CEA Comité Español de Automática
XLV JORNADAS AUTOMÁTICA MÁLAGA 2024
UNIVERSIDADE DA CORUÑA

Jornadas de Automática

Control robótico inteligente para extracción de elementos flexibles

Tapia Sal Paz, B.[a,*], Sorrosal, G.[a], Mancisidor, A.[b], Cabanes, I.[b]

[a]*Ikerlan, Centro tecnológico de investigación, Basque Research and Technology Alliance (BRTA), 20500 Arrasate, España.*
[b]*Departamento de Ingeniería de Sistemas y Automática, Escuela de Ingeniería de Bilbao, Universidad del País Vasco (UPV-EHU), 48013, Bilbao, España.*

To cite this article: Tapia Sal Paz, Benjamín., Sorrosal, Gorka., Mancisidor, Aitziber., Cabanes Itziar. 2024.Intelligent robotic control for flexible element extraction.
Jornadas de Automática, 45. https://doi.org/10.17979/ja-cea.2024.45.10927

Resumen

La automatización de tareas de desensamblaje presenta grandes desafíos, principalmente relacionados con las caracteristicas dinámicas y no estructuradas de la tarea, en donde se necesitan acciones adaptativas para asegurar la interacción adecuada entre el robot y el entorno de la tarea. En este trabajo se propone un control basado en aprendizaje por refuerzo para la automatización de tareas de extracción de elementos flexibles mediante el uso de robots, buscando así enfrentar las dificultades de trabajar en estos entornos desestructurados y dinámicos. Para lograr eso, el control propuesto aprenderá a tomar acciones adecuadas en el movimiento del robot que llevarán a la extraction de elementos flexible a través de trayectorias de baja fuerza. Como resultado, este trabajo demuestra cómo la integración de un controlador basado en aprendizaje por refuerzo puede abordar los desafíos de la extracción de elementos flexibles, contribuyendo asi al avance de procesos de desensamblaje inteligentes mediante el uso de robots.

Palabras clave: Aprendizaje por refuerzo, Sistemas robóticos autónomos, Robótica inteligente, Manipuladores robóticos.

Intelligent robotic control for flexible element extraction

Abstract

The automation of disassembly tasks presents significant challenges, mainly related to the dynamic and unstructured characteristics of the task, where adaptive actions are needed to ensure proper interaction between the robot and the task environment. This work proposes a reinforcement learning-based control to automate flexible element extraction tasks using robots, aiming to tackle the difficulties of working in these unstructured and dynamic environments. To achieve this, the proposed control will learn to take appropriate actions in the robot's movement that will extract flexible elements through low-force trajectories. As a result, this work demonstrates how integrating a reinforcement learning-based controller can address the challenges of flexible element extraction, thereby contributing to the advancement of intelligent disassembly processes using robots.

Keywords: Reinforcement learning control, Autonomous robotic systems, Intelligent robotics, Robots manipulators.

1. Introducción

El desensamblaje de un producto consiste en la descomposición del mismo en sus componentes o partes esenciales, con el objetivo de reparar, reciclar o reutilizar. Cumpliendo así un rol fundamental en la gestión del ciclo de vida de diversos productos que van desde dispositivos electrónicos, electrodomésticos, hasta maquinaria industrial.

El desensamblaje es una tarea que posee cierto grado complejidad debido a la variabilidad de productos y situaciones que se pueden encontrar durante el proceso. Los seres humanos son capaces de adaptarse a esos cambios de manera natural gracias a la capacidad de percepción y de toma de decisiones, por ese motivo en la actualidad este tipo de tarea se realiza mayormente de manera manual.

Sin embargo existe la necesidad de implementación de

*Autor para correspondencia: btapia@ikerlan.es

sistemas automatizados para enfrentar las limitaciones de estos métodos, como son la escalabilidad del proceso Kurilova-Palisaitiene et al. (2018) y aspectos relacionados a la seguridad de los operarios Poschmann et al. (2020). El uso de robots es uno de los métodos mas adecuados para este tipo de tarea gracias a la flexibilidad y adaptabilidad que estos pueden aportar. Sin embargo es necesario enfrentar los principales desafíos de las técnicas de control clásicas en tareas desestructuradas y dinámicas Zachares et al. (2021).

