



BRAZO ROBÓTICO CON CÁMARA ÚNICA PARA RECOLECTAR FRESAS MEDIANTE DEEP LEARNING

Trabajo Fin de Grado
David Campoamor Medrano

31 de mayo de 2025



©2025 David Campoamor Medrano

Algunos derechos reservados.

Este trabajo se entrega bajo licencia CC-BY-SA 4.0.

Contenidos

1. Introducción
2. Estado del arte
3. Objetivos
4. Plataforma de desarrollo
5. Descripción del sistema
6. Conclusiones

Introducción

Los robots y la robótica

- Según la Federación Internacional de Robots (IFR), se define robot como mecanismo accionado programado con cierto grado de autonomía para realizar tareas de locomoción, manipulación o posicionamiento.
- Este término fue utilizado por primera vez por Karel Capek en 1920, en el sentido de la obligatoriedad, entendido como servidumbre, trabajo forzado o esclavitud.
- Isaac Asimov utilizó por primera vez el término robótica y postuló las cuatro leyes de la robótica en 1950, coincidiendo con el apogeo de la robótica moderna.
- Atendiendo al propósito y al contexto en el que se utilicen estos robots, se fueron creando varios grupos en función de los que clasificarlos: robots industriales, robots de servicio y robots en medicina.

Inteligencia Artificial (IA)

- Es un área multidisciplinaria de la ciencia donde se realizan sistemas que tratan de hacer tareas y resolver problemas como lo hace un humano.
- Trata de simular de manera artificial las formas de pensamiento y de trabajar del cerebro para la toma de decisiones.

Visión Artificial

- Se define como la ciencia de programar un ordenador para procesar imágenes o vídeos e incluso entenderlos.
- La imagen percibida pasa por los procesos de obtención, caracterización e interpretación de información de imágenes.
- Para mejorar aún más la eficacia de los sistemas de visión artificial, se recurre al aprendizaje automático o *machine learning* (ML).

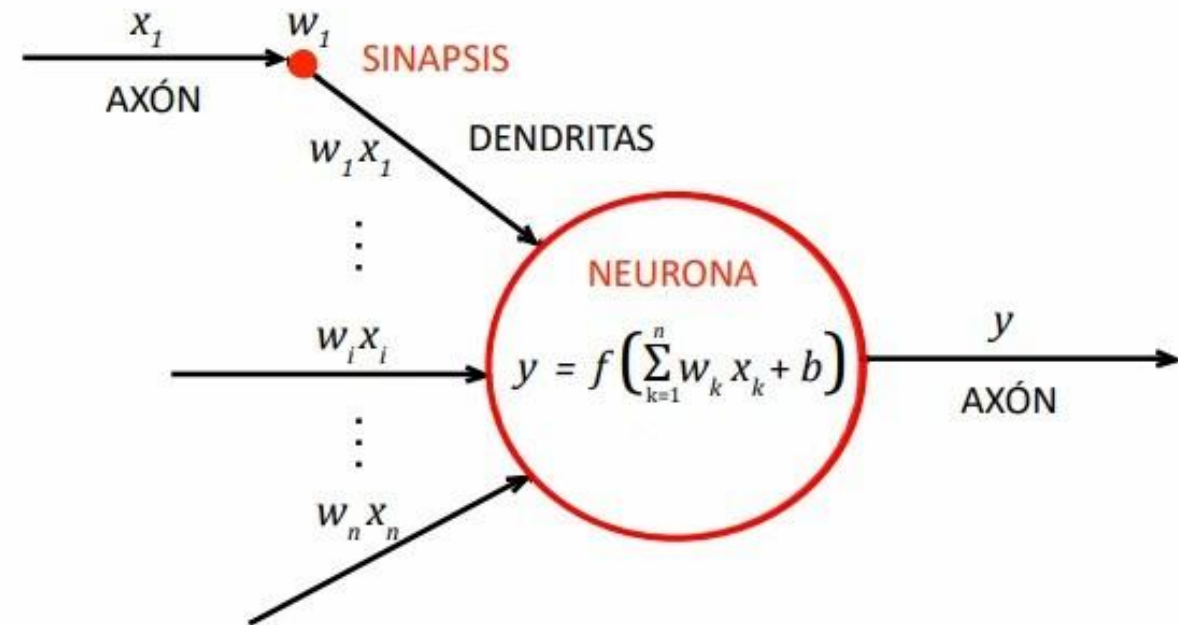
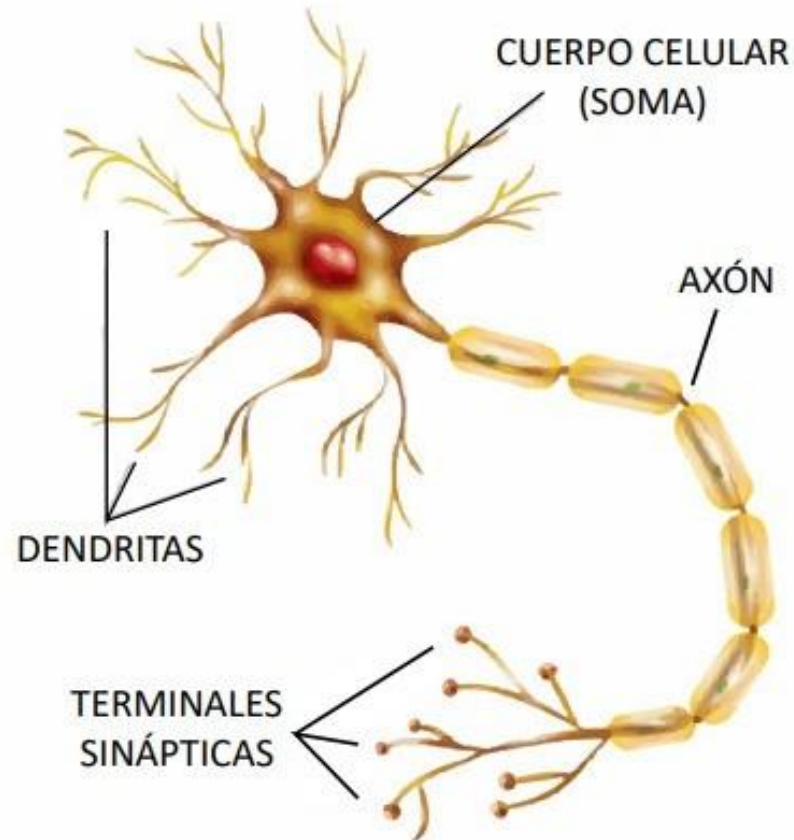
Machine Learning (ML)

- Es una rama en evolución de la IA que se encarga de generar algoritmos que tienen la capacidad de aprender del entorno circundante y no tener que programarlos de manera explícita.
- Este aprendizaje se realiza a partir de la construcción de modelos analíticos.
- Dependiendo de la tarea de aprendizaje, existen varias clases de algoritmos de ML:
 - Aprendizaje superficial o *Shallow Machine Learning*: Se centra en algoritmos simples para tareas específicas.
 - Aprendizaje profundo o *Deep Learning*: Utiliza la construcción y entrenamiento de Redes Neuronales Artificiales (RNA).

