Simposio CEA de Robótica, Bioingeniería, Visión Artificial y Automática Marina 2025

RBVM
Simposio CEA de Robótica, Bioingeniería, Visión por Computador y Automática Marina

# Apunta y Agarra: Teleoperación Asistida de Robots en Realidad Mixta

García-Gómez, M.[a,*], Duque-Domingo, J.[a], Gómez-García-Bermejo, J.[a,b], Zalama, E.[a,b]

[a] *Departamento de Ingeniería de Sistemas y Automática, ITAP, Universidad de Valladolid, C/ Dr Mergelina s/n, 47011 Valladolid, España*
[b] *CARTIF Centro Tecnológico 47151, Valladolid, España*

**Resumen**

El presente artículo aborda la mejora en la teleoperación de robots colaborativos en entornos de Realidad Mixta. La propuesta permite a los operadores seleccionar objetos reales dentro de una interfaz virtual con un simple gesto, tras lo cual el robot realiza la acción de alcance y agarre de forma autónoma. Este enfoque facilita la manipulación de objetos en entornos tridimensionales, mejorando la eficiencia y ofreciendo un control más intuitivo, especialmente para usuarios sin experiencia en robótica. A diferencia de los métodos tradicionales, que requieren un control detallado de cada movimiento del robot, el sistema simplifica la interacción remota. La solución propuesta resulta especialmente útil en aplicaciones donde el operador no puede estar presente, aumentando la accesibilidad y la eficacia en la teleoperación de robots.

*Palabras clave:* Robot manipulador, Teleoperación, Realidad mixta, Telepresencia, Telerrobótica, Tecnología robótica

**Point and Grasp: Assisted Robot Teleoperation in Mixed Reality**

**Abstract**

This paper addresses the improvement of collaborative robot teleoperation in Mixed Reality environments. The proposal allows operators to select real objects within a virtual interface with a simple gesture, after which the robot autonomously performs the reach-and-grasp action. This approach facilitates the manipulation of objects in three-dimensional environments, improving efficiency and offering more intuitive control, especially for users with no robotics experience. Unlike traditional methods, which require detailed control of every movement of the robot, the system simplifies remote interaction. The proposed solution is especially useful in applications where the operator cannot be present, increasing accessibility and efficiency in robot teleoperation.

*Keywords:* Robots manipulators, Teleoperation, Mixed reality, Telepresence, Telerobotics, Robotics technology

## 1. Introducción

La teleoperación de robots móviles con brazos robóticos busca ampliar las capacidades humanas en entornos complejos y remotos. Tecnologías como la Realidad Virtual, Aumentada o Mixta han abierto nuevas vías para un control más intuitivo, aunque persisten desafíos en la precisión y eficiencia al manipular objetos.

El sistema tradicional de teleoperación en Realidad Mixta se basa en un control manual en el que el operador, mediante controladores de Realidad Virtual, mueve sus manos y el robot replica esos movimientos en tiempo real (Su et al., 2022; Ding et al., 2024). Este enfoque, aunque es intuitivo, puede resultar complejo cuando la percepción de profundidad es limitada. Este artículo propone una solución innovadora: un sistema de apuntar y agarrar, donde el operador simplemente apunta a un objeto que ve el robot a través de la cámara y este es capaz de realizar la acción de cogerlo de forma autónoma.

El sistema desarrollado combina una plataforma equipada con dos brazos robóticos y una cámara *RealSense*. Aunque la cámara proporciona datos de profundidad para generar nubes de puntos, la latencia en la transmisión de esta información y el ruido presente han llevado a optar por una visualización en modo 2D mediante una interfaz en *Unity*. Esta configuración limita la percepción de la profundidad, lo que puede dificultar la manipulación de ciertos objetos. Para superar esta limitación, se ha implementado un sistema que permite al usuario seleccionar un objeto en la imagen de la cámara utilizando un puntero láser en el entorno virtual. Una vez seleccionado un objeto, el sistema procesa la imagen segmentándolo mediante técnicas basadas en

*Autor para correspondencia: miguel.garcia.gomez@uva.es

*Segment Anything Model* (SAM) (Kirillov et al., 2023), genera su nube de puntos y estima parámetros críticos que permiten al robot acercarse y orientar su pinza de manera óptima, facilitando la manipulación y recogida del objeto.

