projectiva necessària per passar de coordenades en píxels a coordenades físiques (en centímetres), assumint que tots els punts es troben al mateix pla. Aquesta transformació s'aplica automàticament cada cop que el sistema detecta un instrument quirúrgic per tal de traduir la seva posició dins l'espai de treball real.

Aquest mètode va demostrar una precisió suficient per les necessitats del projecte, tot i que presenta certes limitacions en entorns tridimensionals o amb canvis d'altura notables. En futurs desenvolupaments es podria combinar amb tècniques de visió en 3D o amb models de càmera més precisos per millorar la robustesa de la localització espacial.

### D.3. Bounding Boxes

Per detectar els instruments quirúrgics a la imatge, el sistema de visió genera automàticament una *bounding box* al voltant de cada objecte. Aquesta caixa es defineix pels valors de posició ( $x$ ,  $y$ ), amplada ( $w$ ) i alçada ( $h$ ). A partir d'aquests, es calcula el centre de l'objecte amb la següent fórmula:

$$x_c = x + \frac{w}{2}, \quad y_c = y + \frac{h}{2}$$

*Equació 1: Càlcul centre del Bounding Box*

En aquest projecte s'ha optat per treballar amb centímetres, ja que es tracta d'una unitat intermèdia raonable: proporciona la precisió suficient per ubicar els objectes dins d'un espai de treball de mida A3 (297 × 420 mm) i, alhora, evita els problemes que poden sorgir d'una precisió excessiva (com el mil·límetre) que no aportaria beneficis

significatius però sí més complexitat i possible soroll numèric. A més, aquesta unitat és perfectament compatible amb els sistemes de coordenades del robot UR3e.

Aquesta estratègia és senzilla però efectiva, ja que permet localitzar l'objecte amb precisió suficient per a la manipulació robòtica. El model *YOLOv8* utilitzat es va configurar amb un llindar de confiança de 0.82, valor que assegura una detecció fiable i minimitza els falsos positius.

#### D.4. Conclusió i aplicació final

La fase final del projecte consisteix en la implementació directa del control del robot UR3e mitjançant el sistema de coordenades obtingut per visió i comandes de veu. El codi es va desenvolupar íntegrament a l'entorn *Polyscope*, emprant blocs de programació bàsics per a moviment, lectura de registres *MODBUS* i control de flux.

Es van definir diversos *waypoints* per estructurar el moviment del robot de forma segura i eficient:

- **home**: posició de repòs segura, que també correspon a la posició de *home* predefinida del robot.
- **approach**: posició intermèdia que apropa el braç a la zona de treball abans de la recollida.
- **p\_tool**: coordenada generada a partir del sistema de visió, que indica la ubicació exacta de l'objecte quirúrgic.
- **tray**: posició final on es diposita l'objecte recollit, simulant una safata quirúrgica.

El programa manté el robot en repòs dins d'un bucle principal (loop) fins que es detecta un bit d'activació rebut via *MODBUS*. Un cop rebut aquest senyal, el robot llegeix els valors corresponents a les coordenades x, y, z, així com els bits de signe associats a cada component.

Aquest procediment permet compensar la limitació dels registres *MODBUS*, que només transmeten nombres positius, i garanteix una precisió adequada per a aplicacions robòtiques d'aquest tipus.

Gràcies a aquest esquema modular i controlat, es va aconseguir executar el cicle complet: detecció d'objecte, transformació de coordenades, enviament al robot, desplaçament, recollida i dipòsit. Aquest funcionament integrat constitueix la prova definitiva del sistema desenvolupat.