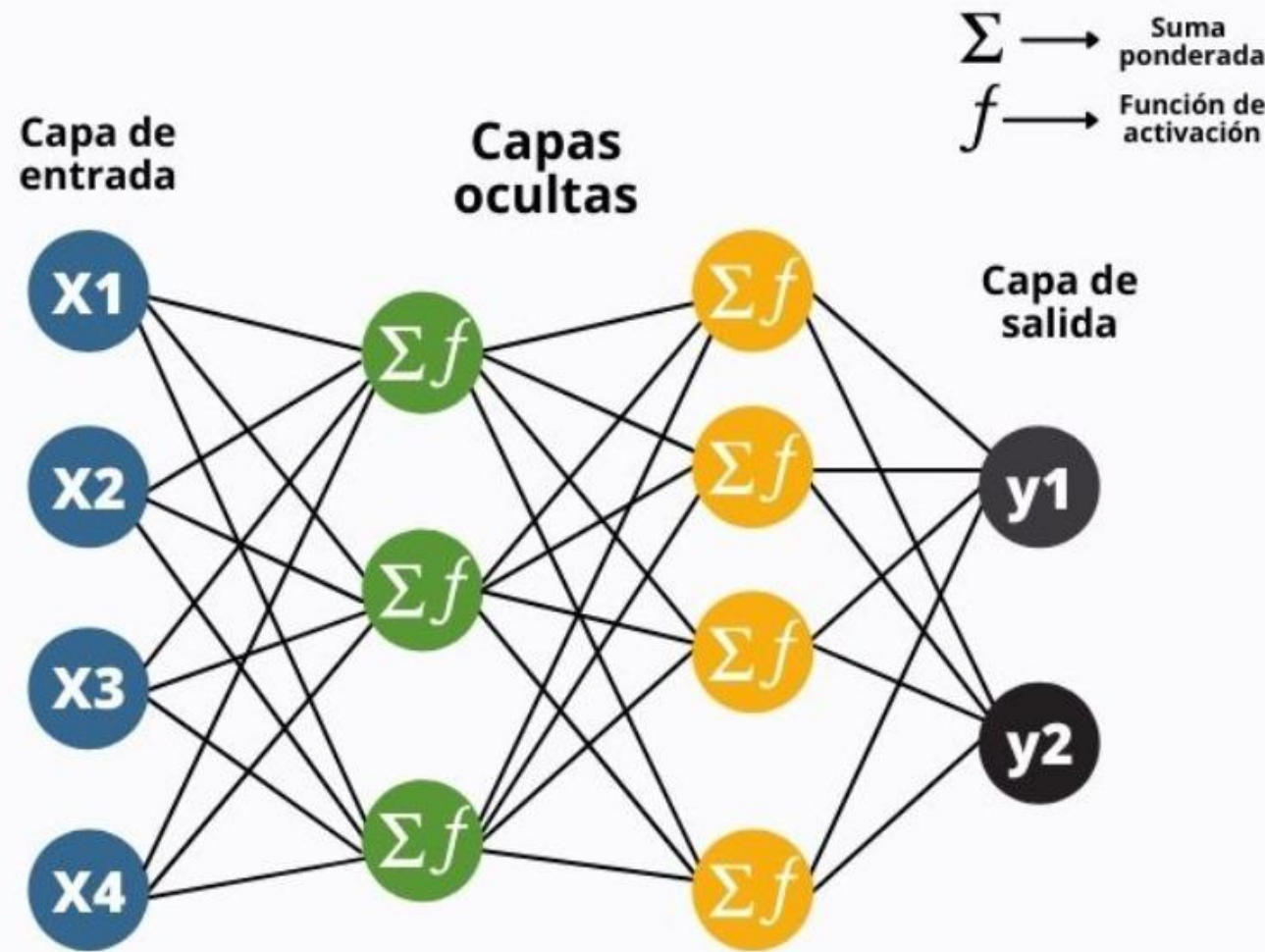
Deep Learning

- Constituye una rama de la IA, incluida dentro del Machine Learning, cuyos modelos computacionales se inspiran en el funcionamiento del cerebro humano.
- Se diseñan con el propósito de adquirir conocimientos y llevar a cabo tareas específicas mediante el procesamiento de datos.
- Estas Redes Neuronales Artificiales (RNA) están inspiradas en las redes neuronales biológicas del cerebro humano, ya que aprenden de la experiencia, generalizan de ejemplos previos a ejemplos nuevos, y abstraen las características principales de una serie de datos.
- En las RNA, la unidad análoga a la neurona biológica es el elemento procesador PE (Process Element).
- Un PE tiene varias entradas y las combina, normalmente, con una suma básica. La suma de las entradas es modificada por una función de transferencia y el valor de la salida de esta función de transferencia se pasa directamente a la salida del elemento procesador.

Deep Learning



Deep Learning

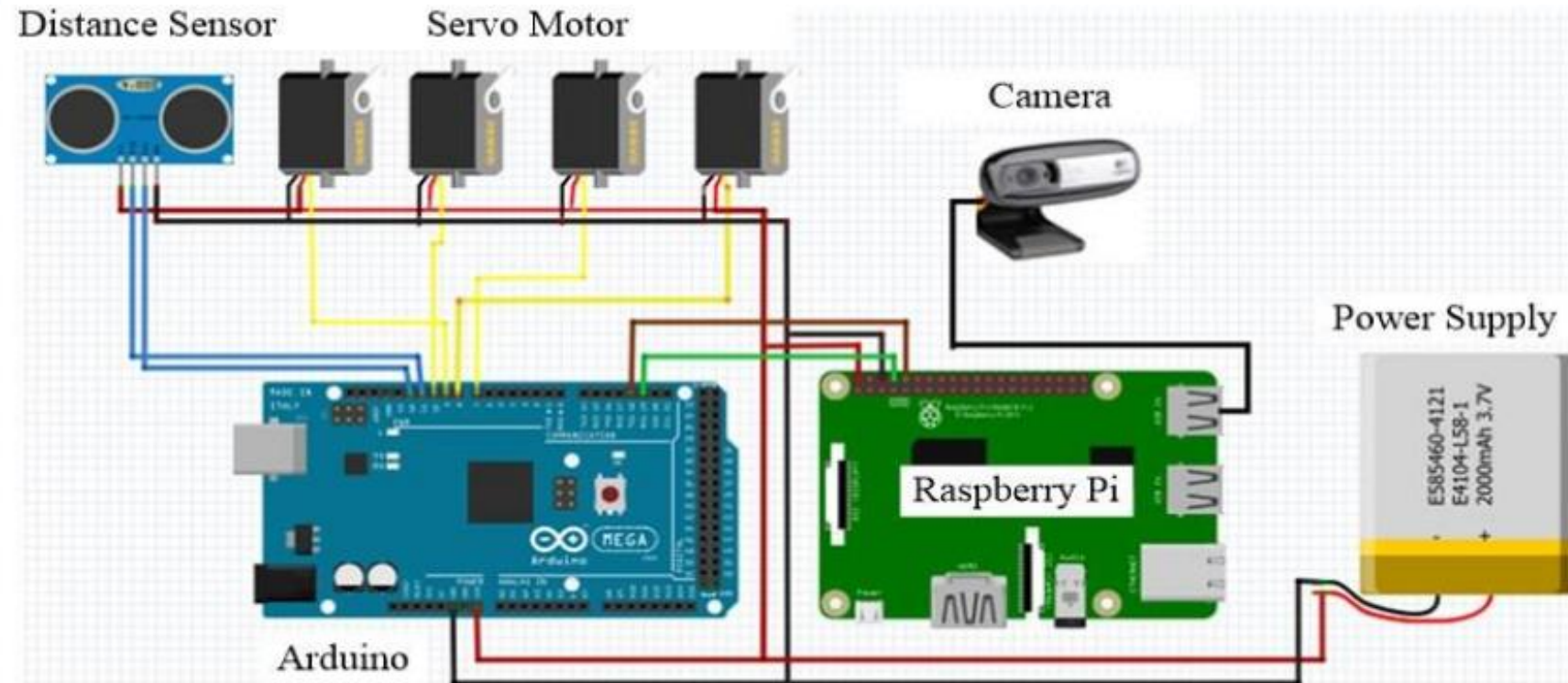
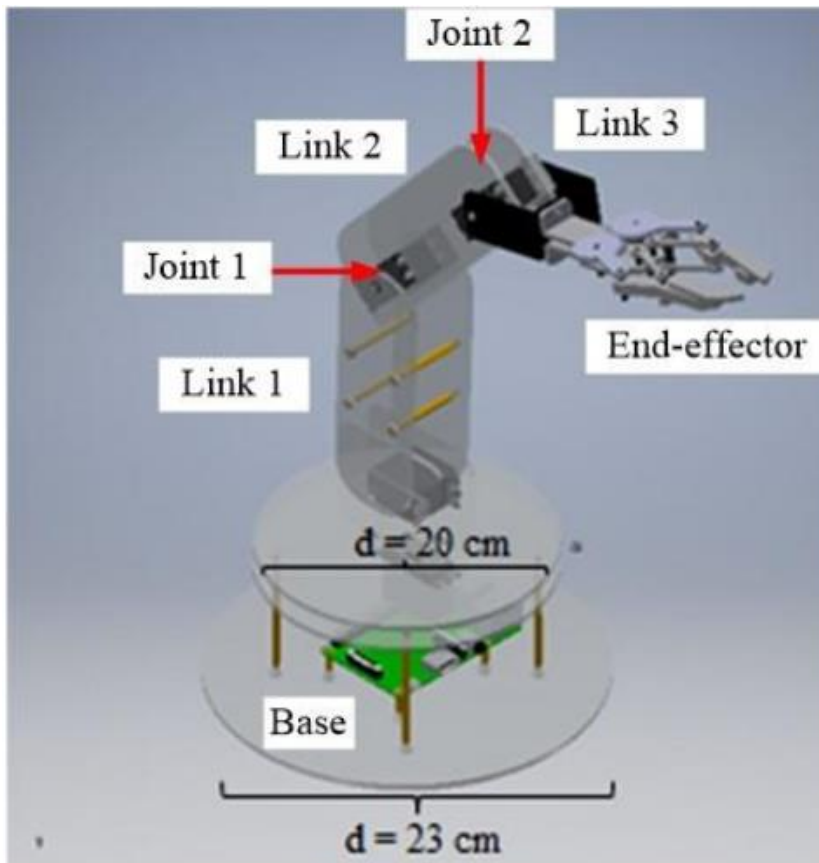


