

## Agarre de objetos desconocidos con un robot manipulador móvil bibraso

Espinoza, G. , Lishan, L. , Mendez, A. , Prados, A. , Barber, R. , Moreno, L. 

RoboticsLab, Departamento de Ingeniería de Sistemas y Automática, Universidad Carlos III de Madrid, Av. de la Universidad, 30, 28911 Leganés, España.

### Resumen

El agarre de objetos desconocidos en entornos domésticos supone uno de los principales retos de los robots asistenciales, donde los manipuladores deben adaptarse a la gran variedad de formas y tamaños de los objetos, y a la incertidumbre de los entornos desestructurados. En este artículo se presenta una implementación de un método *end-to-end* de agarre aplicado al robot asistencial bibraso ADAM en este tipo de condiciones. La implementación cubre el proceso de agarre completo, desde la detección y reconstrucción 3D de objetos a partir de formas supercuádricas, pasando por la síntesis de agarre mediante métodos analíticos no basados en aprendizaje, hasta la selección y planificación del brazo y la ejecución de su movimiento. El método ha sido evaluado en un simulador basado en PyBullet a partir de la reconstrucción de una escena real, sobre objetos de uso cotidiano, y de formas y tamaños variados, demostrando la viabilidad del flujo completo, también transferido a la plataforma robótica real.

**Palabras clave:** Supercuádricas, Síntesis de agarre, Manipulador bibraso, Objetos desconocidos, Agarre y manipulación robótica, Planificación de tareas y movimiento, Percepción y sensores en robótica, Robótica asistencial

### Grasping of unknown objects with a dual-arm mobile manipulator

#### Abstract

Grasping objects in indoor environments poses a major challenge for assistive robots, as mobile manipulators must adapt to a great variety of object shapes and sizes, and to the uncertainty of unstructured environments where the location and orientation of their elements are unknown. In this paper, we present an implementation of an *end-to-end* grasping method applied to the dual-arm assistant robot ADAM under such conditions. Our modular implementation covers the complete grasping process in three stages ranging from the detection and 3D reconstruction of objects using superquadric shapes, to learning-free grasp planning and filtering via analytical methods, and finally, arm selection, motion planning and execution. The method has been evaluated in a PyBullet-based simulator from the reconstruction of the real scene with everyday household objects of varying shapes and sizes. The results demonstrate the feasibility of the full pipeline, which was also successfully transferred to the real robot.

**Keywords:** Superquadrics, Grasp synthesis, Dual-arm manipulator, Unknown objects, Robotic grasping and manipulation, Task and motion planning, Robot perception and sensing, Assistive robotics

### 1. Introducción

Uno de los mayores retos de la robótica asistencial en entornos domésticos es el agarre de objetos. Estos entornos dinámicos y desestructurados contienen una gran variedad de objetos adecuados a las personas, a los que los robots deben adaptarse, y siendo desconocidos no solo en forma, tamaño o posición, sino también en la falta de experiencias previas de agarre que el robot pueda aprender y transferir. Por ello, las estrategias actuales basadas en aprendizaje explotan la información parcial

del objeto para estimar poses de agarre seguras y estables, lo que supone entrenamientos y demostraciones exhaustivos. Esto dificulta el desarrollo de una implementación *end-to-end*, desde la percepción hasta la ejecución, que pueda ser embebida en un robot asistencial que opere en un entorno real.

Por tanto, la finalidad de este trabajo es dotar al robot bibraso de asistencia de personas mayores en entornos domésticos ADAM (*Autonomous Domestic Ambidextrous Manipulator*) (Mora et al., 2024) de la capacidad de agarrar objetos desconocidos, estimando el agarre más estable sin conocimiento

previo del mismo. El método desarrollado trata el proceso de agarre de forma modular en tres fases: la detección y la reconstrucción 3D del objeto en la escena mediante formas supercuádricas; la síntesis del agarre por medio de dos métodos analíticos y su evaluación; y la selección y planificación del brazo óptimo que debe ejecutarlo. Aunque las fases de reconstrucción y planificación se apoyan en métodos consolidados en el estado del arte, la principal aportación se concentra en la fase de síntesis de agarre y toma de decisiones para un robot bíbrazo. En concreto, el trabajo contribuye en los siguientes aspectos:

- **Selección del brazo dominante para el agarre.** Se propone y desarrolla un criterio de evaluación para determinar el brazo óptimo para el agarre del objeto en función de la dirección de acercamiento, penalizando aproximaciones por caras ocluidas y teniendo en cuenta la estabilidad del agarre y la posición de reposo de ambos brazos.
- **Integración de un proceso de agarre.** Si bien existen múltiples flujos *end-to-end* de agarre, este trabajo aporta la integración completa de todas las etapas del proceso de agarre autónomo de objetos, orientadas a su despliegue sobre un robot móvil asistencial bíbrazo.
- **Modularidad.** La implementación está dividida en tres etapas modulares, reemplazables y adaptables a distintos robots y tipos de aplicaciones, que pueden ser sustituidas para mejorar la eficiencia global del método.