Para realizar de forma satisfactoria la tarea de desensamblaje mediante el uso de robots, este debe realizar las diferentes tareas de extracción asegurando la integridad física de los diferentes componentes del entorno y el robot. El principal desafío se encuentra en el no conocimiento de las condiciones o características, las cuales varian de un caso a otro. Por este motivo son necesarias técnicas de control y monitorización, que permitan percebir estas y adaptarse deacuerdo a ellas.

Los elementos flexibles, como cables, gomas o materiales blandos, son componentes comunes en un proceso de desensamblaje, y su extracción es un proceso complejo y delicado. En esta tarea se presentan las principales problemáticas de trabajar en entornos desestructurados y dinámicos, requiriendo una comprensión avanzada de sus propiedades materiales y geometrías.

Este artículo propone una arquitectura de control basada en aprendizaje por refuerzo para realizar este tipo de tareas. Para ello se utiliza la información de posición y fuerza proveniente de los sensores del robot para asi dar información al controlador acerca de la situación actual del entorno y que este puede actuar acordemente. El artículo presenta la siguiente estructura: en la sección 2 se explica la arquitectura de control propuesta, donde el controlador basado en aprendizaje por refuerzo es detallado (2.1). En la sección 3 Se muestra la configuración experimental utilizada para los experimentos descriptos en la sección 3.2, cuyos resultados son presentados y discutidos en las secciónes 4 y 5. Finalmente se presentan las conclusiones de este trabajo en la sección 6.

2. Arquitectura de control para tareas de extracción de elementos flexibles

Para lograr un sistema robótico capaz de adaptarse a entornos de desensamblaje dinámicos y no estructurados, la necesidad de una arquitectura de control sofisticada es necesaria Beltran-Hernandez et al. (2020). Esto se debe a que los métodos de control clásicos no pueden lidiar con estos desafíos, particularmente aquellos que involucran interacciones físicas con el entorno Kristensen et al. (2019). Una de las principales razones es la dificultad en modelar los comportamientos dinámicos complejos y no lineales inherentes a las interacciones físicas de la tarea. Basadas en modelos lineales, las técnicas de control tradicionales pueden tener dificultades para representar con precisión las complejas relaciones entre fuerzas, pares y desplazamientos en escenarios donde la dinámica del sistema es inherentemente no lineal. La incapacidad de considerar estas incertidumbres puede llevar a dificultades para lograr un control estable y robusto durante las interacciones físicas, afectando el desempeño del sistema robótico.

La incorporación de técnicas de aprendizaje por refuerzo en sistemas robóticos enfrenta estos desafíos y ofrece un nuevo paradigma en cómo los robots pueden aprender y adaptarse a estas tareas Kroemer et al. (2021). A diferencia de las técnicas de control clásicas, donde para lograr adaptabilidad en entornos no estructurados es necesario modelar una amplia variedad de situaciones, en esta técnicas el robot aprende autónomamente la tarea a través de la interacción con su entorno Sutton and Barto (2018). Esta autonomía permite que los robots aprendan cómo adaptarse y a optimizar su comportamiento, mostrando buenos resultados en aplicaciones dinámicas y no estructuradas Zhou et al. (2021).