Estado del arte

Sistemas existentes | Proyectos universitarios

- En 2020, la Universidad Politécnica Negeri Sriwijaya de Indonesia, analizó el empleo de un robot recolector de tomates rojos y verdes como proyecto piloto.
- A pesar de la diferencia en el tipo de cultivo, el estudio resulta relevante debido a la similitud en los principios de detección y recolección de frutos.
- Al tratarse de un prototipo básico lo convierte en un punto de partida útil para proyectos de investigación enfocados en la automatización agrícola aplicada a fresas.

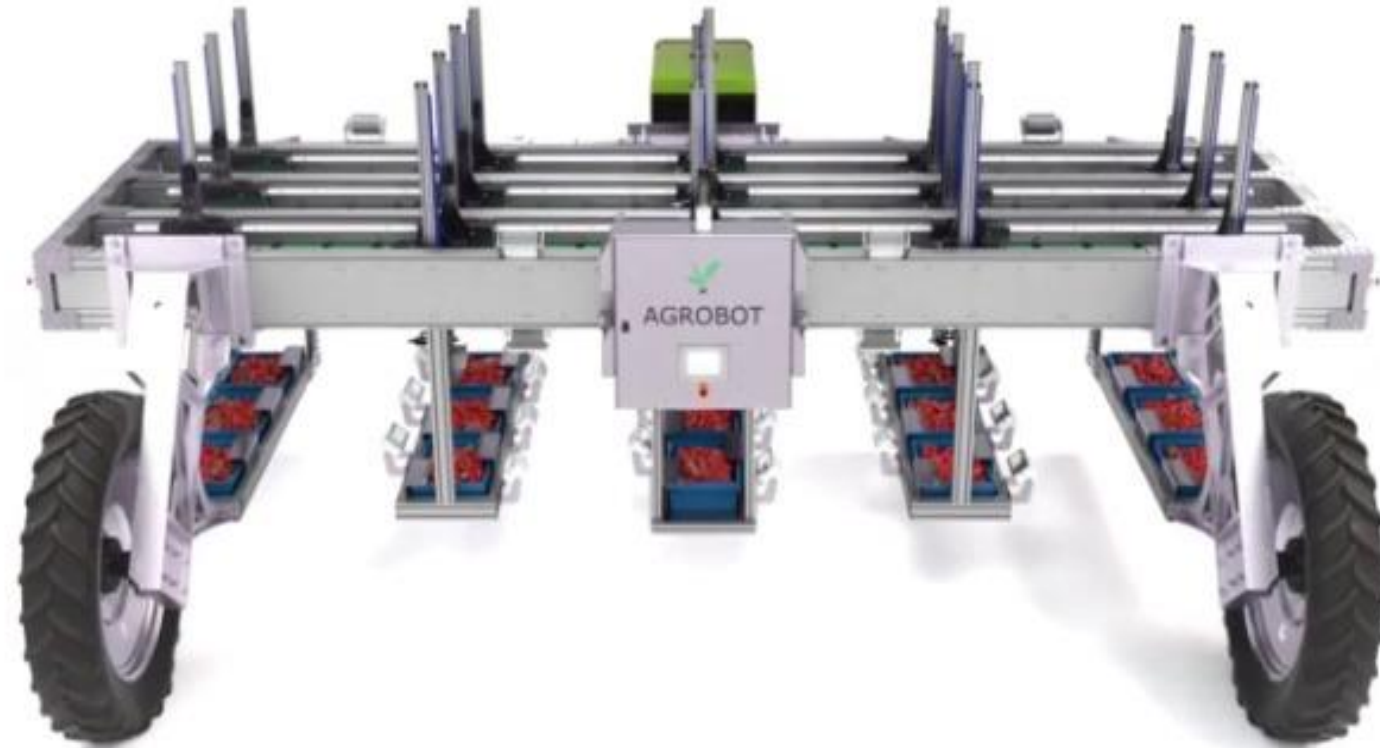
Sistemas existentes | Proyectos universitarios



Sistemas existentes | Agrobot

- Primer prototipo en 2009, en Huelva, a través del Centro de Innovación y Tecnología. Identificaba los frutos maduros y los recogía sin dañarlos, siendo capaz de clasificarlas y colocarlas en los envases que recorren las cintas transportadoras.
- El Agrobot SW 6010, es una cosechadora de fresas que es capaz de recoger de la mata solo la fruta que está madura mediante inteligencia artificial con 30 brazos robóticos que incorporan una cámara con visión artificial que detecta el grado de madurez de la fruta en tiempo real.
- Fabricado en acero inoxidable y aluminio, el Agrobot E-Series se adapta a cualquier configuración agrícola. Con sensores de profundidad infrarrojos y en color integrados, que evalúan el grado de madurez de la fruta, incorporando a su vez, sensores LiDAR que se encargan de la seguridad de los trabajadores circundantes.

