

Título del proyecto

Daniel Gómez

23 de mayo de 2022



# Índice general

<b>I</b>	<b>Introducción</b>	<b>5</b>
1.	<b>Introducción y visión general</b>	<b>7</b>
1.1.	Motivación y justificación . . . . .	7
1.2.	Objetivo . . . . .	9
<b>II</b>	<b>Desarrollo</b>	<b>11</b>
2.	<b>Herramientas utilizadas</b>	<b>13</b>
2.1.	Robotic Operating System . . . . .	13
2.1.1.	Arquitectura de la comunicación mediante ROS . . . . .	14
2.1.2.	¿Por qué ROS? . . . . .	15
2.2.	Mapas de ocupación . . . . .	15
2.3.	Sensores RGBD . . . . .	16
2.3.1.	Imagen de profundidad . . . . .	17
2.4.	Nubes de puntos . . . . .	17
2.5.	YOLO: You only look once . . . . .	18
2.5.1.	Compute Unified Device Architecture . . . . .	19
3.	<b>Diseño e implementación</b>	<b>21</b>
3.1.	Vista general del sistema . . . . .	21
3.2.	Fuente de datos . . . . .	23
3.2.1.	Adaptación del rosbag . . . . .	23
3.2.2.	Otras fuentes tomadas en cuenta . . . . .	25
3.3.	De imagen de profundidad a barrido láser . . . . .	25
3.4.	Generación del mapa de ocupación . . . . .	27
3.5.	Detección de objetos . . . . .	28
3.5.1.	Descompresión de la imagen de color . . . . .	28
3.5.2.	Ejecución de YOLO . . . . .	28
3.6.	Procesamiento de los objetos detectados . . . . .	29



# Parte I

## Introducción



# Capítulo 1

## Introducción y visión general

### 1.1. Motivación y justificación

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 Ingeniería Eléctrica y Electrónica). 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.

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 de disponer de una ruta fijada previamente. Para poder realizarlo, los robots disponen de infinidad 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 de ocupación que 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 2D 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 2D 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.

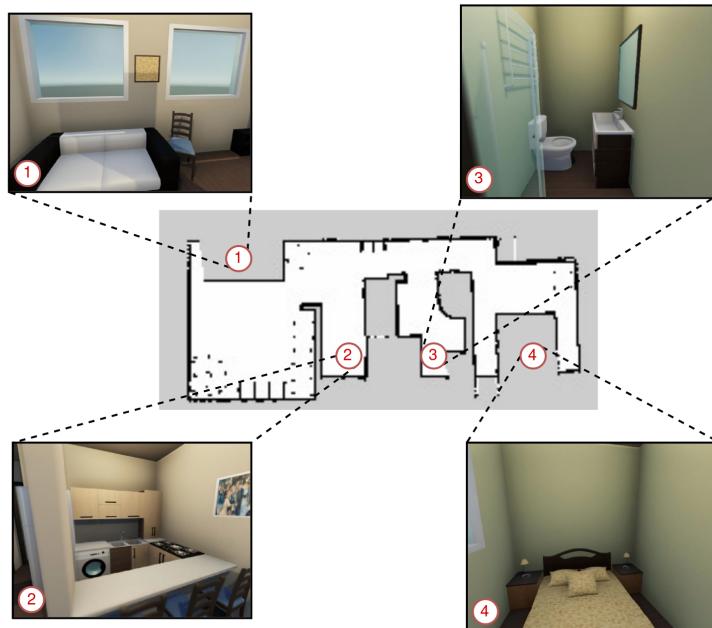


Figura 1.1: Errores típicos con el barrido de un láser 2D

En la figura X 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 la estructura del entorno. De modo que, si los objetos que no forman parte de la estructura del entorno modifican su posición después de la primera fase de análisis, el mapa de ocupación 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 RGB-D 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 RGB-D para generar mapas de ocupación fieles a la estructura invariable del entorno.

## 1.2. Objetivo

El objetivo de este trabajo es utilizar la información proporcionada por sensores RGB-D y utilizarla para la generación 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.