Para validar la propuesta, se ha llevado a cabo una evaluación exhaustiva de los módulos individuales que componen el flujo de agarre sobre la escena reconstruida en simulación y con objetos de uso cotidiano dispuestos aleatoriamente sobre el espacio de trabajo. El método también ha sido probado sobre el robot y objetos reales para comprobar su efectividad.

## 2. Trabajos Relacionados

### 2.1. Agarre de objetos desconocidos

El agarre de objetos desconocidos asume la ausencia de modelos geométricos o de experiencias de agarre previas. Este problema se ha abordado desde dos estrategias principales: la síntesis directa a partir de la información sensorial (2D/3D) parcial del objeto y la generación de poses tras la aproximación de la forma desconocida (Bohg et al., 2013). En la primera, destacan principalmente métodos empíricos generativos (Mousavian et al., 2019; Mayer et al., 2022; Weng et al., 2024), eficaces generando múltiples candidatos de agarre desde la vista parcial. Frente a estos enfoques, que requieren entrenamientos exhaustivos y gran cantidad de datos, otros autores reconstruyen mediante aproximaciones la forma original del objeto, en ocasiones asumiendo cierta simetría. Por ejemplo, Wu et al. (2023) aproximan la forma local del objeto para generar poses de agarre mediante pequeños desplazamientos libres de colisión. Igualmente, Vezzani et al. (2017) ajustan una supercuádrica a la vista parcial del objeto y a la mano, generando agarres mediante el ajuste entre ambas. Makhali et al. (2018) aprovechan ejes de simetría para completar la nube de puntos parcial antes del ajuste de la supercuádrica. En este caso, se generan agarres antipodales en los ejes principales de la supercuádrica.

En nuestro trabajo, como estos últimos, aproximamos formas de objetos a supercuádricas, aunque incluyendo métodos analíticos para la síntesis de agarres, lo que nos permite obtener métricas sobre la calidad del agarre. Esto permite estimar la forma de más de un objeto u obstáculo en escena que pueda ser utilizado a futuro para planificar tareas con el robot ADAM.

### 2.2. Selección del brazo de agarre

La disposición de dos brazos en los robots asistenciales supone una ventaja en la manipulación de los objetos del entorno. Además de permitir coger objetos de gran tamaño o interactuar con ambas manos para realizar una acción sobre un objeto, también permite el agarre de objetos con el brazo más conveniente sin necesidad de desplazar todo el robot. Dentro de este último caso, existen múltiples enfoques para la selección del brazo dominante. Las estrategias más comunes se dividen en métodos heurísticos basados en la proximidad espacial al objeto, lo que puede requerir reagarrs (Tsuda et al., 2012; Harada et al., 2012), y métodos indirectos basados en funciones de coste. Estos últimos evalúan múltiples candidatos asignados a cada brazo y seleccionan el agarre y el brazo óptimo según la estabilidad del agarre y la viabilidad cinemática en la trayectoria libre de colisión (Haustein et al., 2019), o minimizando la distancia recorrida desde la posición de reposo por los brazos (de Quirós et al., 2024). Nuestro enfoque, al igual que los anteriores, escoge brazo de forma indirecta, aunque con un enfoque híbrido que combina la información del agarre y la posición del objeto, lo que permite anticipar y facilitar la planificación de la trayectoria del brazo sin sacrificar la calidad del agarre.

### 2.3. Planificación *end-to-end* del agarre

Los algoritmos *end-to-end* planifican íntegramente el agarre, desde la percepción hasta el control del brazo, para dotar de autonomía a los robots en entornos domésticos reales desconocidos. En la última década, el problema del *end-to-end grasping* ha sido abordado mayormente desde el aprendizaje profundo (Bohg et al., 2013), mediante arquitecturas que extraen características de la nube de puntos para la detección, muestreo y clasificación de agarres (Zhao et al., 2021; Alliegro et al., 2022), y principalmente técnicas de *Reinforcement Learning (RL)* para el aprendizaje a partir de información RGB-D (Singh et al., 2025). También existen métodos no basados en aprendizaje como en (Rosell et al., 2019), donde se utilizan sinergias de agarre para muestrear agarres con forma humana y posteriormente se calcula la trayectoria del brazo, o enfoques puramente analíticos como GeoGrasp (Zapata-Impata et al., 2019), que propone un cálculo rápido de los puntos de contacto basándose en reglas geométricas extraídas directamente de la nube 3D de puntos. El trabajo en (Paez-Ubieta et al., 2023) parte de modelos de detección de objetos para la segmentación de la nube de puntos, mostrando la eficacia de estos enfoques analíticos en exteriores. Nuestro planteamiento sigue la filosofía de estos últimos enfoques geométricos-analíticos para resolver un problema *end-to-end*. En comparación con los basados en aprendizaje, el flujo de agarre desarrollado no requiere de demostraciones, bases de datos o GPU y, de igual forma, parte de la captura de imágenes RGB-D hasta la planificación y ejecución con el brazo óptimo, lo que facilita que el proceso completo pueda ser computado sobre el propio robot.