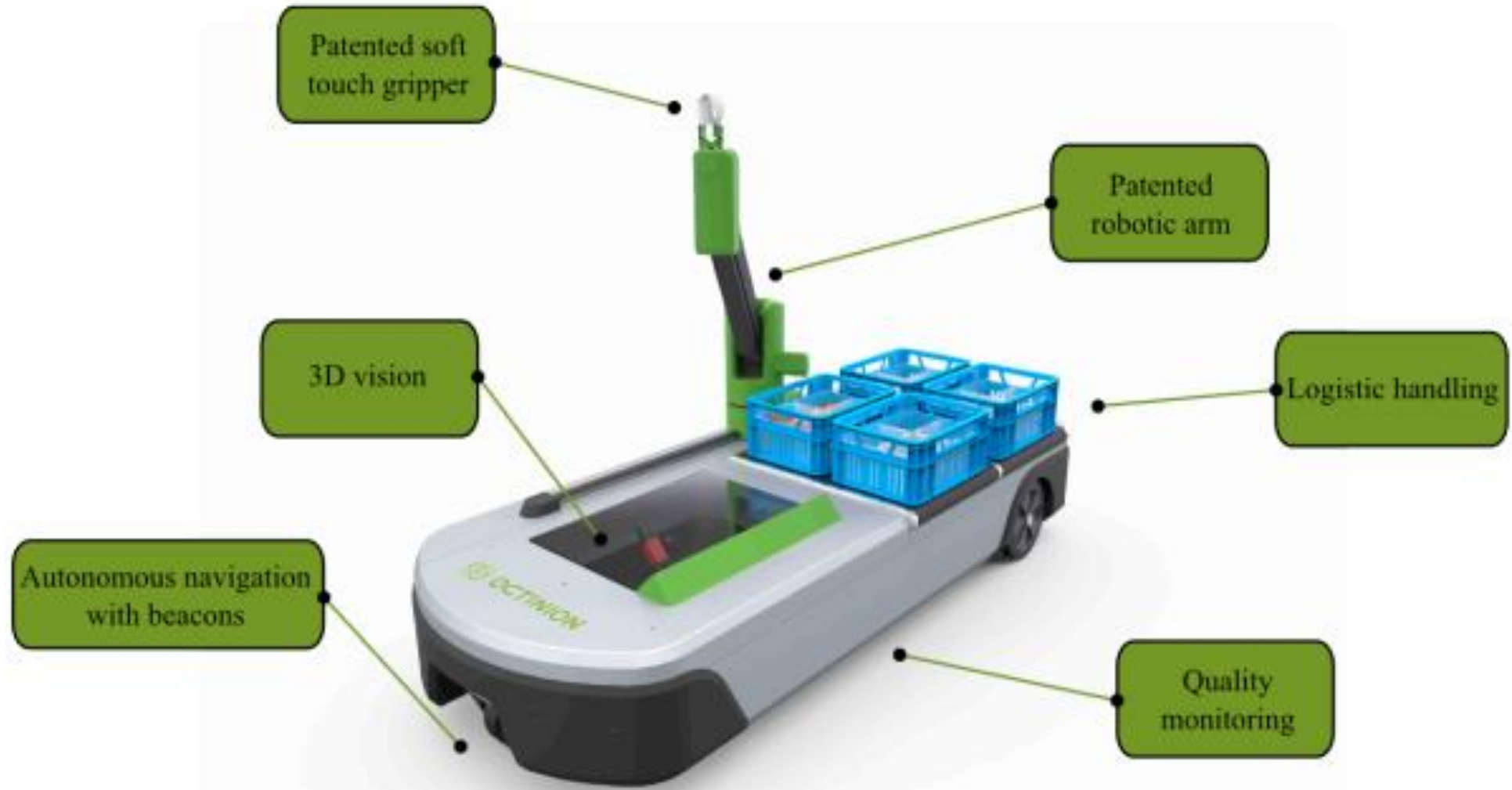
Sistemas existentes | Agrobot



Sistemas existentes | Octinion

- La empresa belga de I+D agrícola Octinion desarrolló un prototipo de robot recolector de fresas en 2017, que recoge los frutos de forma totalmente autónoma basándose en el método de cultivo habitual (sobremesa).
- Este robot, está formado por:
 - Una plataforma eléctrica con una batería recargable
 - Un sistema de localización (codificadores de rueda, giroscopio y sistema de posicionamiento en interiores de banda ultra ancha (UWB))
 - Tres cámaras RGB para la detección de las fresas mediante visión artificial
 - Un brazo robótico diseñado a medida
 - Pinza que se acopla al extremo del brazo robótico
 - Un módulo de gestión o manipulación logística consistente en varias cestas
 - Un módulo de control de calidad que clasifica las fresas detectadas en función de su madurez, forma y tamaño.
- Todo esto le permite al robot recoger al menos el 70 % de las fresas maduras sin dañarlas y depositarla en una cesta (caja en la que se colocan las fresas).

Sistemas existentes | Octinion



Sistemas existentes | Dogtooth

- El sistema Dogtooth de Cambridge (Reino Unido) fue desarrollado para mejorar las operaciones de recogida de frutos rojos siendo capaz de navegar por hileras de fresas y frambuesas, detectar y localizar las maduras, así como recoger y comprobar la fruta antes de colocarla en un cesto.
- Parte de un robot industrial
- Se encuentra en fase de prototipo

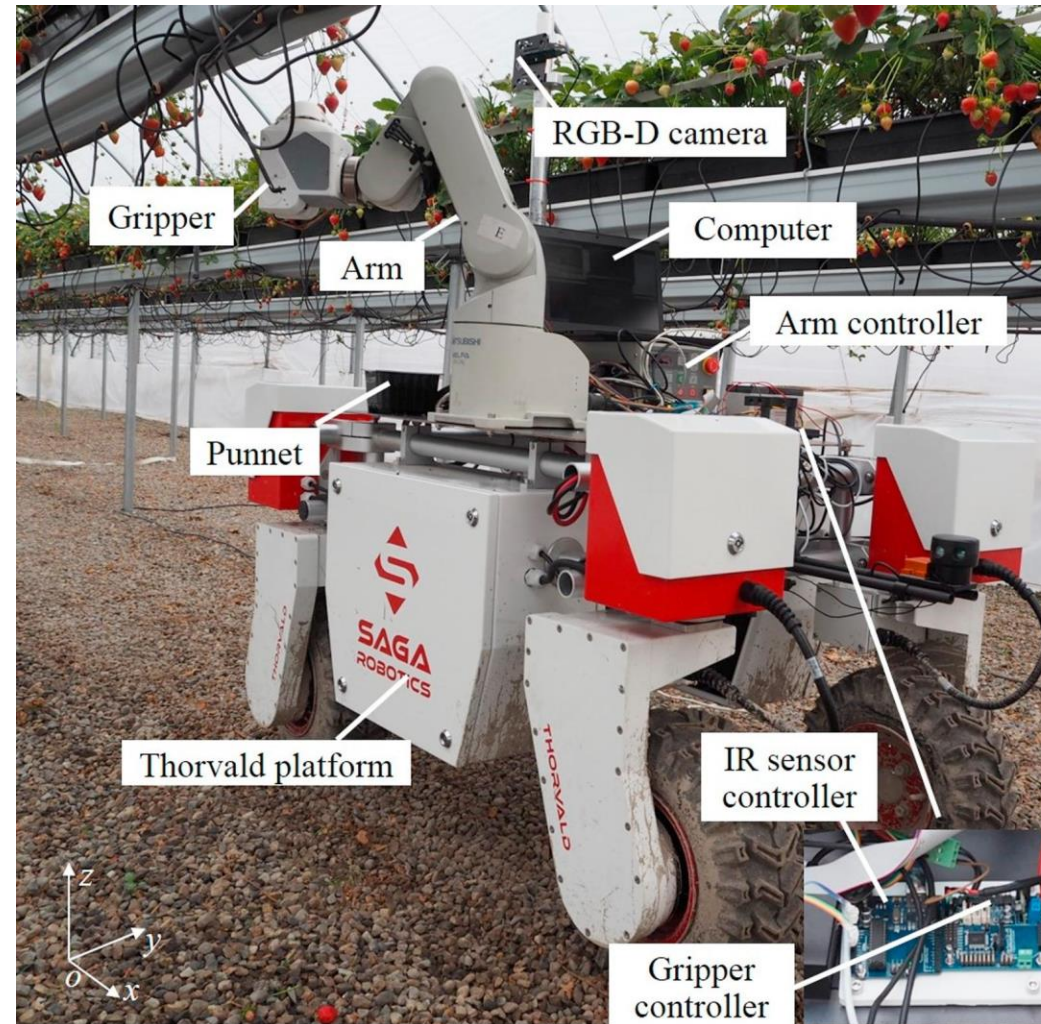
Sistemas existentes | Dogtooth



Sistemas existentes | Thorvald

- La plataforma Thorvald de SAGA Robotics, está desarrollada para el uso de un robot para la recolección de fresas cultivadas en invernaderos.
- El robot encargado de realizar esta tarea está compuesto por una pinza montada en un brazo industrial que a su vez está montado en una base móvil junto con una cámara RGB-D.
- El sistema de visión se basa en el umbral de color combinado con el cribado del área del objeto y el rango de profundidad para seleccionar las fresas maduras y alcanzables.
- La pinza está diseñada para apuntar a la fruta y no al tallo, por lo que sólo requiere la ubicación de la fruta para la recolección.
- Esta pinza también está equipada con sensores internos, por lo que la pinza puede detectar y corregir errores de posición, y es resistente a los errores de localización introducidos por el módulo de visión.