Además de mejorar la experiencia de teleoperación con tecnologías inmersivas, este enfoque introduce un modelo de interacción sencillo que no requiere conocimientos técnicos avanzados. Su aplicación está especialmente orientada a contextos de asistencia remota: personas en situación de dependencia podrían recibir ayuda de cuidadores a distancia, sin necesidad de desplazamientos. Esta capacidad no solo incrementa la autonomía del usuario, sino que también reduce la carga logística del cuidador. Por ello, el sistema ha sido diseñado con un enfoque centrado en la usabilidad, buscando una teleoperación accesible, intuitiva y eficaz.

## 2. Trabajos Relacionados

La manipulación robótica en entornos no estructurados ha cobrado un creciente interés en los últimos años, impulsada por los avances en percepción tridimensional y en los sistemas de interacción humano-robot.

Las interfaces inmersivas han redefinido la forma en que los operadores interactúan con sistemas robóticos remotos, ofreciendo una percepción enriquecida del entorno y una mayor naturalidad en el control. (Omarali et al., 2020) introdujo una arquitectura que combina cámaras de profundidad, *Unity* y Realidad Virtual, permitiendo la telemanipulación remota mediante nubes de puntos y retroalimentación visual. En esta línea, (Xu et al., 2022) desarrollan una interfaz en Realidad Virtual donde el operador guía iterativamente el efector final de un brazo robótico mediante *waypoints*. La escena se visualiza con nubes de puntos y video en una pared virtual. Aunque el sistema es efectivo para tareas de *pick and place*, los autores señalan que aún puede mejorarse en fluidez y autonomía.

El sistema *Open-TeleVision* (Cheng et al., 2024) permite al operador telecontrolar un robot humanoide combinando visión estereoscópica y control de orientación de la cabeza. Esto genera una experiencia inmersiva más realista y precisa, que mejora la percepción espacial. Además, el sistema permite controlar en tiempo real tanto brazos como manos multifuncionales del robot.

En cuanto a la planificación de agarres, se han desarrollado técnicas que extraen información contenida en las nubes de puntos de los objetos para encontrar regiones de contacto estables sin necesidad de modelos del objeto diseñados por ordenador. (Gualtieri et al., 2017) desarrollaron un sistema capaz de detectar poses de agarre en escenas con alta densidad de objetos. A través de análisis de curvatura y normales el sistema es capaz de identificar zonas aptas para el agarre incluso cuando los objetos están parcialmente ocultos o solapados.

En esta línea, (Liu et al., 2022) proponen una metodología que combina el análisis de la forma del objeto con el criterio de *force closure*, descomponiendo objetos complejos en formas básicas como esferas, cilindros y prismas rectangulares. Esta aproximación permite determinar de manera eficiente las posturas óptimas de agarre, eliminando la necesidad de entrenamientos complejos.

## 3. Descripción del Sistema

El sistema, mostrado en la Figura 1, consiste en un robot con dos brazos robóticos *Kinova Gen3* montados sobre una columna con desplazamiento vertical, diseñado para instalarse sobre una plataforma móvil que le otorgue capacidad de navegación en el entorno. A continuación, se detallan los componentes principales del sistema y su integración.

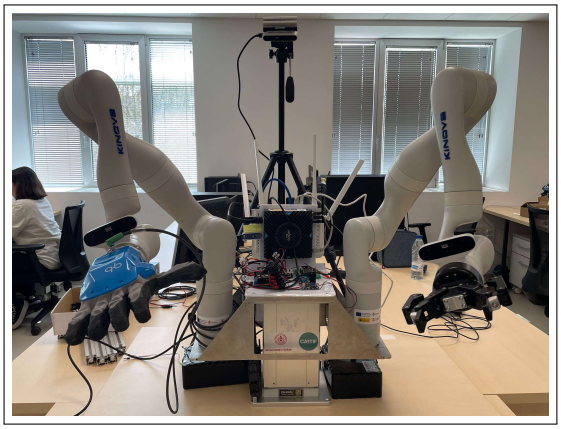

Figura 1: Configuración del sistema real.

