

Título del proyecto

Daniel Gómez

26 de mayo de 2022

Índice general

I	Introducción	5
1.	Introducción y visión general	7
1.1.	La localización como una necesidad	7
1.2.	Objetivo	12
1.3.	Estructura del documento	13
II	Desarrollo	15
2.	Herramientas utilizadas	17
2.1.	Robotic Operating System	17
2.1.1.	Arquitectura de la comunicación mediante ROS	17
2.1.2.	¿Por qué ROS?	19
2.2.	Mapas de ocupación	20
2.3.	Sensores RGBD	20
2.3.1.	Imagen de profundidad	21
2.4.	Nubes de puntos	22
2.5.	YOLO: You only look once	23
2.5.1.	CUDA: Compute Unified Device Architecture	24
3.	Diseño e implementación	27
3.1.	Vista general del sistema	27
3.2.	Fuente de datos	29
3.3.	De imagen de profundidad a barrido láser	30
3.4.	Generación del mapa de ocupación	32
3.5.	Detección de objetos	33
3.5.1.	Descompresión de la imagen de color	33
3.5.2.	Ejecución de YOLO	34
3.6.	Procesamiento de los objetos detectados	34

Parte I

Introducción

Capítulo 1

Introducción y visión general

En este capítulo se hará una breve introducción para entender el contexto de este trabajo y el porqué de la necesidad del mismo. Se comentará la importancia de la localización para los seres humanos y se hará una introducción sobre los mapas utilizados por los robots para localizarse. Finalmente se expondrá la estructura del documento.

1.1. La localización como una necesidad

Que levante la mano quien no se ha perdido alguna vez por los despachos de alguna de las facultades, sobre todo en esos ambientes prácticamente idénticos entre sí. Todo ser humano alguna vez se ha parado en seco, ha mirado a un lado y a otro con cara de desaprobación y se ha preguntado: *¿dónde estoy?* Estas situaciones, a veces incómodas, se deben esencialmente a que todo ser vivo con capacidad para desplazarse por un entorno necesita localizarse y orientarse si quiere alcanzar un objetivo, más aún si dicho entorno es totalmente desconocido.

A lo largo de los años, los seres humanos hemos utilizado multitud de métodos para la orientación, incluso en las situaciones más desfavorables. Un estudio para la revista *The Royal Society* de la Universidad de Rennes sostiene que los vikingos usaban una variedad de calcita para calcular dónde se encontraba el sol utilizando la polarización de la luz dispersada por las nubes. Esta piedra solar se llama “Espato de Islandia” y les permitía orientarse incluso en fechas donde el sol parecía no querer mostrarse, algo que sucede muy frecuentemente en los mares nórdicos [1]. Hoy en día, con el uso de la tecnología GPS y de aplicaciones que la implementan como Google Maps, hemos cambiado esa piedra solar por un dispositivo electrónico que, no solo nos localiza y orienta en prácticamente cualquier situación, sino que lo hace

con un error casi despreciable. Pese al avance de las tecnología de localización como GPS, existen ciertas situaciones en las que es complicado aprovecharse de ellas, como por ejemplo en los entornos indoor. En estos caso, los satélites no son capaces de comunicarse con nuestros dispositivos, por lo que necesitamos de otro mecanismo para localizarnos.

En la figura 1.1 se muestra el plano de la red de metro de la Comunidad de Madrid. A primera vista, uno se pierde entre tanto entresijo de líneas de colores pero si se establece una estación de origen y otra de destino, este plano facilita, e incluso posibilita, concretar el camino a tomar. Además, en las estaciones suele haber una gran cantidad de señalizaciones que hacen aún más fácil la localización. Entonces, si los humanos, seres inteligentes con percepción visual y capacidad de razonamiento, necesitamos localizarnos para alcanzar un objetivo, ¿porqué no lo iban a necesitar los robots?



Figura 1.1: Porción del plano del metro de Madrid [2]

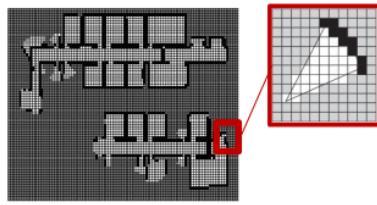
La robótica móvil ha ido evolucionando a un ritmo vertiginoso desde que comenzó a desarrollarse. En los años sesenta se diseñó el apodado como *SHAKY* o “el primer robot inteligente móvil del mundo” según el IEEE (Instituto de Ingenieros Eléctricos y Electrónicos, del inglés *Institute of Electrical and Electronics Engineers*). Este robot era el primero de su generación capaz de navegar en un entorno no controlado, sirviéndose de múltiples dispositivos que le proporcionaban información sobre la distribución de los elementos que le rodeaban, dándole la posibilidad al robot de evitarlos durante el trayecto [3].

Un robot autónomo móvil (ARM) es un robot capaz de navegar a través de un entorno sin supervisión directa por parte de un operador humano ni la necesidad



Figura 1.2: Robot SHAKEY, 1968 [4]

de disponer de un ruta fijada previamente. Para poder llevarlo a cabo, los robots disponen de multitud de sensores que les permiten percibir e interpretar el entorno, dotando al robot de la capacidad de navegar por el mismo evitando obstáculos, tanto fijos como móviles. Para poder realizar una navegación de alto nivel que les permita ir desde un punto origen a otro punto destino, los robots se sirven de mapas. Un mapa es una representación bidimensional del entorno. Existen dos tipos de mapas de aplicación a la robótica: los mapas métricos y los mapas topológicos. Los mapas métricos (figura 1.3, izquierda) dividen el entorno en una rejilla donde cada celda representa una probabilidad de estar ocupada o de no estarlo. Si se discretiza esa probabilidad, el resultado es un mapa de ocupación, (también llamado *floorplan* o “plano de suelo”, en un ámbito menos técnico), si no, se obtiene un mapa probabilístico. El mapa métrico más simple es el formado por un conjunto de referencias visuales o *landmarks* de posición conocida que ayudan al robot a estimar su posición. Los mapas topológicos (figura 1.3, derecha) están basados en grafos y representan la conectividad entre cada uno de sus nodos. Existe una tercer tipo de mapa que es un híbrido entre los comentados y es el más utilizado para los robots: utilizan los grafos para la generación de trayectorias (navegación) y el mapa métrico para la localización.



(a) Ejemplo de mapa métrico: mapa de ocupación



(b) Mapa topológico: líneas del metro de Málaga [5]

Figura 1.3: Ejemplos de tipos de mapas

Los mapas métricos más utilizados en la robótica móvil son los mapas de ocupación o *floorplans*. Son generados en una primera fase de inspección del entorno, empleando los mismos sensores que luego servirán para la navegación autónoma. Tradicionalmente, estos mapas de ocupación son generados empleando un láser 2D o LIDAR (dado su gran alcance y amplio campo de visión), generando por tanto mapas bidimensionales que especifican aquellas zonas del entorno que están libres de obstáculos y por tanto susceptibles de navegación, y aquellas que no lo están. No obstante, dado que un láser 2D no puede distinguir entre los objetos detectados, los mapas de ocupación bidimensionales generados con este sensor contienen, no sólo los elementos estructurales del entorno como paredes, puertas, columnas, etc., si no que incluyen multitud de objetos como mesas y sillas, camas, cajas e incluso personas si estaban presentes en el momento de generar el mapa.