Sistemas existentes | Thorvald



Objetivos

Descripción del problema y objetivo principal

- La necesidad de implementar soluciones tecnológicas que automaticen y optimicen las tareas de recolección incrementando la eficiencia, mejorando la calidad del producto y disminuyendo costes asociados surge del problema de la escasa mano de obra disponible y el proceso manual que esto conlleva.
- **Objetivo principal:** utilizar un robot colaborativo que junto con el sistema de detección elaborado con materiales de bajo coste, sea capaz de reconocer las fresas maduras de un sistema de cultivo agrícola vertical, para su posterior recolección por el brazo robótico, gracias a la comunicación establecida entre el sistema de visión y el robot.

Subobjetivos

Con el fin de alcanzar este objetivo principal, se han establecido los siguientes subobjetivos:

1. Investigar las soluciones actuales que cumplen con las características y objetivos establecidos.
2. Seleccionar la técnica de inteligencia artificial de reconocimiento de frutas y seleccionar los componentes hardware necesarios para desarrollar el sistema de visión de bajo coste más eficiente.
3. Optimizar la técnica escogida y adaptarla de tal manera que sea capaz de funcionar en nuestra plataforma.
4. Realizar el entrenamiento con varios algoritmos de Machine Learning de clasificación.
5. Seleccionar el protocolo de comunicación entre el sistema de visión y el robot y llevar a cabo pruebas.
6. Dar soporte software al robot mediante un sistema de reconocimiento de fresas, que guarde las posiciones y la distancia de estas a la posición de la cámara, para su posterior envío al brazo robótico.
7. Realizar pruebas de la aplicación final.

Plan de trabajo

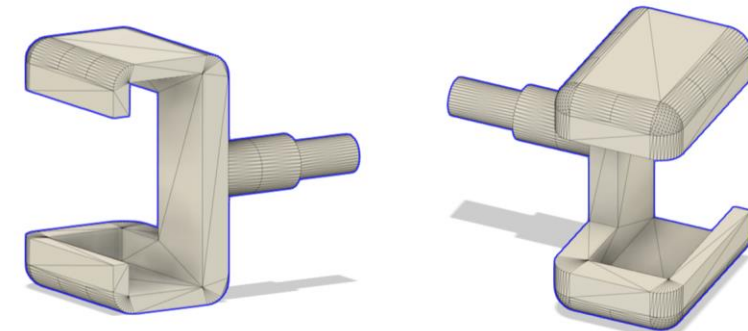
1. Investigación inicial
2. Diseño y desarrollo del sistema de visión artificial
3. Pruebas en entorno simulado
4. Pruebas en entorno real
5. Escritura de la memoria

Plataforma de desarrollo

Hardware

- Cámara Logitech C270 HD
- Soporte de brazo articulado
- Soporte de impresión 3D
- Ordenador principal: Lenovo Legion 5 IMH05
- Robot de Universal Robots de la gama e-series
- Comunicaciones: cables de red y switch

Hardware



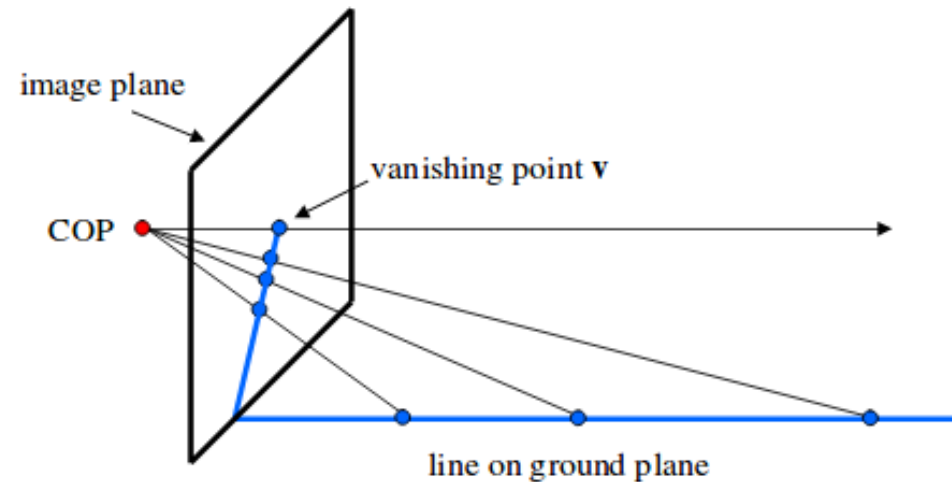
Software

- Ubuntu
- Polyscope
- Anaconda
- Python
- PyTorch
- NumPy
- OpenCV
- OpenGL
- SocketTest
- XML-RPC
- YOLOv3

Descripción del sistema

Hipótesis suelo adaptada al plano vertical

- La estimación de coordenadas tridimensionales se basa en la hipótesis suelo, y se realiza a partir de una única cámara estenopeica RGB fija para poder realizar esta estimación sin recurrir a sensores adicionales.
- Esta técnica de simplificación asume que los objetos de interés se encuentran sobre un plano conocido y fijo respecto a la cámara.



Hipótesis suelo adaptada al plano vertical

- La cámara sigue el modelo *pinhole* o estenopeico, lo que implica que la proyección de un punto en el espacio bidimensional de una imagen al tridimensional se realiza usando los parámetros intrínsecos de la propia cámara y parámetros extrínsecos como la rotación o la traslación de la misma.
- Al asumir que los puntos están sobre el plano $Z=0$ mediante la hipótesis suelo, este modelo estenopeico permite estimar las coordenadas X e Y reales a partir de la imagen captada por la cámara mediante una serie de fórmulas matemáticas.

Parámetros de cámara	Definición
$K (3 \times 3)$	Parámetros intrínsecos
$R (3 \times 3)$	Rotación de la cámara
$T (3 \times 1)$	Traslación de la cámara

$$w \cdot \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = K \cdot [R | T] \cdot \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}$$

- En esta ecuación se observa la relación entre un punto tridimensional (X, Y, Z) y su proyección bidimensional (u, v) multiplicada por un factor de escala w que representa la profundidad del punto proyectado en la imagen captada por la cámara.