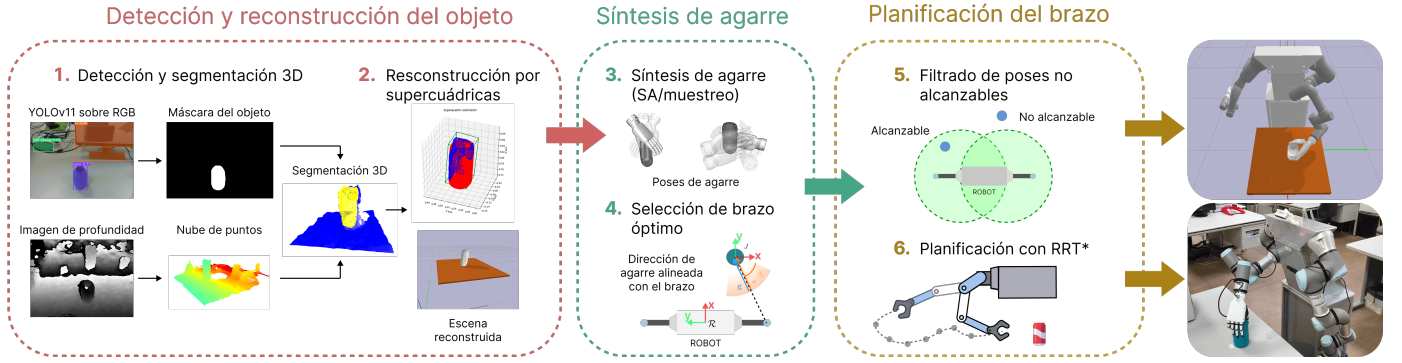


Figura 1: Implementación del método *end-to-end* de agarre. Primero, la información RGB-D es procesada para segmentar el objeto en 3D mediante el modelo *pin-hole* aplicando la máscara obtenida por YOLO/Moondream. La reconstrucción a partir de formas supercuadradas se utiliza para estimar la pose de agarre mediante métodos analíticos. Las poses se evalúan según la dirección de agarre y la pose óptima alcanzable es ejecutada sobre el brazo.

### 3. Conceptos previos y plataforma

En primer lugar, detallamos las características de la plataforma robótica utilizada que influyen en el problema, así como los fundamentos teóricos acerca de la calidad del agarre sobre los que se sustentan los métodos analíticos.

#### 3.1. Descripción y simulación del robot ADAM

Utilizamos el robot asistencial ADAM, un manipulador móvil en cuyo torso se encuentran dos brazos UR3 acabados en dos manos antropomórficas RH56DFX de Inspire Robots de 6 grados de libertad actuados, y, para la visión, una cámara RGB-D RealSense D435i. Como entorno de simulación y validación de la propuesta, se utiliza el simulador ADAMSim (Prados et al., 2025) basado en PyBullet, que integra un modelo virtual fiel a la plataforma real y con funcionalidades para el cálculo de cinemáticas, movimiento de los brazos y simulación física. Adicionalmente, se emplea GraspIt! (Miller and Allen, 2004) para el cómputo de la síntesis y de la calidad del agarre.

#### 3.2. Force Closure y calidad del agarre

Dada una pose del efector final  $g \in SE(3)$  y un objeto  $J$ , se busca cuantificar la estabilidad y calidad a partir de los contactos entre ambos con el fin de comparar poses de agarre generadas por métodos de síntesis y sobre brazos distintos. Una métrica habitualmente utilizada es  $\varepsilon - quality$ , o métrica Ferrari-Canny (Ferrari et al., 1992), computada sobre el espacio de esfuerzos de agarre (*Grasp Wrench Space*, *GWS*). El GWS se define como la envolvente convexa del conjunto de esfuerzos (*wrenches*)  $\omega_i = [f_i^T \ \tau_i^T]^T$ , donde  $f_i$  es la fuerza de contacto normalizada ejercida por el dedo  $i$ , y  $\tau_i$  es el correspondiente torque generado respecto al centro de masa del objeto. Un agarre se considera estable si se encuentra en condición de *force closure*, lo cual ocurre cuando el origen  $O$  del espacio de esfuerzos está estrictamente contenido en el GWS. Bajo esta condición, la métrica  $\varepsilon - quality$ , se define como el radio de la mayor esfera centrada en el origen, totalmente contenida en el GWS. Físicamente,  $\varepsilon$  representa la magnitud de la mayor perturbación externa que el agarre es capaz de resistir en la dirección más desfavorable. Por otro lado, la métrica *volume-quality*, calculada como el volumen del GWS, representa una estimación media de la capacidad del agarre para resistir esfuerzos.

### 4. Método

En esta sección se presenta el método desarrollado para el proceso de agarre autónomo de objetos desconocidos aplicado al robot asistencial ADAM. Este algoritmo queda dividido en módulos adaptados a la morfología del robot y a la configuración de sensores, brazos y efector final, centrándonos en manos antropomórficas. Se agrupan en tres etapas: la detección y reconstrucción del objeto a partir de la escena real, la síntesis de agarre y la planificación del movimiento del brazo. La Figura 1 muestra el flujo del proceso, indicando la información transmitida entre las distintas subetapas.