La arquitectura de control que se presenta en este trabajo esta compuesta de dos niveles de planificación (uno global y otro local) (Figura 1). De esta manera, se busca combinar comportamientos de control pre-planificados y reactivos, incrementando así la adaptabilidad, eficiencia y autonomía del sistema Paz et al. (2023). El framework hybrid planning de ROS2 (Figura 1) se utiliza para implementar esta arquitectura de control. Este framework se basa en el uso de planificadores globales y locales para cumplir con el objetivo global (pre-planificado) y reaccionar a la interacción dinámica (reactivo) respectivamente. En la solución propuesta, para cumplir con todo el proceso de desensamblaje, el planificador global se invoca al comienzo de la tarea para generar una trayectoria de referencia donde se especifican todos los puntos de agarre necesarios ($G_{posición}$). Luego, el planificador global ejecuta los movimientos para posicionar el robot en la posición de agarre y realizar la tarea de agarre. A partir de este momomento comienza la interacción, por lo que el planificador local se activa para así reaccionar a las fuerzas de interacción hasta que se logre la extracción. Una vez que este punto de agarre se extrae, se repite este procedimiento hasta que todos los puntos de agarre son extraidos y la tarea de desensamblaje se considera finalizada.

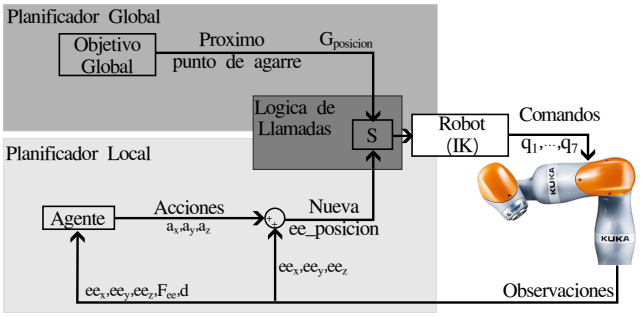

Figura 1: Propuesta de control híbrido basado en Aprendizaje por Refuerzo.

2.1. Control basado en aprendizaje por refuerzo

La elección del algoritmo de aprendizaje depende de las características específicas de la aplicación robótica, incluyendo la naturaleza del entorno, la complejidad de la tarea y los tipos de acciones que el robot necesita realizar. Para la extracción de elementos flexibles, donde se requieren acciones continuas y precisas, se decidió utilizar el algoritmo Proximal Policy Optimization (PPO) ya que muestra resultados adecuados para tareas con características similares en la literatura Lillicrap et al. (2015). Para ayudar al agente del aprendizaje por refuerzo a aprender la tarea de extracción de elementos flexibles, los espacios de estado y observación son conceptos cruciales que definen la acción esperada y la información dis-

ponible del sistema. La elección de estos espacios influye directamente en la capacidad del robot para aprender y realizar la tarea de manera eficiente.

En este trabajo, el espacio de acción propuesto A (Ec. (1)) está integrado para los desplazamientos cartesianos desde la posición actual del efector (a_x, a_y, a_z). Estas acciones son continuas y van de 0 a 0.05 metros,

$$A \rightarrow a_x, a_y, a_z \in \Re[0, 0{,}05] \tag{1}$$

Para el espacio de observación O (Ec. (2)), se utilizan cinco mediciones. Estas son las posiciones cartesianas del efector (ee_x, ee_y, ee_z), la fuerza resultante en el mismo (F_{ee}), y también la distancia (d) entre el punto de agarre $(G_{posición})$ y la posición actual del efector $(ee_{posición})$,

$$O \longrightarrow \begin{array}{l} ee_x, ee_y, ee_z \\ F_{ee} = \left\| F_{x_{ee}} + F_{y_{ee}} + F_{z_{ee}} \right\| \\ d = \left\| G_{posición} - ee_{posición} \right\| \end{array} \tag{2}$$

A través de estos dos espacios, el agente interactúa con el entorno y recibe retroalimentación de las acciones realizadas. Esta retroalimentación se calcula con la función de recompensa, que juega un papel central en guiar el proceso de aprendizaje del sistema, donde el objetivo está implícitamente definido. Para lograr esto, la función de recompensa asigna un valor numérico a cada par estado-acción, indicando el beneficio inmediato asociado con la acción realizada por el agente. El objetivo del agente es aprender una política que maximice la suma de recompensas a lo largo del episodio. La función de recompensa (R) propuesta (Ec. (3)) para la tarea de extracción de elementos flexibles se compone de dos componentes. La primera es la distancia entre el punto de agarre y la posición actual del efector (d) (definido en la Ec. (2)) que incentiva al sistema a extraer el elemento, es decir, que el efector separe el elemento del punto de agarre. El segundo componente es la fuerza resultante en el efector (F_{ee}), que guía al sistema a extraer el elemento a través de trayectorias de baja fuerza,