La inclusión de estos elementos no estructurales en el mapa de ocupación puede ocasionar problemas si se pretende utilizar ese mapa durante un largo período de tiempo. El mapa generado no estaría preparado si el entorno que representa sufre algún tipo de modificación de mobiliario, por lo que sería necesario volver a generar un nuevo mapa, con todo lo que eso conlleva. Sin ir más lejos, en la Escuela de Ingenierías Industriales hay muchos elementos que no pertenecen a la estructura del edificio que, si hoy mismo se utilizara un robot equipado con un láser 2D para generar un mapa, ocasionarían los problemas ya comentados. Estos elementos son, por ejemplo, las impresoras que hay en algunas zonas, las máquinas de refrescos y snacks, papeleras, carteles publicitarios, las nuevas mesas instaladas en la zona central, etc. (figura 1.4).

En la figura 1.5 se muestra un ejemplo de un mapa generado de un entorno virtual a través de un láser 2D de 360º. Como se puede observar hay zonas, marcadas con numeración, en las que el mapa no representa fielmente las dimensiones reales de la vivienda. De modo que, si los objetos que no forman parte de la estructura modifican su posición después de la primera fase de análisis, el mapa de ocupación



Figura 1.4: Máquinas de refrescos de la Escuela de Ingenierías Industriales

generado no podría ser utilizado por el robot para la navegación.

Dado que estos métodos de generación de mapas de ocupación mediante láser 2D sufren frente a la variabilidad en la distribución, surge la idea de utilizar otros sensores que permitan discernir entre los distintos elementos que pudieran formar parte del entorno. Los sensores RGBD son cámaras que, además de proporcionar información sobre el color, ofrecen la distancia a la que se encuentra cada píxel captado en la imagen con respecto a la cámara. Por lo tanto, ofrecen información en 3D del dominio, lo que abre más posibilidades a la hora de percibir e interpretar la información captada mediante los sensores.

En este proyecto se pretende analizar una posible solución al problema de fiabilidad en la representación de los mapas de ocupación realizados con sensores LIDAR mediante el uso de sensores RGBD para generar *floorplans* fieles a la estructura invariable del entorno. Entre las utilidades de conseguir un mapa sin los objetos móviles, las más destacadas son:

- La más importante es la posibilidad de tener un mapa que permita al robot localizarse aún con los cambios que puedan surgir a lo largo del tiempo en los elementos que conforman el entorno. Así se evita tener que generar un nuevo mapa cada vez que surja algún cambio significativo.
- En aplicaciones de realidad aumentada para el diseño de interiores esto permitiría obtener las dimensiones reales de la vivienda y así tener una idea en

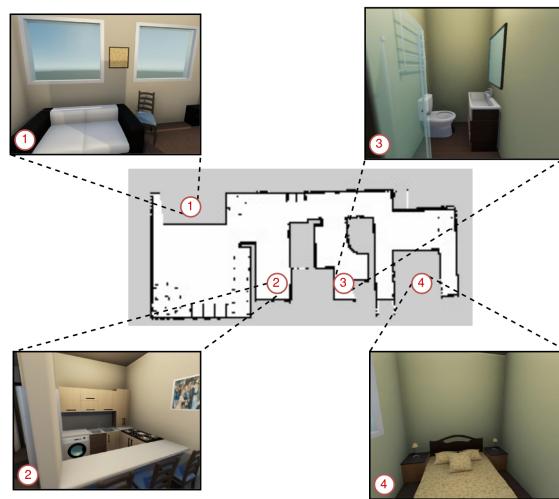


Figura 1.5: Mapa generado con el barrido de un láser 2D

primera persona del cambio de mobiliario o de una nueva organización del mismo.

- Normalmente existe una clara disparidad entre el plano de una vivienda o de un edificio diseñado por el arquitecto y el producto final. Con este mapa generado sería posible comparar y obtener un plano fiel a la realidad y así validar las dimensiones reales de la infraestructura.

1.2. Objetivo

Dado que los robots utilizan los mapas de ocupación o *floorplans* para localizarse y navegar en un entorno, es necesario que estos sean lo más fiel posible a la estructura real del edificio o de la vivienda. Esto resulta imposible si durante la fase de obtención del mapa existen elementos que no pertenecen a la infraestructura, ya que provocan que el mapa resultante no sea fiel las dimensiones reales del entorno.

Debido a la necesidad y utilidad de obtener mapas de ocupación sin elementos cambiantes, el objetivo de este trabajo es utilizar la información proporcionada por sensores RGBD para la generación de *floorplans*, solucionando problemas de variabilidad en el entorno que perjudican seriamente la fiabilidad del mapeado mediante sensores láser.

Se obtendrá información acerca de la profundidad de los píxeles captados y se escogerán aquellos puntos menos restrictivos (más alejados) para generar un láser artificial que permita la creación del mapa de ocupación mediante algoritmos de mapeado.

Paralelamente, se utilizará un modelo entrenado de inteligencia artifical, YOLO, para detectar los elementos de la imagen de color y así llevar un registro de los objetos que se ha encontrado el robot en su trayectoria.

Finalmente, se hará una comparativa de los resultados obtenidos únicamente con el LIDAR con los obtenidos mediante el procesamiento de la información de la cámara. Además se hará un estudio de los objetos detectados durante el mapeado y de la fiabilidad del modelo de detección.

1.3. Estructura del documento

Este documento está dividio en tres[cuatro?] partes. En la primera parte se hará una introducción al contexto del proyecto y se explicará por qué se ha desarrollado.

La segunda parte consta de tres capítulos. En el primero de ellos se expondrán las herramientas utilizadas y se explicarán brevemente cómo se utilizan y porqué se han escogido. En un segundo capítulo, el más importante, se desarrollará el trabajo y se explicará cómo se ha llevado a cabo. En el tercer y último capítulo se expondrán los resultados y se compararán con los ideales, comentando sus diferencias y problemas encontrados.

En la tercera [y última?] parte se mostrarán las conclusiones obtenidas tras la consecución del proyecto y se comentarán algunas posibles líneas de trabajo abiertas de cara al futuro, así como posibles mejorías que se pudieran implementar.

[Cuarta parte. ANEXOS. No sé si me dará tiempo a una cuarta parte de Anexos]

Parte II

Desarrollo

Capítulo 2

Herramientas utilizadas

2.1. Robotic Operating System

Robot Operating System (ROS) es un framework para el desarrollo de software de robots. Se desarrolló originariamente en 2007 por el Laboratorio de Inteligencia Artificial de Stanford para dar soporte a su proyecto STAIR¹ [7]. Es software libre bajo términos de licencia BSD² que, pese a no ser un sistema operativo, provee los servicios estándar de éstos entre los que se incluyen la abstracción del hardware, el control de dispositivos de bajo nivel, la implementación de funcionalidad de uso común, el paso de mensajes entre procesos y el mantenimiento de paquetes. Está basado en una arquitectura de grafos donde el procesamiento toma lugar en los nodos que pueden recibir, mandar y multiplexar mensajes de sensores, control, estados, planificaciones y actuadores, entre otros.