### 3.1. *Componentes Hardware*

El núcleo del sistema está constituido por dos brazos robóticos, seleccionados por su capacidad de manipulación y su diseño compacto e integrado, con la controladora en el propio brazo. La visión se realiza mediante una cámara *RGB-D Intel RealSense D415*, que permite capturar la profundidad y generar nubes de puntos tridimensionales que representan con detalle la forma y estructura del entorno. La mano robótica *SoftHand2* sustituye una de las pinzas estándar y permite una manipulación más versátil, ideal para objetos con geometrías complejas. Todos estos componentes están gestionados mediante un *Mini PC* montado directamente sobre el robot.

Por último, las gafas de Realidad Virtual *Meta Quest 3* constituyen el dispositivo principal para la teleoperación inmersiva del sistema.

### 3.2. *Software de Control y Visualización*

El sistema desarrollado utiliza una arquitectura que combina *Unity*, para la interfaz inmersiva que se ejecutará en el dispositivo de Realidad Virtual, y *ROS*, como *middleware* para el control de los componentes robóticos. Esto permite la integración modular y la comunicación en tiempo real entre los distintos componentes del sistema. A través de *topics* y servicios, *ROS* conecta los datos provenientes de los brazos robóticos y la cámara con el sistema de visualización en *Unity*.

### 3.3. *Entorno Virtual*

El entorno virtual ha sido desarrollado en *Unity* con el objetivo de replicar en tiempo real la información visual captada por la cámara. Dentro del espacio tridimensional del visor de Realidad Virtual, se ha integrado una pantalla flotante que permite al operador visualizar la escena en formato 2D.

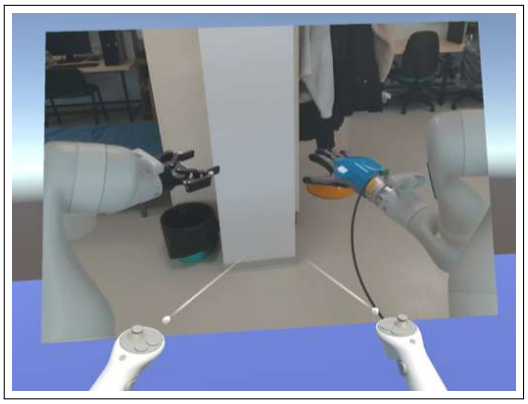

Figura 2: Entorno virtual en modo Teleoperación

En este mundo virtual, el usuario puede controlar los brazos del robot real mediante el movimiento de los controladores de las gafas. Además, se ha desarrollado un modo de teleoperación asistida, en el cual el usuario tiene la capacidad de seleccionar objetos visibles por el robot simplemente apuntando hacia ellos en la pantalla. Una vez seleccionado el objeto, el robot procede automáticamente a recogerlo.

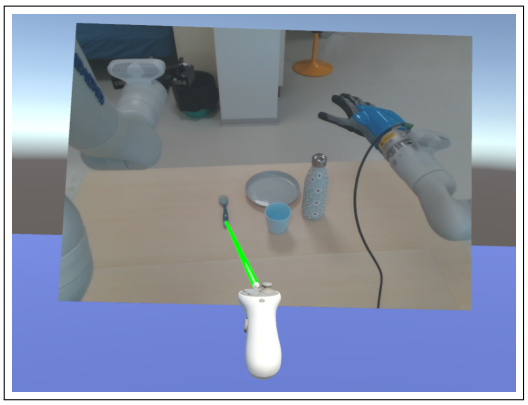

Figura 3: Entorno virtual en modo Teleoperación Asistida

Las Figuras 2 y 3 muestran vistas representativas de este entorno virtual. En la Figura 2 se observa la perspectiva del usuario al teleoperar el robot mediante movimientos manuales. En la Figura 3, el usuario ha pasado al modo de teleoperación asistida, en el cual un rayo proyectado desde el controlador permite seleccionar objetos que aparezcan en la pantalla.

### *3.4. Comunicación entre el usuario y el robot*

El sistema de teleoperación se basa en una arquitectura distribuida que separa el procesamiento gráfico de la interfaz inmersiva del control del robot.

La aplicación desarrollada en *Unity* se ejecuta en el dispositivo de Realidad Virtual. Esta aplicación actúa como cliente visual del sistema, permitiendo al operador visualizar la escena y enviar comandos gestuales mediante los controladores de mano.

Por otro lado, el *ROS Master* se ejecuta en el PC del robot. Este equipo centraliza el control de todos los componentes del robot: brazos robóticos, cámara y manipulación de la mano robótica. Este *Mini PC* opera como nodo maestro dentro del ecosistema *ROS*, gestionando la comunicación entre dispositivos y coordinando la ejecución de comandos.