$$R = d - \beta * {F_{ee}}^2 \tag{3}$$

donde el parámetro β y la forma cuadrática para el componente de fuerza se introducen para lograr una función resultante con un gradiente y un valor máximo, guiando al agente a lograr el comportamiento deseado. Esto puede verse en la Figura 2 donde los valores más altos corresponden a trayectorias de baja fuerza.

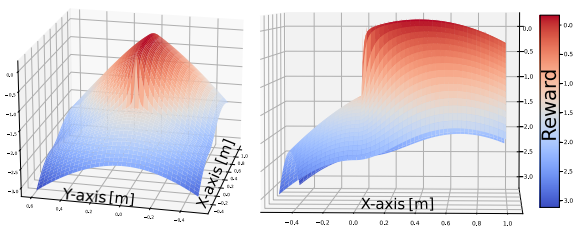

Figura 2: Distribución del valor de la función de recompensa considerando un plano que contiene el punto de agarre y la dirección preferencial de extracción.

3. Configuración Experimental

En la Figura 3(a) se puede observar una tarea típica de extracción de elementos flexibles, donde se necesita extraer la junta de la puerta de un refrigerador. Siguiendo lo presentado en 2, la tarea se divide en dos objetivos para realizar este proceso. Por un lado, tenemos un objetivo general de despegar todo el elemento de su carcasa, es decir, agarrar la goma en diferentes posiciones hasta que se extraiga completamente. Por otro lado, un objetivo más específico en cada agarre del elemento flexible, que es encontrar trayectorias de extracción de baja fuerza.

3.1. Entorno Simulado

Para evaluar el rendimiento de la metodología de control propuesta, se ha establecido un entorno simulado para la validación, donde se pueden destacar tres componentes principales: la cinemática y dinámica del robot, el espacio de trabajo del caso de uso y las fuerzas de interacción de la tarea.

Para todas las implementaciones de software, se utiliza el framework ROS2, que facilita el trabajo con simulaciones físicas (Gazebo) y las herramientas cinemáticas y dinámicas del robot KUKA iiwa14, que se utiliza para todas las implementaciones. Luego, para replicar el escenario del caso de uso, el arreglo experimental propuesto se muestra en la Figura 3(b), que replica la situación que se puede encontrar durante la ejecución de las tarea. En este, el elemento flexible se agarra en la pinza, donde se impone una dirección de extracción preferencial (mejor dirección de extracción).

Finalmente, se simulan las fuerzas que el elemento flexible ejerce sobre el sistema. Este componente de simulación representa las fuerzas en el efector final durante la extracción. Esta fuerza se puede dividir en dos componentes principales. Por un lado, está la fuerza de reacción de la pinza $(F_{Soporte})$ sobre el elemento flexible, y por otro lado, la fuerza elástica $(F_{elástica})$ del elemento flexible. Estos comportamientos se replican utilizando los modelos matemáticos de las Ecuaciones 4 y 5.

Para el componente elástico, se considera un modelo de resorte,

$$F_{elástica} = K_{elástica} * d \tag{4}$$

donde la fuerza elástica $(F_{elástica})$ es proporcional a la constante elástica del material $(K_{elásticaa})$ y la distancia (d) del efector final desde el punto de agarre $(d = 0 \rightarrow F_{elástica} = 0)$.