#### 4.1. Detección y reconstrucción del objeto

La primera etapa consiste en la reconstrucción de la escena real, el espacio de trabajo del robot, generando un modelo simulado del objeto desconocido sobre el obstáculo de la mesa. La Figura 2 muestra la secuencia de pasos en esta etapa.

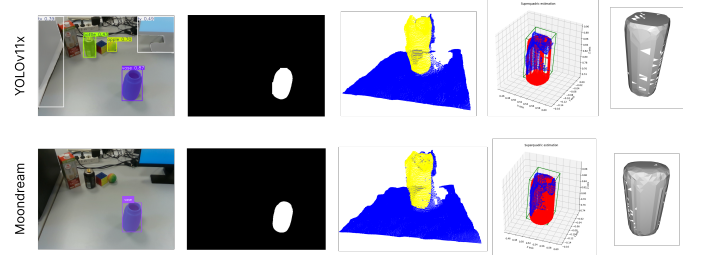


Figura 2: Secuencia de pasos en la reconstrucción del objeto. De izquierda a derecha: detección del objeto y extracción de máscara sobre imágenes 2D; aplicación de la máscara sobre la nube de puntos (modelo *pin-hole*); ajuste de forma supercuadrada; malla reconstruida.

##### 4.1.1. Detección y segmentación 3D

El proceso comienza capturando imágenes RGB y de profundidad de la escena. Se utilizan dos modelos de segmentación de objetos por instancias: YOLOv11 (Ultralytics, 2024) y Moondream AI (Moondream, 2024). El resultado de la segmentación es una máscara del objeto que se aplica a la imagen de profundidad alineada, asumiendo un modelo de cámara *pinhole* para extraer la instancia segmentada 3D de la nube de puntos de la escena. La nube de puntos está referenciada al eje de la cámara, por lo que se transforma al eje de la base del robot, que se utiliza como referencia para el cálculo de las poses de agarre.

#### 4.1.2. Reconstrucción mediante formas supercuádricas

La nube de puntos del objeto segmentada es una vista parcial del objeto. Sin embargo, la síntesis mediante métodos analíticos y la estimación de la calidad de agarre requieren de un modelo lo más fiel posible del objeto. Por ello, se hace uso de una reconstrucción del objeto mediante el método de aproximación con formas supercuádricas presentado en (Vezzani et al., 2017). Esta representación permite estimar la forma del objeto ajustando a formas con cierta simetría, incluyendo prismas, prismas de revolución y esferas, entre otras. Se define por la siguiente función de su superficie:

$$F = \left( \left( \frac{p_x}{\alpha_1} \right)^{2/\epsilon_2} + \left( \frac{p_y}{\alpha_2} \right)^{2/\epsilon_2} \right)^{\epsilon_2/\epsilon_1} + \left( \frac{p_z}{\alpha_3} \right)^{2/\epsilon_1} \quad (1)$$

donde un punto  $p_i = (p_x, p_y, p_z)$  de la nube de  $N_j$  puntos pertenece a la superficie de la supercuádrica si  $F = 1$ , se encuentra en su interior si  $F < 1$  o fuera de ella si  $F > 1$ . Los parámetros  $\epsilon_1$  y  $\epsilon_2$  influyen en la forma de la supercuádrica y en la curvatura de las aristas, y los parámetros  $\alpha_i$  modifican las dimensiones en los tres ejes. Por tanto, la reconstrucción del objeto se realiza mediante la optimización de los parámetros  $\theta = (\epsilon_1, \epsilon_2, \alpha_1, \alpha_2, \alpha_3, x, q)$  donde  $x$  y  $q$  son la posición y orientación en ángulos de Euler de la supercuádrica respecto a la base del robot:

$$\min_{\theta} \sum_{i=0}^{N_j} \left( \sqrt{\alpha_1 \alpha_2 \alpha_3} \times (F(p_i, \theta) - 1) \right)^2 \quad (2)$$

Esta reconstrucción permite obtener la supercuádrica que mejor se ajusta a la vista parcial del objeto, a la vez que permite obtener una estimación de su pose global. A partir de la supercuádrica ajustada se genera la malla del objeto y se reconstruye en la escena sobre el obstáculo de la mesa.

#### 4.2. Síntesis de agarre

El proceso continúa con la síntesis de agarre en GraspIt! (Miller and Allen, 2004) a partir de la escena reconstruida. Se generan poses de agarre para los brazos derecho (R) e izquierdo (L) mediante dos métodos: la planificación mediante *Simulated Annealing* (SA) (Kirkpatrick et al., 1983) y el muestreo de poses mediante heurísticas. En ambos casos, las poses de agarre  $g_{R/L,i}$  contienen únicamente la pose en  $SE(3)$  de la palma de la mano, y los dedos se cierran simultáneamente hasta detectar el contacto con el objeto, asumiendo agarres antipodales para reducir la dimensionalidad del espacio de búsqueda y calcular su calidad  $q(g_{R/L,i})$ .