Ecuaciones parámetros intrínsecos y extrínsecos

$$K = \begin{pmatrix} F_x & 0 & C_x \\ 0 & F_y & C_y \\ 0 & 0 & 1 \end{pmatrix}$$

$$T = \begin{pmatrix} t_x \\ t_y \\ t_z \end{pmatrix}$$

$$R(\theta)_X = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos(\theta) & \sin(\theta) \\ 0 & -\sin(\theta) & \cos(\theta) \end{pmatrix}$$

$$R(\theta)_Y = \begin{pmatrix} \cos(\theta) & 0 & -\sin(\theta) \\ 0 & 1 & 0 \\ \sin(\theta) & 0 & \cos(\theta) \end{pmatrix}$$

$$R(\theta)_Z = \begin{pmatrix} \cos(\theta) & -\sin(\theta) & 0 \\ \sin(\theta) & \cos(\theta) & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

Ecuaciones obtención coordenadas y distancia

$$w \cdot \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = K \cdot [R | T] \cdot \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}$$

$$Z = Z_0$$

$$\text{distancia} = \sqrt{X^2 + Y^2 + Z^2}$$

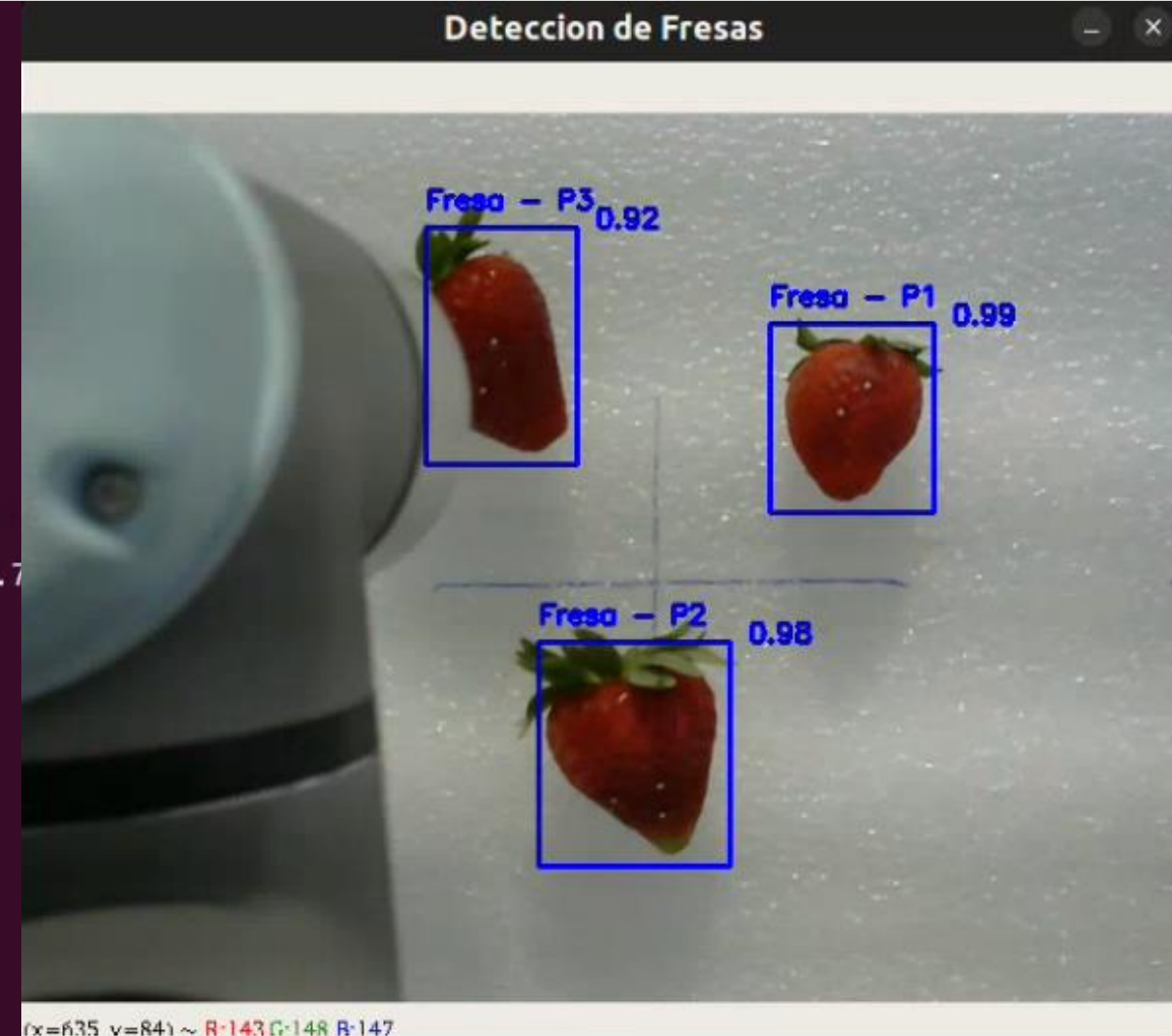
$$\begin{bmatrix} X \\ Y \\ Z_0 \end{bmatrix} = R^{-1} \cdot \left(K^{-1} \cdot w \cdot \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} - T \right)$$

Detección de fresas mediante Deep Learning

- La detección de las fresas maduras se ha resuelto mediante el uso de técnicas de visión artificial basadas en deep learning, concretamente, se ha utilizado el modelo YOLOv3 (You Only Look Once).
- Este modelo ha sido entrenado específicamente para reconocer fresas maduras en las condiciones del entorno de trabajo y se ha integrado en un script que analiza el flujo de vídeo en tiempo real y devuelve las coordenadas y la distancia a la cámara de cada fresa detectada en la imagen, junto con la clase del objeto y la confianza de detección.
- Para mejorar la precisión a la hora de estimar la posición de la fresa, se ha utilizado el centro del *bounding box* de la fresa detectada como punto de referencia para la proyección sobre el plano vertical, conforme a la hipótesis suelo.

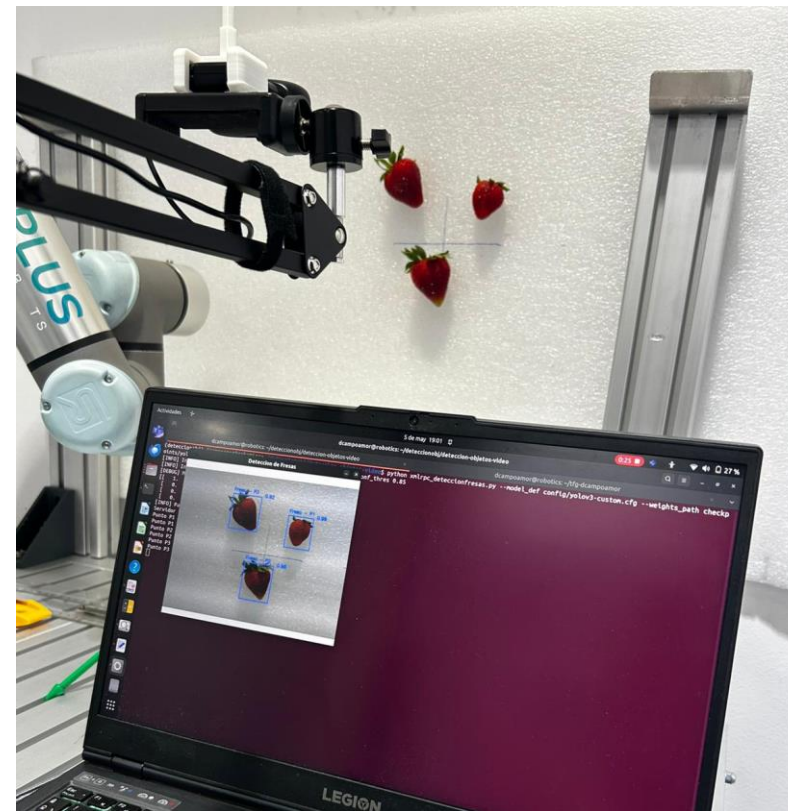
Detección de fresas mediante Deep Learning

```
[INFO] Iniciando programa...
[INFO] Iniciando detección de fresas en el frame actual...
[DEBUG] Matriz RT de la cámara:
[[ 1.  0.  0.  0.]
 [ 0.  1.  0.  0.]
 [ 0.  0.  1. -355.]
 [ 0.  0.  0.  1.]]
[INFO] Parámetros de la cámara cargados correctamente.
Servidor XML-RPC corriendo en el puerto 50000...
Punto P1 - Coordenadas 3D: X=60.77, Y=-31.44, Z=355.00
Punto P1 - Distancia al punto: 349.76 milímetros
Punto P2 - Coordenadas 3D: X=8.43, Y=47.15, Z=355.00
Punto P2 - Distancia al punto: 346.33 milímetros
Punto P3 - Coordenadas 3D: X=-28.20, Y=-48.77, Z=355.00
Punto P3 - Distancia al punto: 347.60 milímetros
[DEBUG] Enviando última posición detectada: (-28.20182882193582, -48.7
192.168.23.214 - - [05/May/2025 19:01:15] "POST / HTTP/1.1" 200 -
Punto P1 - Coordenadas 3D: X=60.93, Y=-31.53, Z=355.00
Punto P1 - Distancia al punto: 349.79 milímetros
Punto P2 - Coordenadas 3D: X=8.96, Y=47.30, Z=355.00
Punto P2 - Distancia al punto: 346.36 milímetros
Punto P1 - Coordenadas 3D: X=60.77, Y=-31.40, Z=355.00
Punto P1 - Distancia al punto: 349.75 milímetros
Punto P2 - Coordenadas 3D: X=7.73, Y=47.31, Z=355.00
Punto P2 - Distancia al punto: 346.33 milímetros
Punto P3 - Coordenadas 3D: X=-24.18, Y=-47.87, Z=355.00
Punto P3 - Distancia al punto: 347.17 milímetros
```