Luego, el modelo presentado en la Ecuación 5 se usa para incorporar las fuerzas producidas por la pinza. En este modelo, se tienen en cuenta dos situaciones. La primera es para la situación en la que el robot está tirando en la dirección preferencial ($\alpha < 90$ y $\alpha > -90$), donde la fuerza resultante es una modulación sinusoidal del modelo elástico. En el otro caso, cuando el robot está tirando en una zona más restringida (dirección opuesta a la dirección preferencial: $\alpha > 90$ y $\alpha < 270$), la fuerza resultante corresponde al modelo elástico más una constante de restricción (a),

$$\begin{cases} \text{si } \alpha > 90 \quad y \quad \alpha < 270: \\ \qquad F_{Soporte} = a + d * K_{elástica} \\ \text{de otro modo}: \\ \qquad F_{Soporte} = \|sin(\alpha) * d * K_{elástica}\| \end{cases} \tag{5}$$

(a) Tarea de extracción de un elemento flexible de la puerta de un refrigerador de forma manual.

(b) Propuesta de validación para la tarea de extracción de elementos flexibles .

Figura 3: Tarea de referencia y arreglo experimental

Como resultado, en la Figura 4 se observa la distribución de fuerzas, en donde existe una zona de baja fuerza (representada en azul) que corresponde con la dirección preferente. A través de esta zona se espera que el sistema aprenda a realizar la extracción.

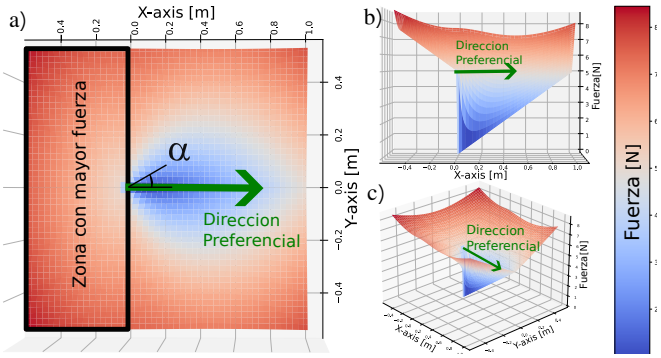

Figura 4: Distribución del valor de la fuerza. Vista superior (a); Vista lateral (b); y Vista isométrica (c) de la distribución espacial de la fuerza en un plano que contiene el punto de agarre y la dirección preferencial de extracción.

3.2. *Experimentos*

Para evaluar el comportamiento del sistema en diferentes entornos, se diseñan una serie de experimentos. Se comienza con la evaluación en entornos más estructurados y se finaliza la evaluacion en entornos más flexibles y desestructurados.

En el primer caso de estudio, el agente se entrena para una posición de agarre fija y una dirección de extracción preferencial fija (a lo largo del eje x). En el segundo caso, los entornos entre episodios varian la posición de agarre pero se mantiene la dirección de extracción preferencial. En el tercer caso, el entorno a lo largo de los episodios varía tanto en la posición de agarre como en la dirección de extracción preferencial.

Los valores de los principales hiperparámetros como son la tasa de aprendizaje (0.003), el tamaño del buffer (10^6), el batch size (64), el coeficiente de actualización ($\tau = 0{,}005$) y la tasa de descuento ($\gamma = 0{,}99$), se basan los valores usados en aplicaciones con características similares Lillicrap et al. (2015).

Para validar la estrategia aprendida en una operación de extracción de elementos flexibles, se evalúa una comparación la fuerza resultante en el efector durante la ejecución de la tarea. Esto compara las estrategias aprendidas por el agente con otras dos; la primera corresponde a una situación en la que se tiene un conocimiento completo del entorno, por lo que la trayectoria tomada sigue la dirección preferencial. Mientras que la otra trayectoria corresponde a una trayectoria (a 45° de la dirección preferencial) y la del peor escenario (zona restringida), las cuales se observan en la Figura 7.

A través de estos experimentos, se busca una evaluación del impacto en el aprendizaje y la performance del sistema frente a diferentes entornos.