##### 4.2.1. Síntesis de agarre mediante *Simulated Annealing*

El planificador de agarres en GraspIt! utiliza *Simulated Annealing*, un algoritmo probabilístico en el que en cada iteración la pose de agarre es trasladada en una dirección aleatoria. La nueva pose es aceptada si disminuye la energía  $E$  del sistema, es decir, si la calidad de agarre aumenta. En caso contrario, se acepta con una probabilidad  $P = e^{-\Delta E/T(t)}$  donde  $T(t)$  es la temperatura del sistema que disminuye con el tiempo, reduciendo progresivamente la probabilidad de aceptar agarres más inestables para evitar caer en mínimos locales. Mediante esta técnica se genera un conjunto de  $N$  agarres por brazo  $G_s = \{g_{R,0}, \dots, g_{R,N}, g_{L,0}, \dots, g_{L,N}\}$ .

##### 4.2.2. Síntesis de agarre mediante muestreo

Alternativamente, se estudia la generación de agarres mediante muestreo a partir de heurísticas. En concreto, para cada mano se generan una serie de candidatos horizontales  $G_h$ , donde el eje longitudinal de la mano es perpendicular al eje  $z$  de la supercuádrica (Figura 3a), y verticales  $G_v$ , con el eje longitudinal paralelo al mismo eje (Figura 3b). En primer lugar, para ambas direcciones de agarre se aprovechan las dimensiones máximas del objeto, dadas por los parámetros  $\alpha_i$  estimados de la supercuádrica, para generar las poses iniciales  $g_{v,R/L,0}$  y  $g_{h,R/L,0}$ , y se desplaza hacia el objeto en la dirección de agarre al punto más cercano sin colisión. A continuación, los agarres horizontales y verticales iniciales se rotan sobre el eje  $z$  de la supercuádrica en ángulos  $\Delta\theta_z = \pi/6$  y se añaden las poses sin colisión a  $G_v$  y a  $G_h$ . Además, se generan más muestras horizontales, desplazando las poses en distancias de  $\Delta d_z = \pm 0,01$ m para cada ángulo rotado. Todas las muestras de cada mano se guardan en un conjunto  $G_m = \{G_h \cup G_v\}$ .

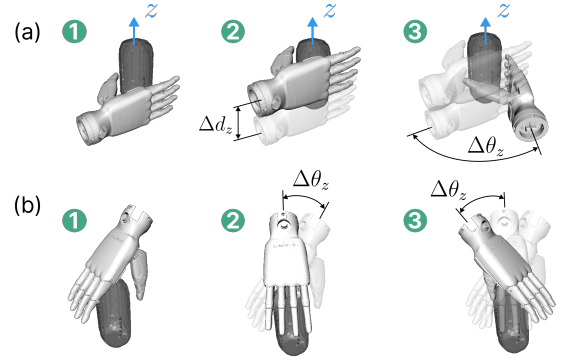


Figura 3: Síntesis de agarre mediante muestreo de poses. (a) Agarres horizontales desplazados  $\Delta d_z$  sobre el eje  $z$  de la supercuádrica y  $\Delta\theta_z$  alrededor. (b) Agarres verticales desplazados por  $\Delta\theta_z$ .

##### 4.2.3. Selección del brazo dominante para el agarre

Tras la síntesis se obtiene un conjunto de muestras, bien  $G = G_s$ , o bien  $G = G_m$ , que contienen candidatos de agarre de cada brazo  $g_{R/L,i}$  con su respectiva calidad de agarre  $q(g_{R/L,i})$ . Esta calidad es independiente de la posición del brazo del robot o de la orientación del propio torso, por lo que se define la dirección de agarre como el eje longitudinal de la mano robótica. Se penalizan aquellas poses cuya dirección de agarre provenga por la cara no vista del objeto, al ser usualmente poses de agarre más difíciles de alcanzar por el brazo y menos naturales.

En primer lugar, se toma la proyección  $\mathbf{g}_i$  de la dirección de agarre sobre el plano  $X - Y$  (ver Figura 4). Por otra parte, se define una dirección óptima de agarre  $\mathbf{o}_{R/L} = \mathbf{c}_o - \mathbf{r}_{R/L}$ , donde  $\mathbf{c}_o = (c_{o,x}, c_{o,y})$  es el centro proyectado del objeto, y  $\mathbf{r}_{R/L} = (r_{R/L,x}, r_{R/L,y})$  es la posición de reposo del brazo derecho o izquierdo. Se define el ángulo de alineamiento como  $\beta_i = \angle(\mathbf{g}_i, \mathbf{o})$ . Por tanto, se aplica un peso  $\lambda(\beta_i)$  a la calidad del agarre, penalizando aquellos agarres menos alineados con la dirección óptima de agarre según la ecuación:

$$\lambda(\beta_i) = 1 - \frac{1}{1 + e^{-\sigma(\beta_i - \beta_c)}} \quad (3)$$

donde  $\beta_c$  representa la amplitud del cono alrededor de  $\mathbf{o}$  con agarres mejor orientados, suavizado por una función sigmoide

con factor de suavizado  $\sigma$ . Finalmente, la calidad total de una pose de agarre viene dada por:  $Q(g_{R/L,i}) = \lambda(\beta_i) \times q(g_{R/L,i})$ .