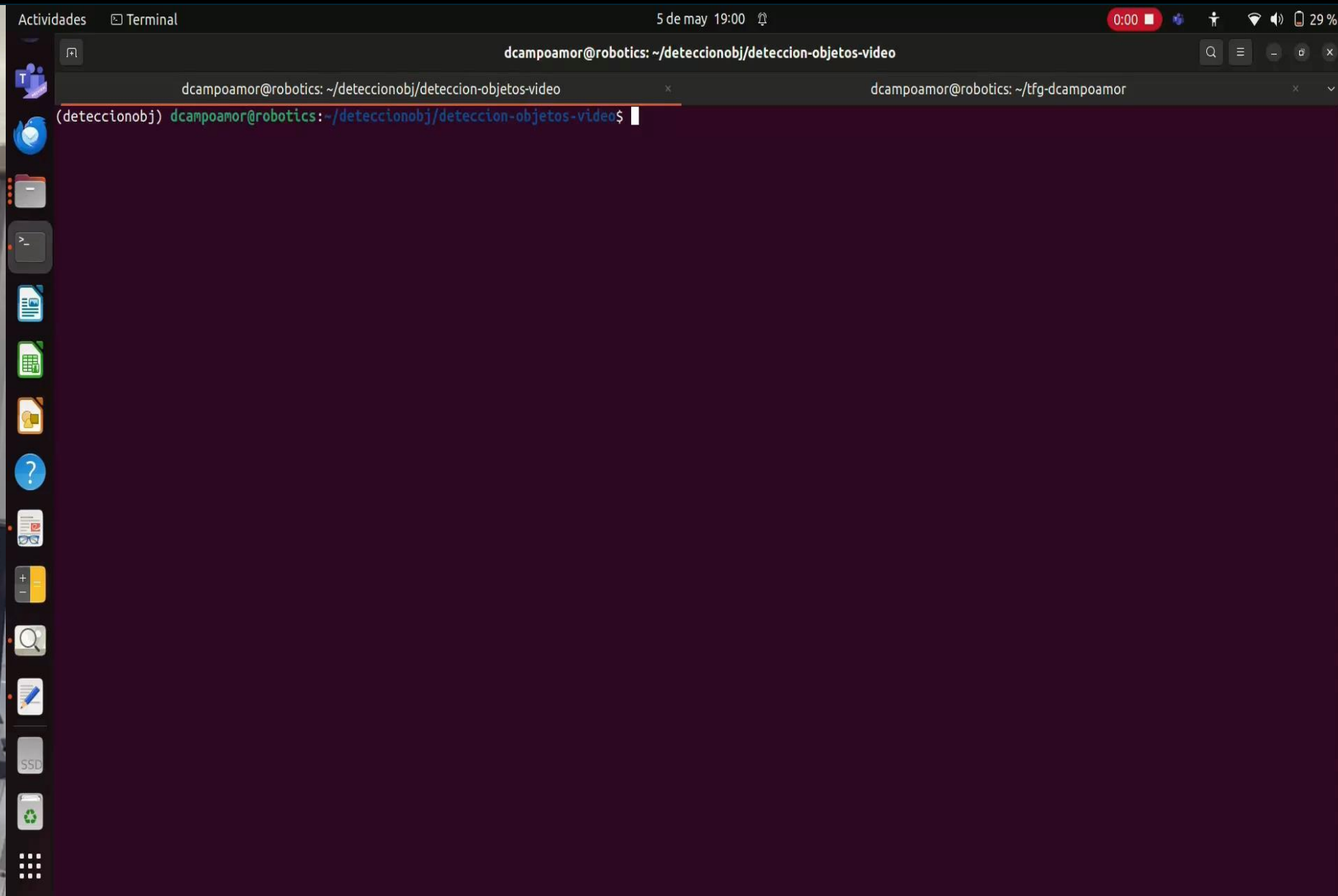
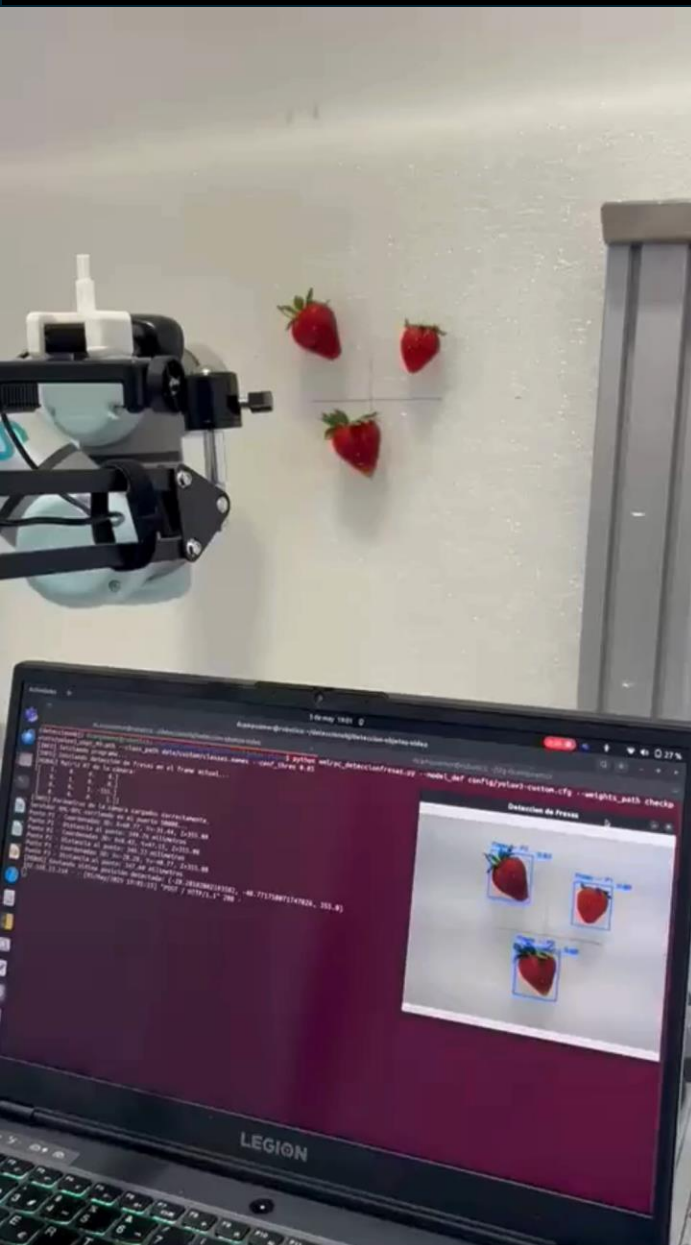
Arquitectura del sistema

- El prototipo final se ha implementado utilizando una combinación de componentes hardware y software.
- Esta integración ha sido diseñada para garantizar la detección precisa de fresas, el cálculo de su localización espacial en coordenadas tridimensionales y la ejecución del movimiento del robot sobre estas posiciones.