La comunicación entre ambos sistemas se establece mediante una red local *WiFi*. En esta configuración, *Unity* actúa como nodo de *ROS*, transmitiendo eventos de interacción a través de diferentes *topics*. Por un lado, puede enviar las velocidades asociadas al movimiento de los controladores para que las manos del robot las imiten en tiempo real. Por otro lado, si el usuario desea agarrar un objeto de manera automática, puede señalar un objeto directamente en la pantalla flotante; *Unity* transmite la posición del píxel seleccionado, y el *Mini PC* interpreta esta información para ejecutar la acción correspondiente.

El esquema de conexiones del sistema se puede observar en la Figura 4. Este esquema permite una teleoperación con baja latencia en la visualización y respuesta rápida ante eventos del operador. Además, separa responsabilidades entre interfaz e inteligencia de control, lo que facilita el mantenimiento y escalabilidad del sistema.

## 4. Metodología

La metodología se presenta en dos partes diferenciadas: en primer lugar, se describe el proceso de teleoperación manual del robot; y en segundo lugar, se detalla el sistema desarrollado para simplificar la interacción del usuario y facilitar la recogida de objetos.

### *4.1. Teleoperación*

El sistema de teleoperación ha sido diseñado para ofrecer una experiencia inmersiva e intuitiva, aprovechando las capacidades de las gafas de Realidad Virtual *Meta Quest 3*. El usuario sostiene un controlador en cada mano, y el sistema calcula en tiempo real las velocidades lineales y angulares de sus movimientos.

Esta información se convierte en comandos que permiten mover de forma precisa los efectores finales de cada brazo robótico, replicando los gestos del usuario. Cada controlador está vinculado a un brazo robótico, lo que posibilita una teleoperación bimanual natural.

#### *4.1.1. Control de movimiento y suavizado*

El control del movimiento de los brazos se basa en la cinemática diferencial, la cual establece la relación entre las velocidades del efector final y las velocidades articulares del robot. Esta metodología resulta ideal para entornos dinámicos, ya que permite una adaptación continua a los comandos del operador. Dado un conjunto de velocidades deseadas $[\dot{X}]$ en el espacio cartesiano (posición y orientación), la cinemática diferencial emplea el Jacobiano analítico $[J(q)]$ para obtener las velocidades articulares $[\dot{q}]$ mediante la Ecuación 1.

$$[\dot{q}] = [J(q)]^{-1} \cdot [\dot{X}] \tag{1}$$

Esto permite que el robot replique directamente los movimientos del controlador en tiempo real, generando una interacción continua, fluida y con bajo retardo. La ventaja clave de este enfoque frente a métodos basados en trayectorias discretas es que no requiere precomputar puntos intermedios, lo cual resulta especialmente útil en tareas de manipulación reactiva o en entornos desconocidos.

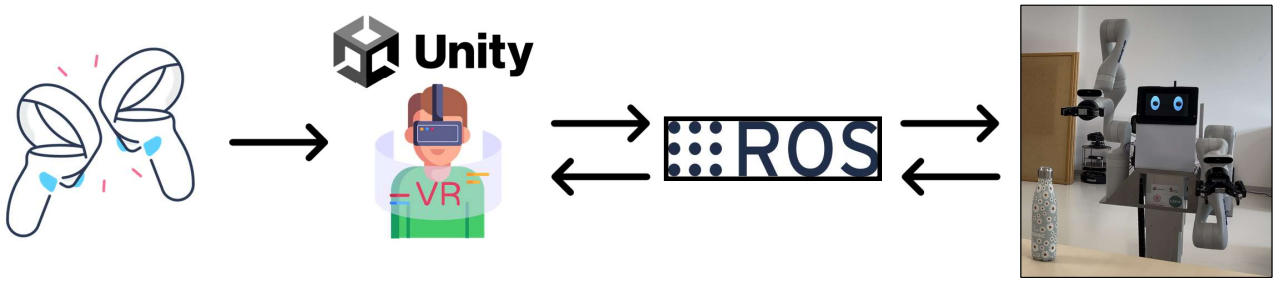


Figura 4: Esquema del sistema. Las gafas se encargan de realizar el seguimiento de los mandos. La aplicación de *Unity* manda los comandos vía WiFi al PC con ROS, encargado de gestionarlos. Desde ROS se envían los mensajes de la cámara a *Unity*.