Los candidatos de ambos brazos se unifican y ordenan descendientemente según su calidad total, seleccionando para la ejecución del agarre el brazo con mayor puntuación. Cabe destacar que sería posible acoplar la síntesis de agarre a las restricciones cinemáticas y por obstáculos del entorno, evitando el cálculo de la calidad de agarre en poses no alcanzables. Sin embargo, esto implicaría el cálculo de la cinemática inversa del brazo para cada candidato de agarre y sus colisiones, lo que puede ser ineficiente. El desacople permite así el reducir el cálculo de cinemáticas y colisiones al comprobar partiendo del candidato de mejor calidad hasta encontrar una solución válida.

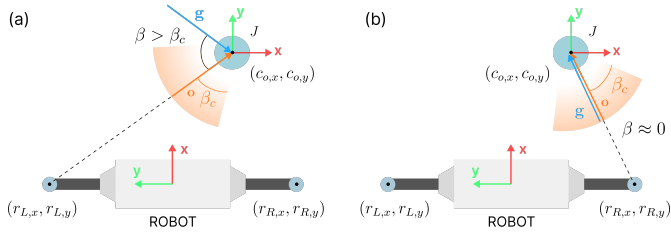


Figura 4: (a) Se penaliza la dirección de agarre no alineada con la dirección óptima para el brazo izquierdo. (b) La dirección de agarre queda alineada con la orientación óptima para el brazo derecho y su peso es mayor.

#### 4.3. Planificación del brazo

La etapa final del proceso se centra en evaluar la viabilidad cinemática y ejecutar el movimiento del brazo. Primero, se selecciona el primer candidato de agarre  $g_{R/L,0}$  de la lista ordenada y se calcula una pose de pre-agarre aplicando un desplazamiento de 10 cm a lo largo del eje longitudinal de la mano, en sentido opuesto al objeto. A continuación, se evalúa si existe solución de cinemática inversa para las poses de pre-agarre y agarre. En caso de no ser viable, el candidato se descarta y se itera secuencialmente sobre la lista hasta encontrar un agarre ejecutable.

Una vez seleccionado el candidato, se planifica la trayectoria en el espacio de configuración del brazo seleccionado para el agarre mediante el algoritmo *Rapidly-Exploring Random Tree Star (RRT\*)*. Este algoritmo se computa en el simulador ADAMSim (Prados et al., 2025), lo que permite detectar colisiones con obstáculos como la mesa. El resultado es una trayectoria discretizada en una serie de configuraciones  $q_j$  desde la posición inicial de reposo  $q_0$  hasta la posición de agarre  $q_g$ , pasando por el pre-agarre. Esta trayectoria es enviada al control del brazo en simulación para la ejecución del agarre.

## 5. Resultados

El proceso de agarre se evalúa en un espacio de trabajo donde el robot ADAM se posiciona frente a una mesa sobre la que en cada prueba se coloca un objeto en una posición y orientación desconocidas, sin obstáculos adicionales que obstruyan la manipulación. Las siguientes secciones presentan los resultados obtenidos en cada fase del proceso y los resultados globales, probados sobre 6 objetos (jarrón, naranja, pelota, lata, cubo de Rubik y brik de leche) en el simulador ADAMSim. El código es completamente ejecutado sobre una CPU Intel Core i7-12700H.

### 5.1. Detección y reconstrucción del objeto

Se han utilizado dos modelos para la identificación y segmentación 2D de los objetos en la imagen de la escena: YOLOv11-x y Moondream. El primero es un modelo mucho más ligero, obteniendo unos tiempos de inferencia de  $1,81 \pm 0,17s$  frente a  $103,37 \pm 2,69s$  en el segundo. No obstante, Moondream es una VLM (*Vision-Language Model*) por lo que no está acotado a un determinado número de clases aprendidas previamente, permitiendo segmentar una mayor variedad de objetos.

A partir de las máscaras extraídas, se aplica el modelo *pin-hole* sobre las nubes de puntos y se ajustan las supercuádras. La supercuádras reconstruida tiene una distancia de Chamfer de entre 0,003 y 0,005m y una intersección sobre la unión (*IoU*) media del 64,7% con respecto a la nube del objeto segmentada manualmente. Por tanto, se obtienen reconstrucciones fieles a la parte vista del objeto, y los parámetros optimizados de la supercuádras permiten extraer su tamaño, posición y orientación con respecto al centro geométrico real de los objetos.

### 5.2. Síntesis de agarre y planificación del brazo

Sobre los modelos reconstruidos, se realizan las síntesis de agarre mediante SA y muestreo, y se aplica la pose de agarre con mayor  $Q(g_{R/L,i})$  que sea alcanzable. La síntesis SA logra planificar al menos un agarre en *force closure* en el 100% de las pruebas con todos los objetos. Por su parte, la generación por muestreo logra la planificación en el 97,2% de las pruebas, sin generar una pose estable en algunos casos con los objetos más pequeños. No obstante, la síntesis de agarre por muestreo con ambas manos se completa en un tiempo mucho menor de 18,7s frente a 88,2s de media.