2.1.1. Arquitectura de la comunicación mediante ROS

La unidad básica de una comunicación en ROS es el *nodo*. Un nodo es un programa diseñado en Python o C++ que se comunica con otros nodos principalmente mediante *topics*. Cuando varios nodos se complementan para cumplir una determinada función, se suelen agrupar en paquetes. Un paquete contiene, entre otras cosas, el nodo o los nodos que lo conforman, las dependencias necesarias para la ejecución

¹STAIR (Stanford Artificial Intelligence Robot) es un proyecto llevado a cabo por la Universidad de Stanford para desarrollar robots capaces de navegar en entornos indoor e interactuar con objetos y personas mediante inteligencia artificial. Para su versión 2.0 desarrollaron el framework de ROS [6]

²La licencia BSD es la licencia de software libre otorgada principalmente para los sistemas BSD (Berkeley Software Distribution). Es una licencia permisiva que permite la redistribución libre o privativa [8]

de alguno de sus nodos, información acerca del propio paquete, etc. ROS tiene sus propios paquetes incluidos de forma estándar con la instalación pero además se pueden instalar paquetes de terceros, desarrollados por otros usuarios.

Existe un nodo especial, llamado nodo *master*, que se encarga de registrar las direcciones de cada nodo y qué y dónde publica o se suscribe. Además, se encarga de alojar el servidor de parámetros, accesibles y modificables en cualquier momento. Es necesario lanzar este nodo previamente al lanzamiento de cualquier otro.

Los topics son el método más común de comunicación entre nodos. Está basado en el modelo de publicador/suscriptor. Los nodos pueden publicar (Publisher) o recibir (Subscriber) mensajes de un único tipo, establecidos por el nodo publicador. En un mismo topic pueden publicar y suscribirse tantos nodos se necesiten³.

Otro método de comunicación son los servicios o *services*. Están basados en un modelo de llamada/respuesta. Mientras que los topics permiten a los nodos obtener continuas actualizaciones de datos, los servicios solo ofrecen respuesta tras ser llamados. Permiten la inclusión de parámetros tanto en la solicitud como en la respuesta al servicio. Este método es bloqueante, es decir, el nodo solicitante queda bloqueado al momento de enviar la solicitud hasta que reciba una respuesta, por lo que están pensados para ser utilizados en tareas que no ocupen mucho tiempo.

El otro método de comunicación que existe en ROS son las acciones o *actions*, pensadas para tareas que necesiten una gran cantidad de tiempo para poder ser realizadas. Las acciones funcionan con un modelo cliente/servidor. A diferencia que los servicios, las acciones no son bloqueantes y permiten ser canceladas mientras se están ejecutando.

Como se ha comentado, el método más usual de comunicación es a través de topics, donde los nodos envían y reciben información continuamente, sin peticiones ni esperas. Los servicios y acciones se utilizan en función de los requerimientos del software a desarrollar.

³Realmente, ROS especifica que no existe límite de nodos suscritos a un mismo topic. Sin embargo, se ha demostrado que a partir de 10 nodos, la fiabilidad se ve reducida y la latencia aumentada de forma significativa, pudiendo incluso ocasionar que algunos mensajes no sean recibidos por algunos suscriptores [9]

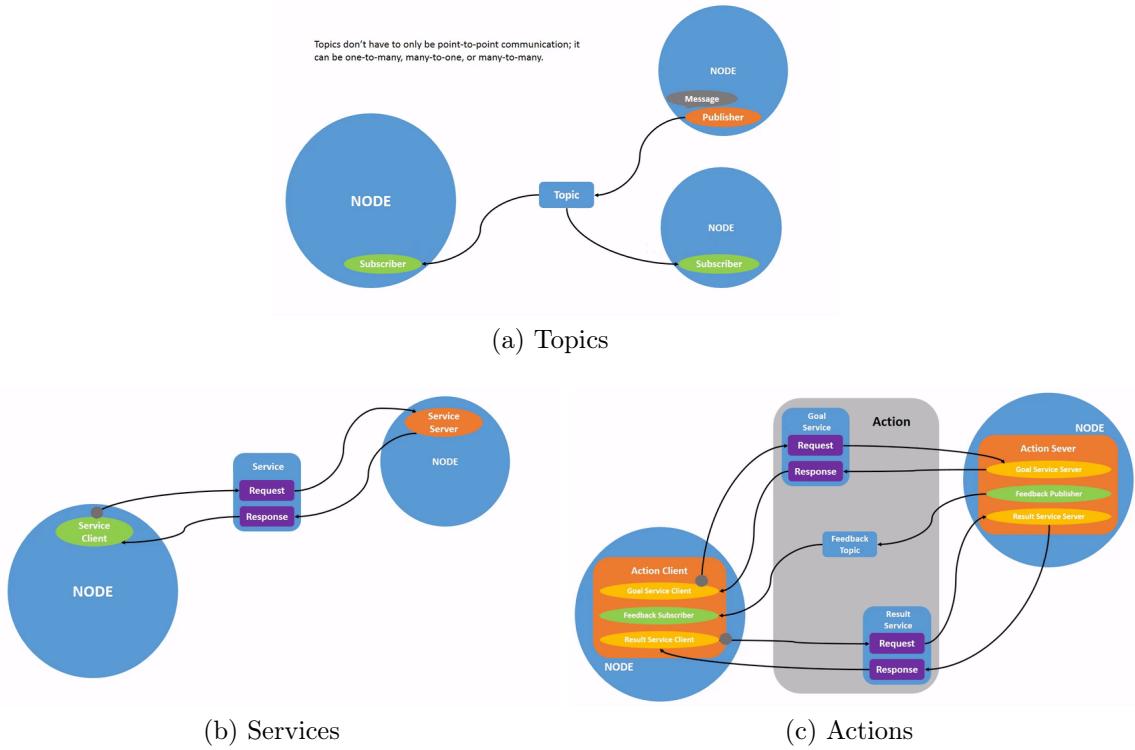


Figura 2.1: Mecanismos de comunicación en ROS

ROS también implementa una herramienta muy útil a la hora de probar los programas que se desarrollan. Un *rosbag* de ROS es una herramienta que permite capturar todos los mensajes que se publican por determinados topics para luego reproducirlos. Esto permite ejecutar el robot una única vez capturando todos los topics y luego reproducir los mensajes de estos topics cuantas veces queramos, permitiendo probar los nodos que se estén desarrollando sin tener que hacerlo directamente sobre el robot. Toda esta información se guarda en un archivo .bag.

2.1.2. ¿Por qué ROS?

La comunidad robótica ha evolucionado considerablemente a lo largo de los años. Pese a este rápido progreso, los robots aún presentan auténticos desafíos para los desarrolladores de software. ROS nace precisamente para facilitar muchas de las dificultades que surgen en la comunicación entre los distintos elementos que conforman un robot.

Muchos de los sistemas robóticos modernos necesitan de un sistema de comunicaciones que sirva de enlace entre los diferentes procesos. Estos procesos suelen

dividirse en varias computadoras, lo que complica aún más la comunicación. Las diferentes vías de comunicación que ROS pone a disposición brindan a los desarrolladores infinidad de posibilidades.