Para lograr un movimiento estable de los brazos, especialmente frente a pequeñas vibraciones del usuario cuando la mano del usuario permanece casi estática, se ha implementado un suavizado exponencial simple sobre las velocidades angulares $[\dot{\omega}]$. Esto evita oscilaciones no deseadas en los efectores y permite una operación más precisa durante tareas de manipulación fina. Basado en (García-Gómez et al., 2024), la ecuación utilizada para el suavizado es la mostrada en la Ecuacion 2.

$$[\dot{\omega}]'_t = \alpha \cdot [\dot{\omega}]_t + (1 - \alpha)[\dot{\omega}]'_{t-1} \tag{2}$$

donde:

- $[\dot{\omega}]'_t$ es la velocidad angular suavizada en el tiempo $t$, es decir, el valor que efectivamente se usará para mover el brazo.
- $[\dot{\omega}]_t$ es la velocidad angular medida en el instante actual $t$, es decir, el valor crudo que viene del mando.
- $\alpha$ es el factor de suavizado, que controla cuánto peso se les da a los datos más recientes. Ha sido ajustado a 0.5 como valor óptimo.

### 4.2. *Apunta y Agarra*

Uno de los principales desafíos en la teleoperación robótica es la estimación precisa de la pose de un objeto cuando el operador solo dispone de una visualización monocular en 2D. Para mitigar esta limitación, se ha desarrollado un sistema que permite al usuario seleccionar un objeto en la pantalla de *Unity*, apuntando con el mando de las gafas a un píxel de interés dentro de la imagen proyectada por la cámara.

Figura 5: Elección de objeto (izquierda) y segmentación (derecha)

El píxel seleccionado representa el punto por el que se quiere agarrar el objeto deseado, y sus coordenadas $(x, y)$ son enviadas desde *Unity* al PC de control del robot mediante un tópico de ROS. Una vez recibidas las coordenadas del píxel, el PC del robot obtiene la última imagen publicada por la cámara. Esta imagen se utiliza como entrada para el modelo SAM, que segmenta automáticamente el objeto que contiene dicho píxel. Se ha empleado el modelo con *backbone ViT-L* (*Vision Transformer - Large*), el cual ofrece un equilibrio adecuado entre precisión y tiempos de procesamiento en comparación con las otras variantes disponibles. En la Figura 5 se puede observar, a la izquierda, al usuario eligiendo un punto de la botella y, a la derecha, la segmentación de la botella realizada por SAM.

#### 4.2.1. *Conversión de Píxeles a Coordenadas 3D*

A partir del píxel seleccionado $(x, y)$ en la imagen y del valor de profundidad $Z$ correspondiente, se realiza la conversión al espacio 3D. Considerando los parámetros intrínsecos de la cámara (distancias focales $f_x$, $f_y$ y centro óptico $(c_x, c_y)$), la posición tridimensional del punto en el sistema de la cámara se calcula como:

$$X = \frac{(x - c_x) \cdot Z}{f_x}, \quad Y = \frac{(y - c_y) \cdot Z}{f_y}, \quad Z = Z \tag{3}$$

Consiguiendo la posición 3D del punto con respecto a la cámara:

$$\mathbf{p}_{\text{cam}} = \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} \tag{4}$$

#### 4.2.2. *Segmentación Automática con SAM*

La segmentación genera una máscara binaria que define el contorno del objeto seleccionado. Esta máscara permite extraer los puntos 3D correspondientes del mapa de profundidad sincronizado que genera la cámara. A partir de esto, se construye una nube de puntos tridimensional del objeto.

La nube de puntos segmentada es filtrada y suavizada para eliminar el ruido y las inconsistencias espaciales. Posteriormente, se procede al análisis geométrico de la misma. Mediante análisis de componentes principales, se estiman:

- La **normal** $\vec{n}$ de la superficie en el punto de interés.
- El **eje principal de inercia** $\vec{i}$, que representa la dirección longitudinal dominante del objeto.

En la Figura 6, estos vectores se han representado sobre la nube de puntos: la flecha roja indica la normal y la flecha azul, el eje de inercia del objeto.