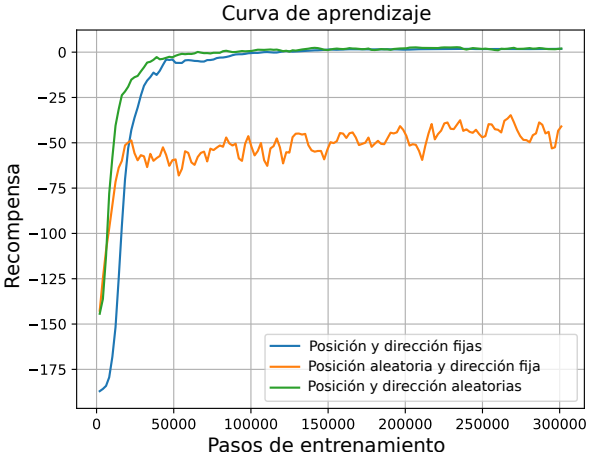

Figura 5: Curva de entrenamiento del valor de recompensa acumulada de tres casos de entrenamiento diferentes: a) Posición de agarre fija y dirección preferencial; b) Posición de agarre aleatoria y dirección preferencial fija; c) Posición de agarre y dirección preferencial aleatorias.

4. Resultados

Los resultados de los experimentos introducidos en la sección 3.2 se presentan a continuación. La curva de aprendizaje de los diferentes casos de entrenamiento se presentan en la Figura 5. Los valores de convergencia principales se muestran en la Figura 6, donde se puede destacar que cuanto más desestructurado es el caso, menor es el valor de la fuerza en estado estacionario. Para el caso de posición de agarre fija y dirección preferencial, la recompensa acumulada es de aproximadamente 7-8[N]; para el caso de posición de agarre aleatoria y dirección preferencial fija, es de 8-9[N]; y para posición de agarre y dirección preferencial aleatorias, es de 16-22[N].

Para la prueba de evaluación de la estrategia aprendida por el robot, se presenta la resultante de la fuerza en la Figura 7. Se puede observar que, para el caso de posición de agarre fija y dirección preferencial fija, la trayectoria ideal (Azul) presenta la menor fuerza ejercida, la trayectoria de 45º (Naranja) ejerce la mayor fuerza, y que la estrategia aprendida se encuentra entre ambas, pero cerca de la trayectoria ideal, alcanzando una fuerza estacionaria similar. Finalmente, se observa que la trayectoria del agente (Verde) se aproxima a la trayectoria ideal.

Método de entreno	Recompensa media del episodio	Fuerza media	Fuerza Final
*Punto de agarre fijo dirección de extracción fija	3.58	7.43 [N]	7.63[N]
*Punto de agarre aleatorio dirección de extracción fija	-2.96	8.37[N]	8.46[N]
*Punto de agarre aleatorio dirección de extracción aleatoria	-42.44	16.40[N]	21.47[N]
Trayectoria Ideal	-	4.24[N]	-
Trayectoria a 45º	-	31.64[N]	-

Figura 6: Resultados del sistema robótico en diferentes entornos de simulación y testeo.

5. Discusión

El uso del algoritmo de control diseñado, presenta resultados satisfactorios para la tarea de extracción de elementos flexibles. El control basado en aprendizaje por refuerzo demostró la capacidad de adquirir habilidades de extracción, encontrando una trayectoria adecuada que minimiza las fuerzas de interacción (Figura 7).

A partir de los resultados obtenidos en las Figuras 6 y 7, se observa que la estrategia aprendida logra fuerzas menores que siguiendo una metodología clásica (trayectoria naranja). De esta manera, luego de la etapa de entrenamiento, el robot mostró un entendimiento de los requisitos de fuerza óptimos. De esta manera se disminuye el estrés mecánico impuesto a los componentes extraidos, evitando daños y extendiendo su vida útil.

El controlador muestra la capacidad para aprender la fuerza óptima requerida para cada tarea de desensamblaje a través de episodios de entrenamiento iterativos. Un aumento constante en el valor de recompensa acumulada significa un aumento en la distancia del efector desde el punto de agarre, seguido de una reducción en la fuerza ejercida. Estos dos aspectos se interpretan como una acción de extracción ejecutada en una trayectoria de baja fuerza. Con este comportamiento, se puede considerar que la función de recompensa propuesta es satisfactoria, ya que conduce a trayectorias de extracción de baja fuerza, y la posición final estacionaria corresponde al punto donde la función de recompensa (Figura 2) alcanza el valor máximo.