Pero la que es quizá la mayor ventaja de ROS frente a otros frameworks y software especializado en robótica es la gran comunidad que tiene detrás. Existen miles de usuarios que contribuyen continuamente con nuevos paquetes que implementan funcionalidades de todas las ramas de la robótica, lo que abre un abanico de opciones para el desarrollador, reduciendo en gran medida la curva de aprendizaje.

2.2. Mapas de ocupación

Un mapa o rejilla de ocupación es una representación bidimensional del entorno que se almacena en el robot para realizar tareas de navegación. Está formado por celdas que discretizan el entorno y determinan si una porción del espacio está ocupado o no. Generalmente, las celdas ocupadas se representan con el color negro, y las celdas libres con el blanco.

Para generar estos mapas, se utiliza la información obtenida de los diferentes sensores que conforman el robot, como pueden ser infrarrojos, cámaras RGBD, sónares o, los más utilizados, láseres 2D. Independientemente de la fiabilidad del sensor, no es posible establecer con certeza si una celda está ocupada o no. En la generación de estos mapas, típicamente se establecen tres suposiciones: una celda puede estar libre u ocupada, cada celda es independiente a las demás y el entorno es estático. Para estimar las distribuciones de probabilidad de cada celda se utiliza generalmente un Filtro de Ocupación Bayesiano (BOF) [10].

2.3. Sensores RGBD

Las cámaras RGBD son sensores visuales que, además de ofrecer información de color del entorno, proporcionan la distancia a la que se encuentra cada píxel de la imagen. Son cámaras a color comunes a las que se le añade un sensor de profundidad (normalmente, sensores infrarrojos) junto con un procesamiento que les permite percibir e interpretar la distancia a la que se sitúan los diferentes elementos del entorno que estén dentro de su campo de visión.

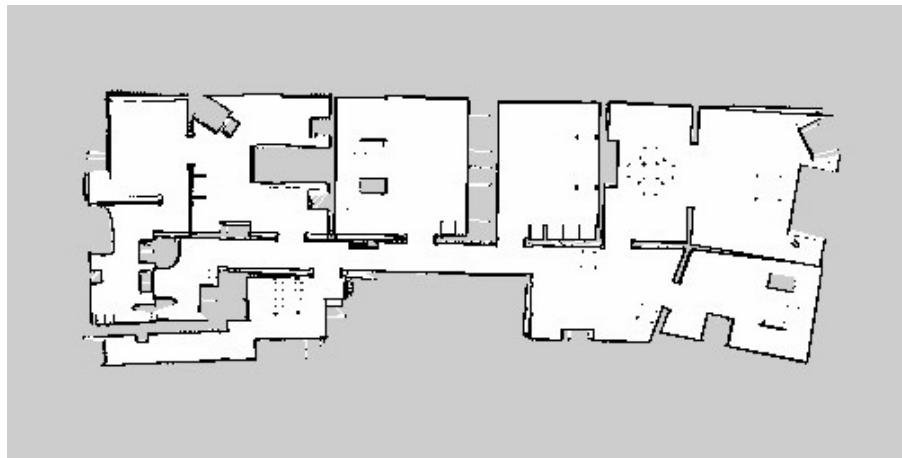


Figura 2.2: Ejemplo de mapa de ocupación



Figura 2.3: Kinect V1

Algunos de los fabricantes más importantes de este tipo de sensores son Microsoft, con la conocida Kinect; ASUS, IFM, StereoLabs, Intel, Orbbec, etc.

2.3.1. Imagen de profundidad

La imagen de profundidad, al igual que cada uno de los canales de color RGB, es una matriz donde cada valor representa un nivel de gris asociado a una distancia y está referido a cada uno de los píxeles que componen la imagen. Las dimensiones de esta imagen vendrán determinadas por la resolución de la cámara, es decir, si por ejemplo la resolución de la cámara es de 640x480 píxeles, la matriz tendrá 480 filas y 640 columnas.

El valor de distancia para cada píxel establece un nivel de gris. Generalmente, los objetos situados más alejados del sensor tendrán valores más claros, mientras que a los más cercanos se le asignan valores más oscuros. Esta información se deco-



Figura 2.4: Ejemplo de imagen de profundidad

difica para obtener así la información absoluta en distancia, típicamente en metros o milímetros, dependerá de la codificación con la que la cámara envía la información.

2.4. Nubes de puntos

Una nube de puntos o pointcloud es un conjunto de vértices en un sistema de coordenadas tridimensional. Estos vértices se identifican habitualmente como coordenadas X, Y y Z y son representaciones de la superficie externa de un objeto. Se crean a partir de escáneres láser tridimensionales o imágenes de profundidad.

Las nubes de puntos tienen infinidad de aplicaciones como la elaboración de modelos 3D en CAD de piezas fabricadas, la inspección de calidad en metrología y muchas otras en el ámbito de la visualización, animación, texturización y aplicaciones de personalización masiva.

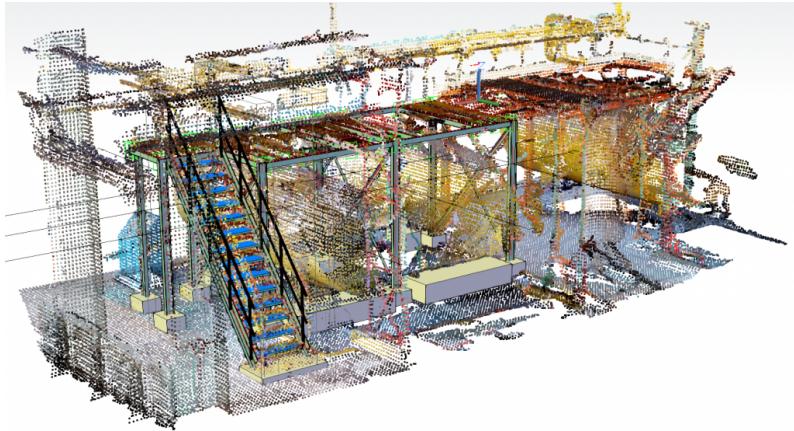


Figura 2.5: Ejemplo de nube de puntos

2.5. YOLO: You only look once

You only look once (YOLO) es un sistema en tiempo real de detección de objetos basado en rede neuronales convolucionales y deep learning. Su principal característica y la que le distingue de otros modelos para la detección de objetos es que solo requiere visualizar una única vez la imagen (haciendo honores a su acrónimo). Esta propiedad le permite ser más rápido que los modelos competidores permitiendo la detección en tiempo real en vídeos de hasta 30 FPS. Según comentan sus creadores, es 1000 veces más rápido que R-CNN y 100 veces más rápido que Fast R-CNN.

El procedimiento llevado a cabo por YOLO es sencillo. Primero divide la imagen en una cuadrícula de $S \times S$ y en cada una de estas celdas precide N posibles bounding boxes y calcula su probabilidad. Tras el cálculo, se eliminan aquellas que no superen un umbral de probabilidad. A las bounding boxes restantes, se les somete a un proceso de supresión de no máximos con el objetivo de eliminar los objetos que fueron detectados por duplicado, obteniéndose el resultado de la figura 2.7.