Funcionamiento

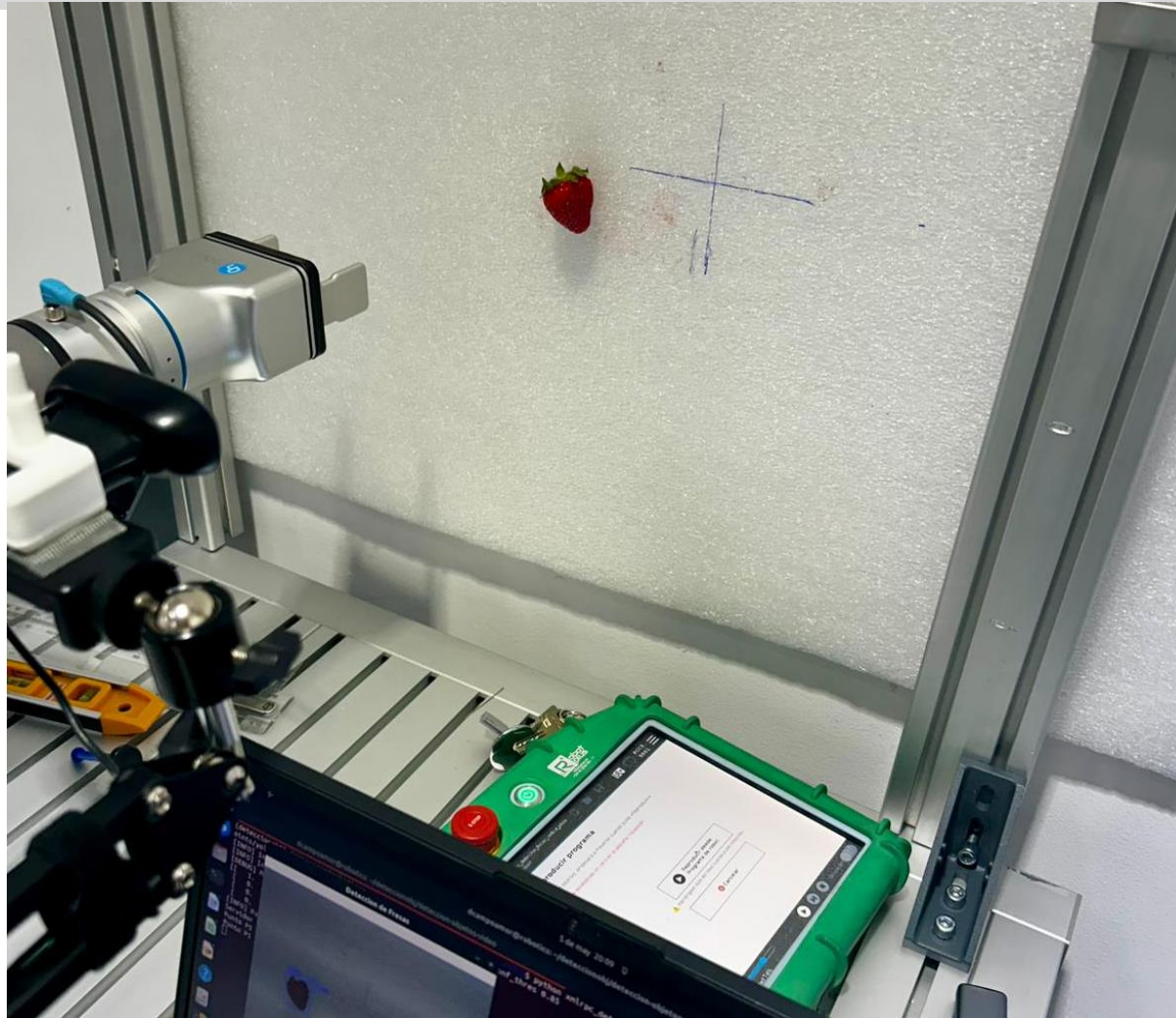
1. La cámara captura imágenes de la escena en tiempo real.
2. El modelo de deep learning detecta las fresas maduras y genera sus posiciones en píxeles gracias al código en Python.
3. Estas coordenadas se proyectan al espacio tridimensional sobre el plano vertical utilizando la función que calcula la intersección del rayo proyectado con el plano Z definido por la posición de la cámara, utilizando como argumento de entrada las coordenadas en píxeles de la detección.
4. El servidor XML-RPC se crea en la dirección IP y puerto especificado, se transmiten las coordenadas al robot y lanza un hilo independiente que queda a la espera de nuevas peticiones.
5. El robot interpreta la posición, calcula una trayectoria y actúa sobre la fresa detectada.



Funcionamiento con efector final

- Para completar la evaluación del prototipo, se realizaron pruebas complementarias utilizando una pinza industrial acoplada al robot con el objetivo de verificar la compatibilidad y robustez del programa desarrollado.
- Estas pruebas permitieron demostrar que el sistema era plenamente funcional y adaptable, y que se podía integrar sin inconvenientes cualquier herramienta de agarre que fuera adecuada para el manejo de fresas, dada su delicadeza.

Funcionamiento con efector final



Conclusiones

Objetivos cumplidos

- Se ha conseguido cumplir con el objetivo principal de desarrollar un sistema de visión artificial de bajo coste, basado en técnicas de inteligencia artificial, capaz de detectar fresas maduras y comunicar su posición y distancia a un brazo robótico para su recolección automatizada.
- Todo ello ha sido probado tanto en entornos simulados como en condiciones reales, demostrando la viabilidad del sistema diseñado.

Subobjetivos cumplidos

1. Se han investigado las soluciones actuales tanto académicas como comerciales relacionadas con la detección de frutos mediante visión artificial .
2. Se ha seleccionado el uso de redes neuronales convolucionales (CNN) a través del modelo YOLOv3, debido a su eficiencia y precisión en tareas de detección en tiempo real.
3. La técnica escogida se ha optimizado y adaptado para funcionar en la plataforma de trabajo, lo que ha requerido la creación de un dataset específico de imágenes de fresas.
4. Se ha realizado el entrenamiento del sistema con distintos algoritmos de clasificación basados en Machine Learning, evaluando su rendimiento mediante pruebas con imágenes reales.
5. Se ha seleccionado el protocolo de comunicación entre el sistema de visión y el robot, implementando un servidor XML-RPC para transmitir de forma efectiva la información de las detecciones.
6. El sistema es capaz de reconocer fresas maduras, calcular su posición en coordenadas del mundo real y estimar su distancia a la cámara, transmitiendo esta información al brazo robótico para su uso operativo.
7. Se ha probado el sistema completo en situaciones tanto simuladas como reales, comprobando su funcionamiento y eficiencia, y sentando las bases para posibles mejoras y aplicaciones futuras.

Líneas futuras

- Entrenamiento con datasets más amplios y variados.
- Optimización del rendimiento en hardware embebido: Adaptar el sistema para funcionar en dispositivos aún más limitados (como Raspberry Pi Zero o NVIDIA Jetson Nano), buscando reducir el consumo energético y el coste del sistema.
- Implementación en entornos agrícolas reales: Validar el sistema en condiciones reales de campo, frente a variables como viento, sombra o vegetación densa para comprobar su fiabilidad y utilidad práctica.
- Diseño o integración de una herramienta de recolección versátil: Investigar o desarrollar una pinza robótica compatible con el sistema de visión y adecuada para manipular distintos tipos de frutos con cuidado y precisión, ampliando así el alcance del sistema a otros cultivos más allá de la fresa.



BRAZO ROBÓTICO CON CÁMARA ÚNICA PARA RECOLECTAR FRESAS MEDIANTE DEEP LEARNING

Trabajo Fin de Grado
David Campoamor Medrano

31 de mayo de 2025