A partir de la Figura 5 y 6, se puede ver que en todos los escenarios el robot aprende a encontrar una trayectoria que disminuye la fuerza ejercida incluso en escenarios más desestructurados. Para el entorno más desestructurado (tercer caso), la fuerza media durante toda la trayectoria de extracción disminuye en un 30 % en comparación con la trayectoria de extracción a 45º. Otro aspecto notable de los diferentes escenarios de entrenamiento es la diferencia en la recompensa acumulada y el valor de la fuerza media. Pudiendose interpretar de la siguiente manera: en el primer caso, el entorno siempre es el mismo (misma posición y dirección de extracción), y una vez que el agente aprende la configuración, se requieren las mismas acciones. En otros casos, hasta que el agente descubre la configuración específica del caso actual, el agente recibe recompensas más bajas, indicando trayectorias incorrectas. Por lo tanto, en escenarios más desestructurados con posiciones de agarre y direcciónes de extracción preferencials aleatorias, el agente recibe las recompensas más bajas. Sin embargo, en todos los escenarios el robot aprende a realizar la tarea de manera satisfactoria reflejando un aprendizaje correcto de la tarea.

6. Conclusiones

Este trabajo propone una arquitectura de control basada en técnicas de aprendizaje por refuerzo para realizar desensamblaje de elementos flexibles. El controlador responde a los cambios dinámicos de las fuerzas de interacción durante la ejecución de la tarea. Así el sistema robótico puede operar de manera efectiva en estos entornos desestructurados, asegurando trayectorias de extracción de baja fuerza.

La capacidad del algoritmo de aprendizaje por refuerzo para generalizar en diferentes situaciones demuestra su potencial para aplicaciones reales. Esta capacidad es crucial para la escalabilidad y adaptabilidad de los sistemas robóticos en diversos procesos como la remanufactura y procesos mecánicos.

Los resultados indican una reducción en la fuerza ejercida durante el entrenamiento con el modelo de aprendizaje por refuerzo propuesto. Además, el sistema muestra adaptabilidad a diferentes entornos, mejorando la eficiencia y contribuyendo a la seguridad general del proceso de desmontaje. Por lo tanto, esta investigación destaca un enfoque prometedor para integrar técnicas de aprendizaje por refuerzo para lograr procesos de desmontaje eficientes y seguros.

Futuras investigaciones se centrarán en explorar más a fondo el comportamiento del modelo del algoritmo de aprendizaje por refuerzo al probar diferentes tipos de algoritmos (Soft Actor Critic y Deep Deterministic Policy Gradient), aumentar el número de observaciones e incorporar estimadores de tarea más avanzados. Así mismo también se realizarán las implementaciones en escenarios de desmontaje en el mundo real, abordando los desafíos asociados con las diferencias entre la simulación y la realidad.