YOLO es fácilmente implementable en ROS gracias al paquete darknet_ros, que permite utilizarlo tanto en GPU como en CPU. La ventaja de usar la GPU para lanzar el modelo es que es 500 veces más rápido que utilizando la CPU. Para utilizar la GPU es necesario instalar el software CUDA de Nvidia [11].

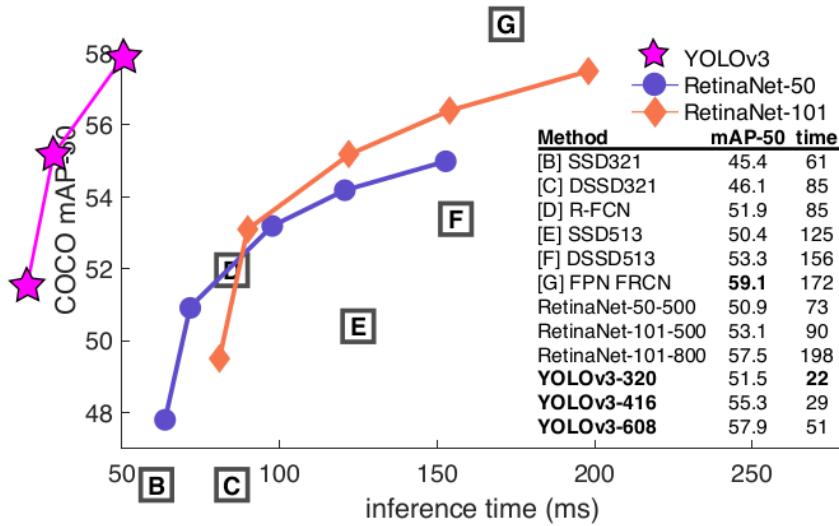


Figura 2.6: Comparación con otros modelos

2.5.1. CUDA: Compute Unified Device Architecture

CUDA es un conjunto de herramientas de desarrollo creadas por Nvidia que permiten a los programadores usar una variación del lenguaje de programación C (CUDA C) para codificar algoritmos GPU de Nvidia. Tiene como objetivo explotar las ventajas de las GPU frente a las CPU de propósito general utilizando el paralelismo que ofrecen sus múltiples núcleos, que permiten lanzar un altísimo número de procesos simultáneos [12].

Las principales ventajas de este sistema de computación son:

- Lecturas dispersas. Se puede consultar cualquier posición en memoria.
- Memoria compartida. CUDA pone a disposición un área de memoria de entre 16KB y 48KB que se compartirá entre hilos del mismo bloque, pudiéndose utilizar como memoria caché.
- Lecturas más rápidas de y hacia la GPU.
- Soporte para enteros y operadores a nivel de bit.

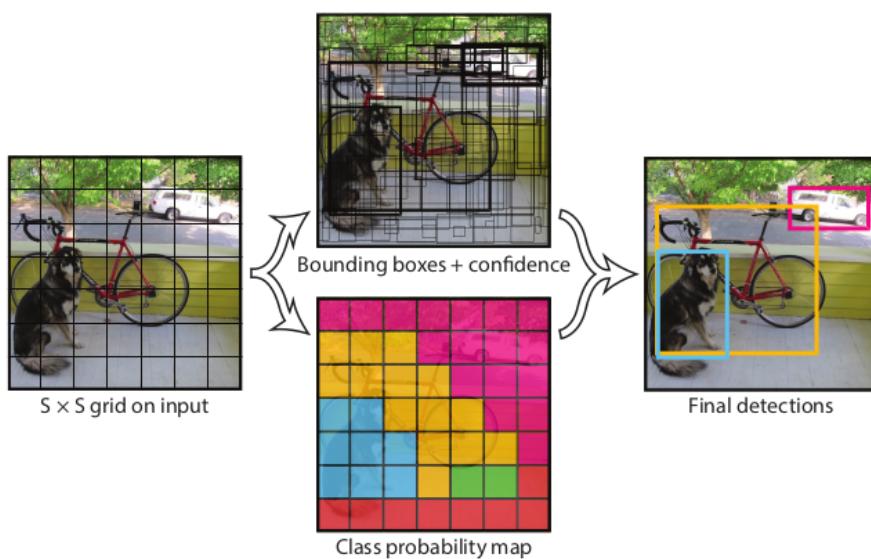


Figura 2.7: Procedimiento de YOLO

Capítulo 3

Diseño e implementación

Para implementar el sistema que nos permita obtener estos mapas de ocupación a partir de la imagen de profundidad generada por una cámara RGB-D, se utilizarán una serie de nodos de ROS. Algunos de estos paquetes y nodos son proporcionados por la comunidad de usuarios de ROS. Otros son diseñados expresamente para este proyecto.

Primeramente se realizará una visión general de todo el conjunto de nodos y paquetes. Se explicará el camino que llevará la información y las formas que tienen los nodos de comunicarse.

Sobre los paquetes de terceros, se hará una breve explicación sobre la función que tienen de fábrica y la funcionalidad que se le han dado en este trabajo, así como el modo de comunicación con todo el árbol de nodos.

3.1. Vista general del sistema

Como se ha comentado, el objetivo es transformar la imagen de profundidad que se obtiene de un sensor RGBD en un láser artificial para generar un mapa de ocupación. En otro proceso, simultáneamente o de forma asíncrona, se generará un archivo de texto o CVS con los elementos detectado por YOLO de la imagen de color para tener un registro y poder analizarlo. En la figura 3.1 se muestra el árbol de nodos y topics de todo el sistema. Los nodos que han sido diseñados para este proyecto se muestran en color azul. Los nodos que han sido obtenidos de otras fuentes pero han necesitado de alguna modificación en su código se muestran en color naranja. Los nodos que no se han modificado se muestran en blanco. Los topics se representan

con una caja de color blanco.

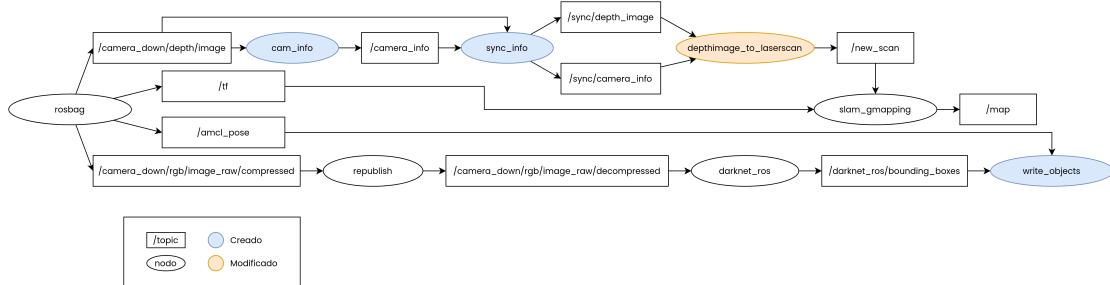


Figura 3.1: Árbol de nodos y topics. Los nodos diseñados para este trabajo son de color azul. Los nodos de terceros cuyo código ha sido modificado son de color naranja. Los nodos descargados que no se han modificado son de color blanco

A continuación se enumeran todos los nodos que se van a utilizar en este proyecto:

- El nodo **cam_info** es un nodo diseñado expresamente para este trabajo. Se encarga de recibir la cabecera de la imagen de profundidad para generar un mensaje con información de la cámara y publicarlo.
- El nodo **sync_info** es otro de los que han sido diseñados para el proyecto. Su función es sincronizar los mensajes de la imagen de profundidad y de la información de la cámara.
- **depthimage_to_laserscan** es un nodo de un paquete homónimo diseñado por Chad Rocke [13] que se encarga, principalmente, de transformar una imagen de profundidad en un barrido láser en función de unos parámetros. Para implementar el nodo en este trabajo modificaremos algunos fragmentos de su código.
- El nodo **slam_gmapping** es un nodo del paquete **gmapping** diseñado por Brian Gerkey [14]. Su función es la de obtener la localización del robot generando el mapa de ocupación.
- **republish** es un nodo que pertenece al paquete **image_transport** diseñado por Patrick Mihelich [15]. Se encarga de descomprimir las imágenes de entrada y publicarlas en formato descomprimido.
- El nodo **darknet_ros** es un nodo de un paquete con el mismo nombre diseñado por Marko Bjelonic [16]. Implementa YOLO en ROS.

- El nodo `write_objects` es un nodo diseñado para este proyecto con el fin de procesar los objetos detectados y escribirlos en un archivo de texto para su posterior análisis.

3.2. Fuente de datos

La mejor manera de comprobar el correcto funcionamiento de todo el conjunto de nodos sería hacerlo directamente sobre el robot, con información real de sensores y del entorno. Sin embargo, debido a la dificultad que supondría utilizar este sistema sobre el robot, se decidió utilizar la herramienta que ROS pone a disposición precisamente para lidiar con este inconveniente: los rosbags.

Para este proyecto se va a utilizar un rosbag generado por uno de los robots del laboratorio de Automática de la Escuela Técnica Superior de Informática. Esta grabación recoge la información de la pose del robot (posición X e Y y el ángulo de orientación), los datos recogidos por una cámara RGB-D (color y profundidad), el barrido de un láser y otra información que típicamente se utiliza en los robots como el árbol de transformadas.

Sin embargo, pese a que este rosbag ha sido el utilizado para comprobar el funcionamiento del algoritmo diseñado, ha sido necesario hacer algunas modificaciones para que se pueda procesar la información correctamente.

Normalmente, cuando se utiliza una cámara RGBD en ROS, el paquete encargado de controlar la cámara envía tres tipos de mensajes pertenecientes al paquete `sensor_msgs`. Estos mensajes son:

- Imagen a color. Información de color de la imagen capturada con un mensaje de tipo `Image`. La mayoría de cámaras ofrece además la imagen rectificada, sin distorsiones.
- Imagen de profundidad. Información sobre la profundidad de la imagen capturada con un mensaje de tipo `Image`. Al igual que con la de color, también se ofrece la imagen rectificada.
- Parámetros de la cámara. Información sobre las propiedades de la cámara con un mensaje de tipo `CameraInfo`. Contiene datos sobre las propiedades de la imagen (ancho y alto) y parámetros de calibración de la cámara.

Debido a algunos problemas que se tuvieron durante la captura de los datos, este último topic no se ofrece en el rosbag, por lo que es necesario generararlo externo

al rosbag. Para ello, se ha diseñado un nodo llamado `cam_info` que se encarga de generar este mensaje que más tarde será utilizado por otros nodos.

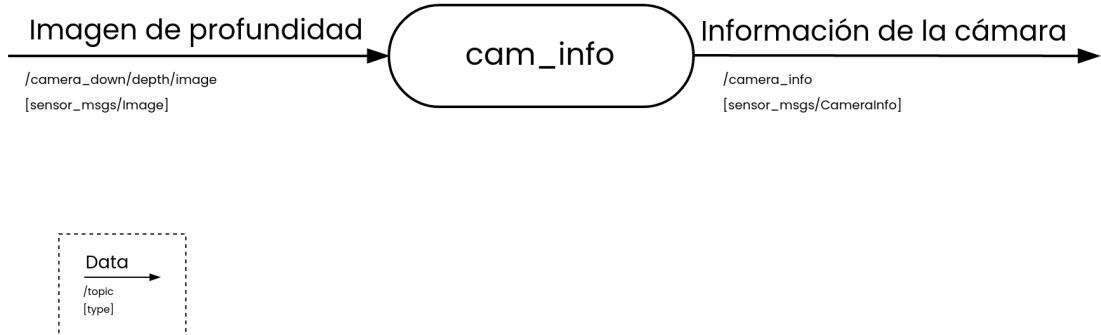


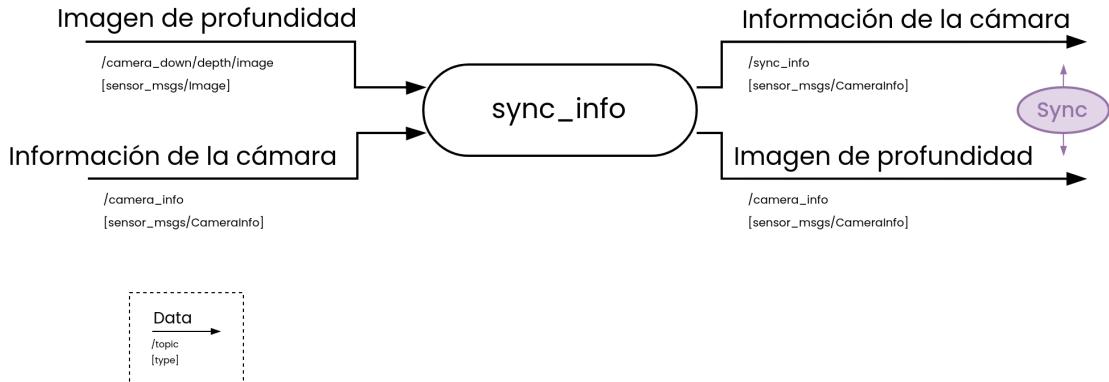
Figura 3.2: Diagrama del nodo `cam_info`

Este nodo, programado en C++, recibe el mensaje de la imagen de profundidad ofrecido por el rosbag, copia su encabezado, crea el mensaje de tipo `sensor_msgs/CameraInfo` con los parámetros de la cámara y lo envía por el topic `/camera_info`.

Finalmente, una vez generado el mensaje con la información de los parámetros de la cámara, es necesario sincronizar estos topics para que puedan ser utilizados por otros nodos. Para ello se ha diseñado un nodo en Python llamado `sync_info` que se sirve del paquete `message_filter` para realizar este proceso. Se ha programado en Python porque resulta más fácil utilizar esta librería en este lenguaje que en C++.

3.3. De imagen de profundidad a barrido láser

Una vez se tiene toda la información ofrecida por el robot y la cámara correctamente sincronizada y configurada, es momento de realizar el procedimiento principal de este proyecto: transformar la imagen de profundidad obtenida mediante la cámara RGB-D en un barrido láser equivalente, cogiendo la distancias más alejadas captadas en la imagen. De esta forma, si tenemos varios objetos delante de una pared, con el algoritmo diseñado no se tendrán en cuenta estos objetos y se seleccionarán siempre las distancias menos restrictivas para generar el láser. Es necesario generar un láser artificial porque los algoritmos que se utilizan para generar el mapa de ocupación utilizan este tipo de datos.