# Parte II

## Desarrollo



# Capítulo 2

## Herramientas utilizadas

### 2.1. Robotic Operating System

Robot Operating System (ROS) es una colección de frameworks 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 STAIR<sup>1</sup>. Es software libre bajo términos de licencia BSD<sup>2</sup> 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.



Figura 2.1: Propiedades

<sup>1</sup>STAIR (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. Fuente: <http://stair.stanford.edu/index.php>

<sup>2</sup>La 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. Fuente: [https://es.wikipedia.org/wiki/Licencia\\_permisiva](https://es.wikipedia.org/wiki/Licencia_permisiva)

### 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 escrito en Python o C++ que se enlaza con otros nodos mediante topics, formando una red de nodos. 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 de alguno de sus nodos, información acerca del propio paquete, etc. ROS tiene sus propios paquetes incluidos ya con la instalación pero además se pueden instalar paquetes desarrollados por otros usuarios.

Existe un nodo especial, llamado nodo master, que es el encargado de dibujar todo el árbol de nodos. Es decir, es el encargado de registrar los nodos, realizar las conexiones entre los publicadores y suscriptores y asignarle el canal de comunicación a través de los topics y de marcar el tipo de mensaje dictaminado por el publicador. Además, se encarga de establecer 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 necesiten<sup>3</sup>.

Otro método de comunicación son los servicios o services. Permiten la comunicación entre nodos mediante el modelo de cliente/servidor. 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.

El otro método de comunicación que existe en ROS son las acciones o actions. Las acciones funcionan de la misma manera que los servicios, solo que en este caso este modelo no bloquea al nodo que solicita información.

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

---

<sup>3</sup>Realmente, 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. Fuente: [https://github.com/ros2/rmw\\_fastrtps/issues/249](https://github.com/ros2/rmw_fastrtps/issues/249)

software a desarrollar.

### 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 binario de estados estáticos de Bayes, basados en probabilidad condicionada.

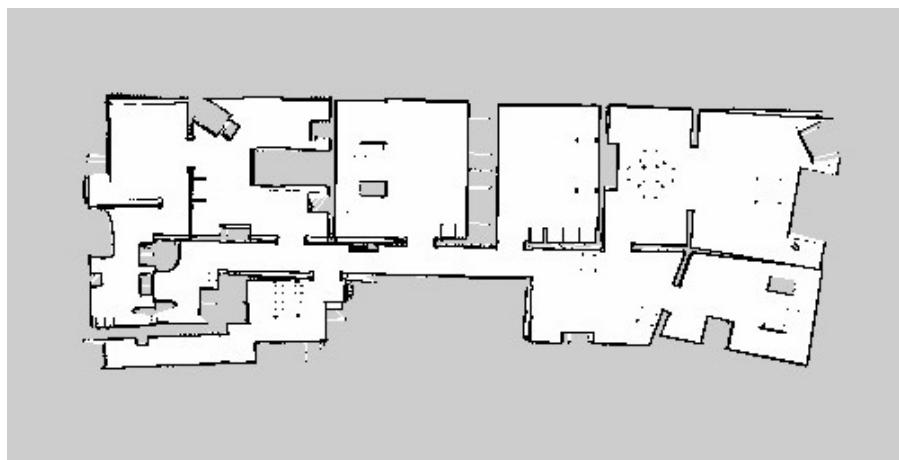


Figura 2.2: Ejemplo de mapa de ocupación

### 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 2.3: 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.



Figura 2.4: Ejemplo de imagen de profundidad

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 decodifica 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.