### 5.3. Proceso de agarre

Finalmente, se evalúa el proceso completo de agarre sobre los 6 objetos de uso cotidiano con 9 pruebas de agarre por objeto y por método de síntesis de agarre. En cada prueba, el objeto se coloca en la mesa frente al robot en una posición y orientación desconocidas, distribuidas dentro de todo el rango de visión de la cámara del robot. Dado que en esta sección la finalidad es evaluar el rendimiento del proceso y los modelos de segmentación por instancias dan resultados próximos, se toma Moondream para detectar una mayor cantidad de objetos.

La Tabla 1 recoge el porcentaje de agarres exitosos, contando como tales aquellos que logran agarrar el objeto y levantarlo 10cm sobre el nivel de la mesa, partiendo desde la detección del mismo. Además, los resultados son desglosados por método de síntesis de agarre y recoge el porcentaje de casos en los que no se ejecuta correctamente el proceso completo debido a que no se logra agarrar el objeto o se desliza tras elevarlo, y aquellos en los que no se encuentra una solución al movimiento del brazo, causado principalmente por una colisión con poses de agarre cercanas a la mesa. El análisis experimental muestra una superioridad de la generación de poses de agarre por medio de muestreos, logrando una tasa de planificación del movimiento del brazo superior (61,11% frente 46,30%). Aunque los módulos individuales demuestran robustez de forma aislada, el acoplamiento en simulación se ve en parte afectado por las físicas de contacto. No obstante, la viabilidad del flujo se demuestra de forma cualitativa en la Figura 5, en la que se muestran casos de agarre con distintos objetos y métodos de síntesis.

Tabla 1: Tasa de éxitos de agarre y planificación del brazo

Objeto	Agarre exitoso		RRT* exitoso	
	Sim. Ann.	Muestreo	Sim. Ann.	Muestreo
Jarrón	55.56 %	44.44 %	66.67 %	44.44 %
Naranja	11.11 %	33.33 %	77.78 %	55.56 %
Pelota	0.00 %	44.44 %	22.22 %	44.44 %
Lata	33.33 %	77.78 %	55.56 %	88.89 %
Rubik	44.44 %	88.89 %	11.11 %	77.78 %
Brik de leche	22.22 %	44.44 %	44.44 %	55.56 %
<b>TOTAL</b>	<b>27.78 %</b>	<b>55.55 %</b>	<b>46.30 %</b>	<b>61.11 %</b>

Finalmente, para verificar que el método es transferible al robot real, se realiza una prueba conceptual con un jarrón cilíndrico en la que la trayectoria calculada a partir del simulador es enviada a través de ROS (*Robot Operating System*) al controlador del brazo para agarrar y elevar el objeto. Esta prueba satisfactoria sugiere la necesidad de una validación experimental exhaustiva para desacoplar el rendimiento del algoritmo de las limitaciones inherentes al entorno de simulación.

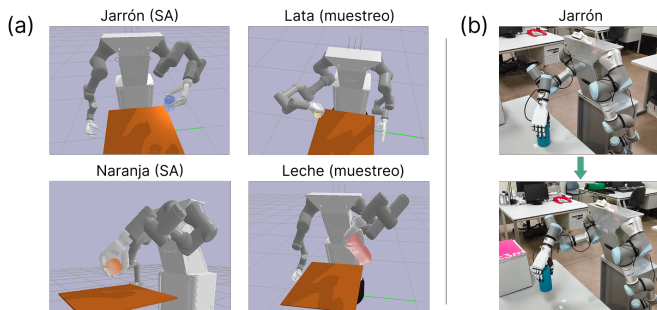


Figura 5: (a) Pruebas en simulación. Ejecución del agarre en simulación con Simulated Annealing (SA) y por muestreo sobre objetos cotidianos. (b) Prueba en entorno real. Transferencia de trayectoria calculada al robot ADAM.

## 6. Conclusiones

Este trabajo presenta un método *end-to-end* para el agarre autónomo de objetos desconocidos en entornos domésticos, aplicado al robot asistencial ADAM. El sistema integra las fases de detección y reconstrucción 3D del objeto mediante formas supercuádricas, la síntesis y planificación del agarre. Además, se define una métrica para seleccionar el brazo óptimo, combinando información sobre la estabilidad del agarre, la posición relativa del objeto respecto al robot y su alcanzabilidad.

Pese a las limitaciones del modelo de contacto simulado, se demuestra la eficacia del flujo completo y su modularidad, comprobada con distintos métodos de detección y agarre. El método propuesto es eficaz, aunque asume objetos aislados, estando también limitado cuando son reflectantes, transparentes o de formas irregulares, lo que perjudica su reconstrucción con supercuádricas. Así, los trabajos futuros buscarán evaluar escenarios con múltiples objetos y más complejos, que puedan tener mayores restricciones espaciales para la manipulación.

## Agradecimientos

Este trabajo ha sido apoyado por el proyecto Advanced Mobile dual-arm Manipulator for Elderly People Attendance (AMME) (PID2022-139227OB-I00), financiado por el Ministerio de Ciencia e Innovación.