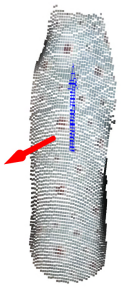

Figura 6: Normal del objeto en el punto seleccionado (flecha roja) y eje de inercia del objeto (flecha azul) en nube de puntos

Basándose en (Suzuki and Oka, 2016), estos vectores se han utilizado para definir la pose de agarre deseada: la normal determina la dirección de aproximación, y el eje de inercia se utiliza para alinear la orientación del efector final del robot, de modo que la pinza quede perpendicular a dicho eje. Esta estrategia de orientación maximiza la estabilidad del agarre en objetos alargados o simétricos.

### *4.3. Ejecución del movimiento*

Una vez obtenida la pose objetivo del objeto, el sistema realiza una serie de pasos que permiten al robot alcanzar y manipular dicho objeto de forma precisa.

#### *4.3.1. Transformación al Sistema del Robot*

Tanto la posición como los vectores direccionales calculados se encuentran en el marco de referencia de la cámara. Para que el robot pueda interactuar con el objeto, se transforman al sistema de referencia de la base del robot utilizando una matriz de transformación homogénea $\mathbf{T}_{\text{cam}\to\text{base}} \in \mathbb{R}^{4\times4}$, obtenida a partir del sistema *TF* de ROS:

$$\mathbf{p}_{\text{robot}} = \mathbf{T}_{\text{cam}\to\text{base}} \cdot \mathbf{p}_{\text{cam}} \tag{5}$$

Los vectores de dirección se transforman aplicando únicamente la submatriz de rotación $\mathbf{R} \in \mathbb{R}^{3\times3}$:

$$\vec{n}_{\text{robot}} = \mathbf{R} \cdot \vec{n}_{\text{cam}}, \quad \vec{i}_{\text{robot}} = \mathbf{R} \cdot \vec{i}_{\text{cam}} \tag{6}$$

#### *4.3.2. Construcción de la Pose de Agarre*

Para definir la orientación del efector final del robot durante el agarre, se construye una base ortonormal a partir de la información geométrica del objeto. Esta base se ajusta teniendo en cuenta la convención de ejes del efector final, así como las restricciones espaciales del sistema.

En el sistema de referencia del efector, se adopta la siguiente convención:

- El eje $\hat{z}$ apunta hacia el exterior de la pinza, es decir, en la dirección de aproximación del agarre.
- El eje $\hat{y}$ apunta hacia la cámara de visión ubicada sobre la pinza, representando la orientación vertical relativa del efector en el entorno.
- El eje $\hat{x}$ completa el sistema ortonormal mediante el producto vectorial $\hat{x} = \hat{y} \times \hat{z}$.

Con esta convención, la base ortonormal se construye a partir de las características del objeto detectado:

- El eje $\hat{z}$ del efector se alinea de forma opuesta a la normal local de la superficie del objeto:

  $$\hat{z} = -\vec{n}_{robot} \tag{7}$$

  Esto garantiza que el robot se aproxime al objeto desde una dirección frontal, favoreciendo un contacto estable.
- El eje $\hat{y}$ se alinea con el eje principal de inercia del objeto estimado mediante PCA:

  $$\hat{y} = \vec{i}_{robot} \tag{8}$$

  Este eje se reorienta, si es necesario, para mantener una orientación coherente respecto a la posición del brazo, apuntando hacia arriba o hacia delante según el contexto de la tarea.
- Finalmente, el eje $\hat{x}$ se calcula como:

  $$\hat{x} = \hat{y} \times \hat{z} \tag{9}$$

  asegurando la ortogonalidad de la matriz de rotación.

Con esta base se define la matriz de rotación deseada para el efector. Esta, junto con la posición 3D transformada al marco del robot, forma la pose de agarre en seis grados de libertad. A partir de esta información, el planificador genera una trayectoria desde la configuración actual del brazo hasta la pose objetivo, teniendo en cuenta restricciones cinemáticas y posibles colisiones.

## 5. Resultados

El sistema desarrollado ha sido sometido a pruebas ejecutando diversas tareas comunes que un cuidador podría realizar en un entorno doméstico remoto. Estas tareas incluyeron acciones sencillas como tomar un objeto y depositarlo en otra localización, y también actividades más complejas como verter agua en un vaso o acercárselo a una persona, simulando asistencia a personas en situación de dependencia.