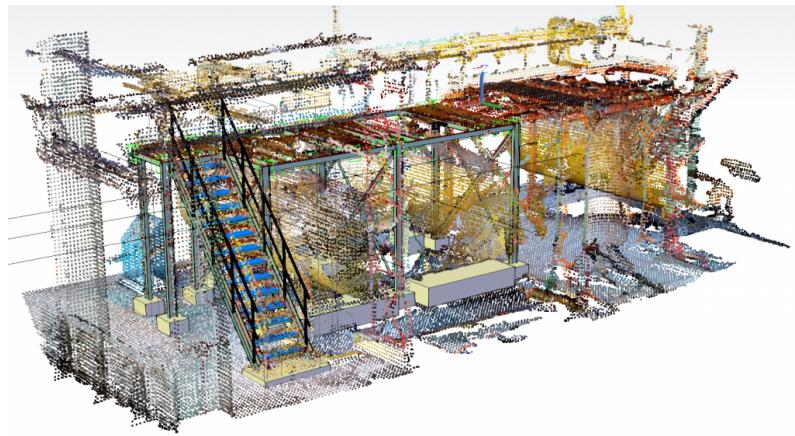


Figura 2.5: Ejemplo de nube de puntos

## 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.

En la figura 2.5 se muestra una gráfica de la comparación de YOLOv3 con respecto a otros modelos. El tiempo de inferencia es mucho menor que el de los otros modelos y la métrica COCO mAP-50 es bastante decente, sobre todo teniendo en cuenta el bajo tiempo de inferencia.

El procedimiento de YOLO es sencillo. Primero divide la imagen en una cuadrícula de  $S \times S$  y en cada una de estas celdas predice N posibles bounding boxes

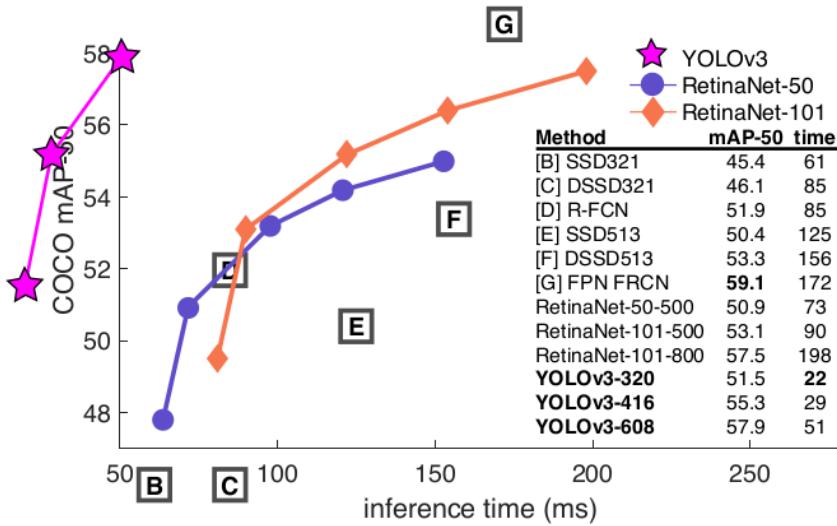


Figura 2.6: Comparación con otros modelos

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 2.5.

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.

(Fuente: <https://pjreddie.com/darknet/yolo/> además tiene un PAPER)

### 2.5.1. 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.

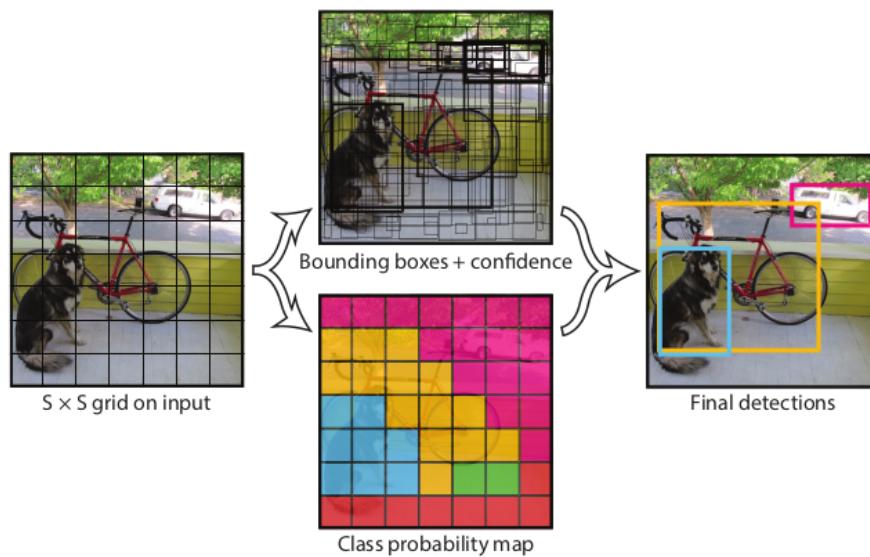


Figura 2.7: Procedimiento de YOLO

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.