Figura 3.3: Diagrama del nodo `sync_info`

El algoritmo utilizado está basado en el paquete de ROS `depthimage_to_laserscan`. Este paquete se encarga de transformar una imagen de profundidad en un barrido láser a partir de los parámetros de la cámara y de una serie de parámetros de entrada configurables. Estos parámetros son:

- **scan_height**. Establece la cantidad de filas que se quieren procesar para generar el láser.
- **scan_time**. Establece el tiempo de actualización entre escaneos. Por defecto está a 0.033 (30 FPS).
- **range_min**. Rango de distancia mínima. Valores medidos menores que este valor se tomarán como -Inf. Por defecto está a 0.45 (metros).
- **range_max**. Rango de distancia máxima. Valores medidos mayores que este valor se tomarán como +Inf. Por defecto está a 10 (metros).
- **output_frame_id**. Establece el id del eje de coordenadas (frame) de salida. Se indica el id del frame del láser.

Este nodo primeramente recibe la imagen de profundidad y la información de la cámara. Se recuerda que estos mensajes los debe recibir a la vez y por eso requería de la sincronización. Una vez recibidos estos mensajes, evalúa la codificación de la imagen. Si la codificación es correcta (admite codificación 16UC1 o 32FC1)¹, proce-

¹El paquete `sensor_msgs` provee de multitud de codificaciones para las imágenes. Este es un parámetro de la imagen y viene dado por una cadena de caracteres [17]. El paquete `depthimage_to_laserscan` solo admite las dos codificaciones comentadas

de a convertir la imagen de profundidad a un mensaje de tipo LaserScan.

Un mensaje de tipo LaserScan se forma a partir de las siguientes características:

- **header**. Encabezado del mensaje. Propiedad que tienen todos los mensajes.
- **angle_min**. Ángulo mínimo que abarca el láser en radianes.
- **angle_max**. Ángulo máximo en radianes.
- **angle_increment**. Ángulo entre medidas en radianes.
- **time_increment**. Tiempo entre medidas en segundos.
- **scan_time**. Tiempo entre escaneos en segundos.
- **range_min**. Rango mínimo del sensor en metros.
- **range_max**. rango máximo del sensor en metros.
- **ranges**. Vector con las medidas tomadas en metros.
- **intensities**. Vector con las intensidades de las medidas. Lo tomaremos como un array vacío porque no es necesario.

El algoritmo se encarga de analizar tantas filas como se establezcan en el parámetro **scan_height** del nodo **depthimage_to_laserscan** y transformar la distancia dada por la imagen de profundidad en una distancia real al sensor. Una vez transformada, se compara con la distancia de las otras filas medidas en la misma columna y, si la distancia es mayor, se añade al vector ranges. Una vez completado el análisis de todas las filas de la imagen, se publica el mensaje generado de tipo **sensor_msgs/LaserScan** en el topic **/new_scan**.

3.4. Generación del mapa de ocupación

Una vez obtenido el láser artificial a partir de la imagen de profundidad, es momento de generar el mapa o rejilla de ocupación. Para este procedimiento se utilizará el paquete **gmapping**.

Este paquete ofrece un sistema de Localización y Mapeado Simultáneo o SLAM (del inglés, *Simultaneous Location and Mapping*) a través de un nodo llamado **slam_gmapping**. Este nodo recibe el láser artificial (por el topic **/new_scan**) y el

árbol de transformadas (por el topic `/tf`).

Al lanzar este nodo, es posible establecer multitud de parámetros que determinan el modo de operación. Entre otros, los parámetros que se tendrán en cuenta para ajustar el resultado acorde al objetivo son:

- **linearUpdate**. El nodo procesará un escaneo cada vez que se alcance la distancia lineal determinada por este parámetro. Su valor por defecto es 1,0. El valor óptimo encontrado tras múltiples pruebas es 0,3.
- **angularUpdate**. Determina el incremento de ángulo por el cuál el robot procesará otro escaneo. Su valor por defecto es 0,5. El valor óptimo es 0,7.
- **temporalUpdate**. Establece cada cuánto se procesa un nuevo escaneo. Su valor por defecto y óptimo es 3,0.

Como se comenta en el listado de parámetros, tienen unos valores por defecto que se han ido modificando para obtener un mapa con más fiable. Esto se comentará más en profundidad en la sección de resultados.

3.5. Detección de objetos

Paralelamente, se ha diseñado un sistema capaz de interpretar la imagen a color proporcionada por el sensor RGBD para detectar los objetos y tener un registro de los mismos. Este sistema, en principio, se puede lanzar simultáneamente junto con el sistema de generación de mapas de ocupación, sin embargo, debido a la potencia necesaria para ejecutar la red neuronal por GPU, resulta tedioso e incluso problemático con algunos ordenadores. Es por ello que se ha diseñado para que pueda ser lanzado de forma asíncrona con respecto al otro sistema.

3.5.1. Descompresión de la imagen de color

En muchas ocasiones, los paquetes de ROS que sirven para utilizar determinadas marcas de sensores RGBD proporcionan las imágenes en un formato comprimido para ahorrar espacio. Con el rosbag que se ha utilizado como fuente de datos ocurre lo mismo, por lo que es necesario un proceso de descompresión.

Para este procedimiento, se ha utilizado un nodo llamado `republish` del paquete `image_transport`. Este nodo se recibe un mensaje en formato `sensor_msgs/-CompressedImage` y lo convierte en un mensaje de tipo `sensor_msgs/Image` para

publicarlo.

3.5.2. Ejecución de YOLO

Una vez la imagen está descomprimida, ya puede ser recibida por la red neuronal. Este nodo recibe la imagen de color en formato `sensor_msgs/Image` y publica en tres topics:

- `object_detector`. Mensaje de tipo `std_msgs/Int8` que indica el número de objetos detectados.
- `bounding_boxes`. Mensaje de tipo `darknet_ros_msgs/BoundingBoxes` que representa un array que proporciona información sobre la posición y el tamaño de los bounding boxes en píxeles.
- `detection_image`. Mensaje de tipo `sensor_msgs/Image` que ofrece las bounding boxes detectadas sobre la imagen a color que ha sido procesada.

Las bounding boxes son unas “cajas” que señalan al elemento detectado. Un mensaje de tipo `darknet_ros_msgs/BoundingBoxes` es un array de elementos con formato `darknet_ros_msgs/BoundingBox`.

Además, este nodo ofrece una salida en la terminal indicando los frames por segundo (FPS) a los que se está ejecutando y un listado de los objetos encontrados junto con sus probabilidades, como se muestra en la figura 3.4. YOLO tiene un umbral de probabilidad establecido por defecto en 0,7 para aceptar una detección como válida o no.

[Cabe mencionar que para utilizar YOLO con CUDA correctamente ha sido necesario un largo y tedioso proceso de configuración que se explicará en el Anexo X.]