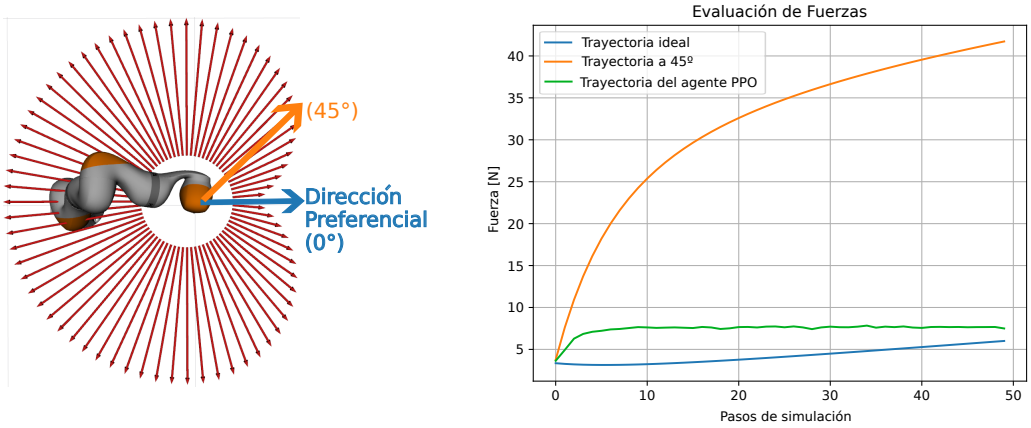

Figura 7: Fuerza resultante del efector del robot durante una tarea de extracción de un elemento flexible en un entorno de prueba con posición de agarre fija y dirección de extracción preferencial fija. Los trayectorias comparadas corresponden a la trayectoria que sigue la dirección preferencial, otra que se desvia a 45° de esta y finalmente la realizada por el agente PPO entrenada en las mismas condiciones de testeo.

Agradecimientos

Este trabajo ha sido financiado por el programa de investigación y innovación Horizon 2020 de la Unión Europea, bajo el acuerdo de beca Marie Sklodowska-Curie No 955681 y por miembros del grupo de investigación de Sensorización Virtual de la Universidad del País Vasco (Basque Goverment Ref. IT1726-22).

Referencias

Beltran-Hernandez, C. C., Petit, D., Ramirez-Alpizar, I. G., Harada, K., 10 2020. Variable compliance control for robotic peg-in-hole assembly: A deep-reinforcement-learning approach. Applied Sciences (Switzerland) 10, 1–17.
DOI: 10.3390/app10196923

Kristensen, C. B., Sørensen, F. A., Nielsen, H. B., Andersen, M. S., Bendtsen, S. P., Bøgh, S., 2019. Towards a robot simulation framework for e-waste disassembly using reinforcement learning. Vol. 38. Elsevier B.V., pp. 225–232.
DOI: 10.1016/j.promfg.2020.01.030

Kroemer, O., Niekum, S., Konidaris, G., 2021. A review of robot learning for manipulation: Challenges, representations, and algorithms.
DOI: 10.48550/arXiv.1907.03146

Kurilova-Palisaitiene, J., Sundin, E., Poksinska, B., 1 2018. Remanufacturing challenges and possible lean improvements. Journal of Cleaner Production 172, 3225–3236.
DOI: 10.1016/j.jclepro.2017.11.023

Lillicrap, T. P., Hunt, J. J., Pritzel, A., Heess, N., Erez, T., Tassa, Y., Silver, D., Wierstra, D., 9 2015. Continuous control with deep reinforcement learning.
DOI: 10.48550/arXiv.1509.02971

Paz, B. T. S., Sorrosal, G., Mancisidor, A., Cabanes, I., 8 2023. Control híbrido para la manipulación robótica de elementos flexibles. Servizo de Publicacións. Universidade da Coruña, pp. 768–772.
DOI: 10.17979/spudc.9788497498609.768

Poschmann, H., Brüggemann, H., Goldmann, D., 4 2020. Disassembly 4.0: A review on using robotics in disassembly tasks as a way of automation.
DOI: 10.1002/cite.201900107

Sutton, R. S., Barto, A. G., 2018. Reinforcement learning : an introduction, 2nd Edition. The MIT Press.

Zachares, P. A., Lee, M. A., Lian, W., Bohg, J., 10 2021. Interpreting contact interactions to overcome failure in robot assembly tasks. Institute of Electrical and Electronics Engineers (IEEE), pp. 3410–3417.
DOI: 10.1109/icra48506.2021.9560825

Zhou, Z., Ni, P., Zhu, X., Cao, Q., 11 2021. Compliant robotic assembly based on deep reinforcement learning. Institute of Electrical and Electronics Engineers (IEEE), pp. 6–9, la precision requerida en operaciones de peg-in-hole siguen siendo un problema abierto.
DOI: 10.1109/mlise54096.2021.00009