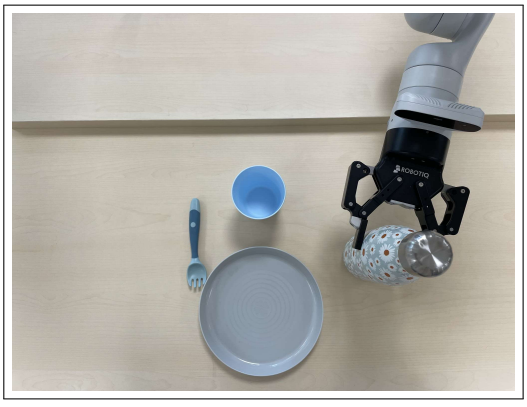

Figura 7: Agarre de objeto requerido

En estas pruebas, se evaluaron los dos modos de operación: **teleoperación manual** y **teleoperación asistida** mediante el sistema de apuntar y agarrar. Mientras que la teleoperación manual requiere mayor destreza y control por parte del operador,

el sistema asistido permite que el usuario complete tareas con menor esfuerzo. La Figura 7 muestra un ejemplo típico de la acción de agarre realizada durante las pruebas.

El índice de acierto global de la estrategia de apuntar y agarrar fue del **82 %**, destacando un **92 %** en objetos alargados como botellas, cubiertos, rotuladores y bolígrafos. Este buen desempeño se debe a la correcta estimación del eje de inercia y la orientación del efector final del robot en el momento del agarre.

En promedio, el tiempo requerido para realizar la segmentación ha sido de **3.52 segundos**, lo que permite una interacción fluida entre el operador y el robot, sin generar retrasos significativos que afecten a la experiencia del usuario. Este rendimiento ha sido posible gracias a la utilización de un servidor externo equipado con una unidad de procesamiento gráfico (GPU), encargado de procesar la segmentación de las imágenes, permitiendo superar las limitaciones del mini PC, en el cual la utilización de SAM resultaba más lento.

Se han identificado ciertas limitaciones en la ejecución de agarres cuando el objeto no presenta un eje de inercia claramente definido. En particular, objetos con geometrías de dimensiones similares en todas las direcciones, como cubos o esferas. En estos casos, es necesario recurrir al modo de teleoperación manual. Asimismo, se ha observado que el algoritmo de estimación de la normal puede, en algunos casos, calcular una dirección orientada hacia el interior del objeto. Esta inversión de la normal compromete la aproximación del efector final, impidiendo una maniobra de agarre efectiva.

## 6. Conclusiones

Los resultados obtenidos demuestran que la integración de teleoperación inmersiva con planificación automática de agarres constituye una estrategia eficaz para la manipulación remota de objetos. En particular, el modo asistido de apuntar y agarrar representa un avance significativo frente a esquemas tradicionales de teleoperación manual, al reducir la carga operativa y facilitar la ejecución de tareas. Este modo muestra un rendimiento notable al trabajar con objetos que poseen geometrías bien definidas, con estructuras alargadas o ejes de inercia marcados.

Por otra parte, este sistema puede utilizarse como herramienta para generar datos, lo cual es especialmente útil en contextos de aprendizaje por demostración (Kamijo et al., 2024; Iyer et al., 2024; Duque-Domingo et al., 2024a,b).

Entre las principales líneas de mejora se encuentra la optimización del sistema con la incorporación de redes de segmentación más rápidas, como *FastSAM* (Zhao et al., 2023), podría reducir los tiempos de segmentación, facilitando una respuesta más ágil del sistema. Como trabajo futuro, se plantea implementar visualización estereoscópica para aumentar la capacidad de percepción tridimensional del usuario en la etapa de teleoperación. Asimismo, se contempla integrar modelos de aprendizaje automático capaces de predecir estrategias de agarre en función de las características geométricas del objeto con el fin de aumentar la tasa de acierto global del sistema (Ten Pas et al., 2017; Vuong et al., 2024; Noh et al., 2024).


## Agradecimientos

La investigación que se presenta en este trabajo ha recibido financiación del proyecto ROSOGAR PID2021-123020 OB-I00 financiado por MCIN/AEI/10.13039/501100011033/FEDER, UE, y del proyecto EIAROB Financiado por la Consejería de Familia de la Junta de Castilla y León - Next Generation EU.