3.6. Procesamiento de los objetos detectados

Como se ha comentado, a partir del nodo que implementa YOLO obtenemos un array de bounding boxes con información sobre la posición y el tamaño en píxeles. Para el desarrollo de este proyecto se ha visto conveniente procesar esa información para realizar un posterior estudio de los objetos que han sido detectados y analizar la fiabilidad de la red neuronal así como la calidad de los datos proporcionado por

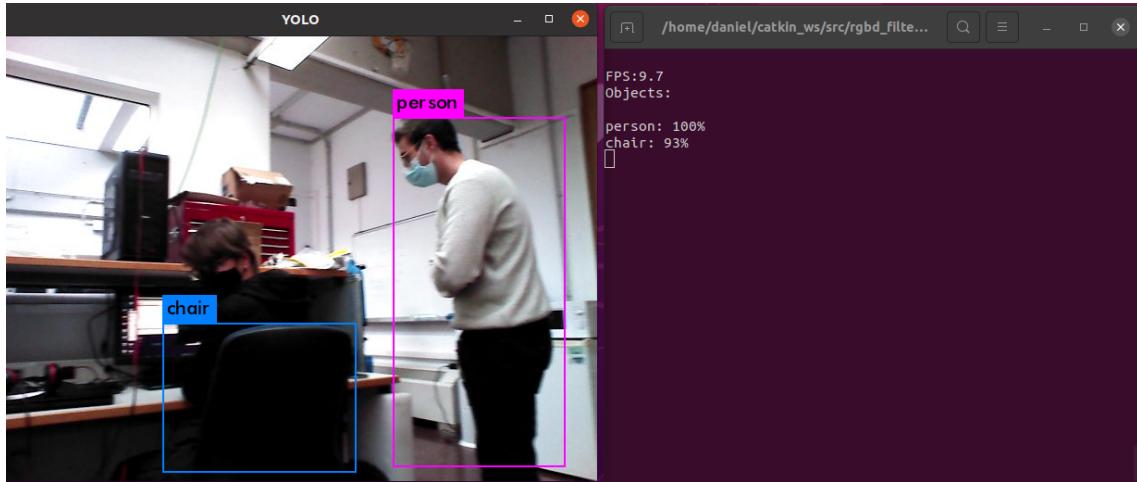


Figura 3.4: Salida en la terminal del nodo `darknet_ros`

el rosbag.

Para este procesamiento se ha diseñado un nodo llamado `write_objects` que se encarga de recibir los mensajes que se publican en `/darknet_ros/bounding_boxes` y en `/amcl_pose` para generar un archivo CSV donde cada fila representa la información con el formato mostrado en el cuadro 3.1. Cada columna utilizará el delimitador de punto y coma (“ ; ”).

Objetos	Probs	Posición	Orientación	Tiempo
obj1:obj2:...:objN	prob1:prob2:...:probN	pX:pY:pZ	oW:oX:oY:oZ	(seg)

Cuadro 3.1: Formato del archivo CSV generado

Toda esta información es obtenida a partir del formato de los mensajes de las bounding boxes y de la pose. Para las bounding boxes es necesario acceder a cada una de ellas y obtener sus propiedades `Class` y `probability`. Para la pose es necesario acceder a su propiedades `pose.pose.position` y `pose.pose.orientation`.

El formato del mensaje de una bounding box tiene las siguientes propiedades:

- **probability**. Probabilidad de que el elemento detectado sea realmente de la clase que se le atribuye.
- **xmin**. Coordenada X mínima de la bounding box en píxeles.

- **ymin**. Coordenada Y mínima de la bounding box en píxeles.
- **xmax**. Coordenada X máxima de la bounding box en píxeles.
- **ymax**. Coordenada Y máxima de la bounding box en píxeles.
- **id**. Identificador único para cada bounding box.
- **Class**. Clase o etiqueta del elemento detectado.

Para la generación del archivo CSV se ha hecho uso de la librería

Bibliografía

- [1] Teresa Guerrero. *El 'GPS' de los vikingos*. Noviembre de 2011. URL: <https://www.elmundo.es/elmundo/2011/11/07/ciencia/1320663197.html>. (acc: 25.05.2022).
- [2] MetroMadrid. *Metro de Madrid*. URL: <https://www.metromadrid.es/es/viaja-en-metro/plano-de-metro-de-madrid>. (acc: 25.05.2022).
- [3] Cristina Sánchez. *Shakey, El Primer Robot Inteligente de la Historia y el abuelo del Coche Autónomo*. Abril de 2017. URL: https://www.eldiario.es/hojaderouter/tecnologia/shakey-robot-inteligencia-artificial-coche-autonomo_1_3466717.html. (acc: 25.05.2022).
- [4] David Cassel. *Remembering shakey, the First Intelligent Robot*. Diciembre de 2021. URL: <https://thenewstack.io/remembering-shakey-first-intelligent-robot/>. (acc: 25.05.2022).
- [5] Wikipedia. *Metro de Málaga*. URL: https://es.wikipedia.org/wiki/Metro_de_M%C3%A1laga. (acc: 25.05.2022).
- [6] Universidad de Stanford. *STAIR*. URL: <http://stair.stanford.edu/index.php>. (acc: 25.05.2022).
- [7] Morgan Quigley, Eric Berger y Andrew Y. Ng. “STAIR: hardware and software architecture”. En: (2007).
- [8] Wikipedia. *Licencia permisiva*. Marzo de 2022. URL: https://es.wikipedia.org/wiki/Licencia_permisiva. (acc: 25.05.2022).
- [9] Github. *Maximum subscribers of a topic*. URL: https://github.com/ros2/rmw%5C_fastrtps/issues/249. (acc: 25.05.2022).
- [10] Chiara Fulgenzi, Anne Spalanzani y Christian Laugier. “Dynamic Obstacle Avoidance in uncertain environment combining PVOs and Occupancy Grid”. En: *Proceedings 2007 IEEE International Conference on Robotics and Automation*. 2007, págs. 1610-1616. DOI: [10.1109/ROBOT.2007.363554](https://doi.org/10.1109/ROBOT.2007.363554).
- [11] Joseph Redmon y col. *You Only Look Once: Unified, Real-Time Object Detection*. 2015. eprint: [arXiv:1506.02640](https://arxiv.org/abs/1506.02640).

- [12] Wikipedia. *Compute Unified Device Architecture*. URL: <https://es.wikipedia.org/wiki/CUDA>. (acc: 26.05.2022).
- [13] Chad Rockey. *ROS WIKI: depthimage to laserscan*. URL: http://wiki.ros.org/depthimage_to_laserscan. (acc: 26.05.2022).
- [14] Brian Gerkey. *ROS WIKI: gmapping*. URL: <http://wiki.ros.org/gmapping?distro=noetic>. (acc: 26.05.2022).
- [15] Patrick Mihelich. *ROS WIKI: image transport*. URL: http://wiki.ros.org/image_transport. (acc: 26.05.2022).
- [16] Marko Bjelonic. *ROS WIKI: darknet ros*. URL: http://wiki.ros.org/darknet_ros. (acc: 26.05.2022).
- [17] ROS. *image-encodings.h File Reference*. URL: http://docs.ros.org/en/jade/api/sensor_msgs/html/image__encodings_8h.html. (acc: 26.05.2022).