## Referencias

- Alliegro, A., Rudorfer, M., Frattin, F., Leonardis, A., Tommasi, T., 2022. End-to-end learning to grasp via sampling from object point clouds. *IEEE Robotics and Automation Letters* 7 (4), 9865–9872.
- Bohg, J., Morales, A., Asfour, T., Kragic, D., 2013. Data-driven grasp synthesis—a survey. *IEEE Transactions on robotics* 30 (2), 289–309.
- de Quirós, C. B. B., de la Casa, S. M., Menéndez, E., 2024. Selección y agarre robótico de objetos basada en el seguimiento de la mirada. In: *Actas del Simposio de Robótica, Bioingeniería y Visión por Computador: Badajoz, 29 a 31 de mayo de 2024*. Servicio de Publicaciones, pp. 127–132.
- Ferrari, C., Canny, J., et al., 1992. Planning optimal grasps. In: *Proceedings., 1992 IEEE International Conference on Robotics and Automation, 1992*. Vol. 3. IEEE, pp. 2290–2295.
- Harada, K., Foissotte, T., Tsuji, T., Nagata, K., Yamanobe, N., Nakamura, A., Kawai, Y., 2012. Pick and place planning for dual-arm manipulators. In: *2012 IEEE International Conference on Robotics and Automation*. IEEE, pp. 2281–2286.
- Haustein, J. A., Hang, K., Stork, J., Kragic, D., 2019. Object placement planning and optimization for robot manipulators. In: *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, pp. 17–24.
- Kirkpatrick, S., Gelatt Jr, C. D., Vecchi, M. P., 1983. Optimization by simulated annealing. *science* 220 (4598), 671–680.
- Makhal, A., Thomas, F., Gracia, A. P., 2018. Grasping unknown objects in clutter by superquadric representation. In: *2018 Second IEEE international conference on robotic computing (IRC)*. IEEE, pp. 292–299.
- Mayer, V., Feng, Q., Deng, J., Shi, Y., Chen, Z., Knoll, A., 2022. Ffhnet: Generating multi-fingered robotic grasps for unknown objects in real-time. In: *2022 International Conference on Robotics and Automation (ICRA)*. IEEE, pp. 762–769.
- Miller, A., Allen, P., 2004. Graspit! a versatile simulator for robotic grasping. *IEEE Robotics & Automation Magazine* 11 (4), 110–122.
- Moondream, 2024. Moondream. The Open Source VLM That Runs Everywhere. URL: <https://moondream.ai/>
- Mora, A., Prados, A., Mendez, A., Espinoza, G., Gonzalez, P., Lopez, B., Muñoz, V., Moreno, L., Garrido, S., Barber, R., 2024. Adam: a robotic companion for enhanced quality of life in aging populations. *Frontiers in Neurobotics* 18, 1337608.
- Mousavian, A., Eppner, C., Fox, D., 2019. 6-dof graspnet: Variational grasp generation for object manipulation. In: *Proceedings of the IEEE/CVF international conference on computer vision*. pp. 2901–2910.
- Paez-Ubieta, I. d. L., Castaño-Amorós, J., Puente, S. T., Gil, P., 2023. Vision and tactile robotic system to grasp litter in outdoor environments. *Journal of Intelligent & Robotic Systems* 109 (2), 36.
- Prados, A., Espinoza, G., Mendez, A., Mora, A., Garrido, S., Barber, R., 2025. Adamsim: Pybullet-based simulation environment for research on domestic mobile manipulator robots. *Jornadas de Automática* (6).
- Rosell, J., Suárez, R., García, N., Din, M. U., 2019. Planning grasping motions for humanoid robots. *International Journal of Humanoid Robotics* 16 (06), 1950041.
- Singh, R., Van Wyk, K., Abbeel, P., Malik, J., Ratliff, N., Handa, A., 2025. End-to-end rl improves dexterous grasping policies. *arXiv preprint arXiv:2509.16434*.
- Tsuda, A., Kakiuchi, Y., Nozawa, S., Ueda, R., Okada, K., Inaba, M., 2012. On-line next best grasp selection for in-hand object 3d modeling with dual-arm coordination. In: *2012 IEEE International Conference on Robotics and Automation*. IEEE, pp. 1799–1804.
- Ultralytics, 2024. Ultralytics YOLO v11. URL: <https://github.com/ultralytics/ultralytics/blob/main/docs/en/models/yolo11.md>
- Vezzani, G., Pattacini, U., Natale, L., 2017. A grasping approach based on superquadric models. In: *2017 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, pp. 1579–1586.
- Weng, Z., Lu, H., Kragic, D., Lundell, J., 2024. Dexdiffuser: Generating dexterous grasps with diffusion models. *IEEE Robotics and Automation Letters*.
- Wu, Y., Liu, W., Liu, Z., Chirikjian, G. S., 2023. Learning-free grasping of unknown objects using hidden superquadrics. *arXiv preprint arXiv:2305.06591*.
- Zapata-Impata, B. S., Gil, P., Pomares, J., Torres, F., 2019. Fast geometry-based computation of grasping points on three-dimensional point clouds. *International Journal of Advanced Robotic Systems* 16 (1), 1729881419831846.
- Zhao, B., Zhang, H., Lan, X., Wang, H., Tian, Z., Zheng, N., 2021. Regnet: Region-based grasp network for end-to-end grasp detection in point clouds. In: *2021 IEEE international conference on robotics and automation (ICRA)*. IEEE, pp. 13474–13480.