## Referencias


Cheng, X., Li, J., Yang, S., Yang, G., Wang, X., 2024. Open-television: Teleoperation with immersive active visual feedback. arXiv preprint arXiv:2407.01512.

Ding, R., Qin, Y., Zhu, J., Jia, C., Yang, S., Yang, R., Qi, X., Wang, X., 2024. Bunny-visionpro: Real-time bimanual dexterous teleoperation for imitation learning. arXiv preprint arXiv:2407.03162.

Duque-Domingo, J., García-Gómez, M., Zalama, E., Gómez-García-Bermejo, J., 2024a. Continuous rapid learning by human imitation using audio prompts and one-shot learning. In: 2024 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). pp. 2572–2577. DOI: 10.1109/IROS58592.2024.10801547

Duque-Domingo, J., García-Gómez, M., Zalama, E., Gómez-García-Bermejo, J., 2024b. Learning by demonstration of a robot using one-shot learning and cross-validation regression with z-score. Electronics 13 (17). DOI: 10.3390/electronics13173365

García-Gómez, M., Duque-Domingo, J., Gómez-García-Bermejo, J., Zalama, E., 2024. Optimización de la teleoperación del robot kinova gen3 mediante realidad mixta. In: Actas del Simposio de Robótica, Bioingeniería y Visión por Computador: Badajoz, 29 a 31 de mayo de 2024. Servicio de Publicaciones de la Universidad de Extremadura, pp. 67–72.

Gualtieri, M., Ten Pas, A., Saenko, K., Platt, R., 2017. High precision grasp pose detection in dense clutter. In: Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). pp. 598–605.

Iyer, A., Peng, Z., Dai, Y., Guzey, I., Haldar, S., Chintala, S., Pinto, L., 2024. Open teach: A versatile teleoperation system for robotic manipulation. arXiv preprint arXiv:2403.07870.

Kamijo, T., Beltran-Hernandez, C. C., Hamaya, M., 2024. Learning variable compliance control from a few demonstrations for bimanual robot with haptic feedback teleoperation system. In: 2024 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). IEEE, pp. 12663–12670.

Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., Xiao, T., Whitehead, S., Berg, A. C., Lo, W.-Y., et al., 2023. Segment anything. In: Proceedings of the IEEE/CVF international conference on computer vision. pp. 4015–4026.

Liu, Y., Jiang, D., Tao, B., Qi, J., Jiang, G., Yun, J., Huang, L., Tong, X., Chen, B., Li, G., 2022. Grasping posture of humanoid manipulator based on target shape analysis and force closure. Alexandria Engineering Journal 61 (5), 3959–3969.

Noh, S., Kim, J., Nam, D., Back, S., Kang, R., Lee, K., 2024. Graspsam: When segment anything model meets grasp detection. arXiv preprint arXiv:2409.12521.

Omarali, B., Denoun, B., Althoefer, K., Jamone, L., Valle, M., Farkhatdinov, I., 2020. Virtual reality based telerobotics framework with depth cameras. In: 2020 29th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN). pp. 1217–1222.

Su, Y.-P., Chen, X.-Q., Zhou, T., Pretty, C., Chase, G., 2022. Mixed-reality-enhanced human–robot interaction with an imitation-based mapping approach for intuitive teleoperation of a robotic arm-hand system. Applied Sciences 12 (9), 4740.

Suzuki, T., Oka, T., 2016. Grasping of unknown objects on a planar surface using a single depth image. Journal of Robotics and Mechatronics 28 (4), 483–490.

Ten Pas, A., Gualtieri, M., Saenko, K., Platt, R., 2017. Grasp pose detection in point clouds. The International Journal of Robotics Research 36 (13-14), 1455–1473.

Vuong, A. D., Vu, M. N., Le, H., Huang, B., Binh, H. T. T., Vo, T., Kugi, A., Nguyen, A., 2024. Grasp-anything: Large-scale grasp dataset from foundation models. In: 2024 IEEE International Conference on Robotics and Automation (ICRA). IEEE, pp. 14030–14037.

Xu, S., Moore, S., Cosgun, A., 2022. Shared-control robotic manipulation in virtual reality. arXiv preprint arXiv:2205.10564.

Zhao, X., Ding, W., An, Y., Du, Y., Yu, T., Li, M., Tang, M., Wang, J., 2023. Fast segment anything. arXiv preprint arXiv:2306.12156.