Fuente: <https://es.wikipedia.org/wiki/CUDA>

# 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 TXT con los elementos detectados por YOLO de la imagen de color para tener un registro y poder analizarlo. En la figura X se muestra el árbol de nodos y topics de todo el sistema. Los nodos que han sido creados 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 han sido modificados se muestran en blanco. Los topics se representan con una caja de color blanco.

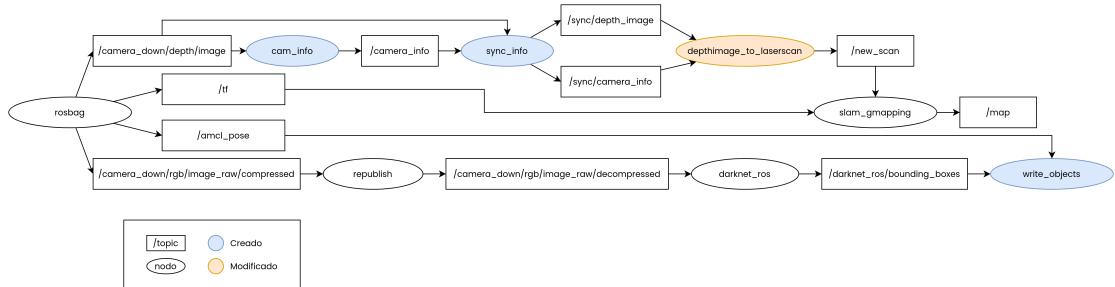


Figura 3.1: Árbol de nodos y topics

Primeramente, se obtiene la información del robot mediante la reproducción de un rosbag (más información en la sección 3.2). De este rosbag se obtienen mensajes en cuatro topics (entre otros):

- `/camera_down/depth/image`. Imagen de profundidad de la cámara.
- `/tf`. Árbol de transformadas.
- `/amcl_pose`. Pose del robot.
- `/camera_down/rgb/image_raw/compressed`. Imagen de color comprimida de la cámara.

El primero de los topics contiene mensajes de tipo `sensor_msgs/Image` con información de la imagen de profundidad captada por la cámara. A partir de este mensaje generamos el mensaje de tipo `sensor_msgs/CameraInfo` con información de la cámara mediante el nodo `cam_info`. Luego se sincronizan estos dos mensajes con el nodo `sync_info` y se envían por los topics `/sync/depth_image` y `/sync/camera_info`. Estos mensajes los recibe el nodo `depthimage_to_laserscan` y los procesa para generar el mensaje de tipo `sensor_msgs/LaserScan` y enviarlos por el topic `/new_scan`. Estos mensajes, juntos con los de tipo `tf2_msgs/TFMessage` del topic `/tf`, los recibe el nodo `slam_gmapping` y los transforma en un mapa o rejilla de ocupación que envía por el topic `/map` con el formato `nav_msgs/OccupancyGrid`.

Con el topic `/camera_down/rgb/image_raw/compressed` se obtienen mensajes de tipo `sensor_msgs/CompressedImage`. Con el nodo `republish` se descomprimen y se envían por el topic `/camera_down/rgb/image_raw/decompressed`. Estos mensajes los recibe la red neuronal mediante el nodo `darknet_ros` y envía mensajes de tipo

`darknet_ros_msgs/BoundingBoxes` por el topic `/darknet_ros/bounding_boxes`. El nodo `write_objects` procesa estos mensajes y los de la pose del robot de tipo `sensor_msgs/PoseWithCovarianceStamped` del topic `/amcl_pose` y les da formato de tabla para pasarlo a un archivo TXT.

## 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 optar por otras vías.

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.

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.

### 3.2.1. Adaptación del rosbag

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 RGB-D 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.

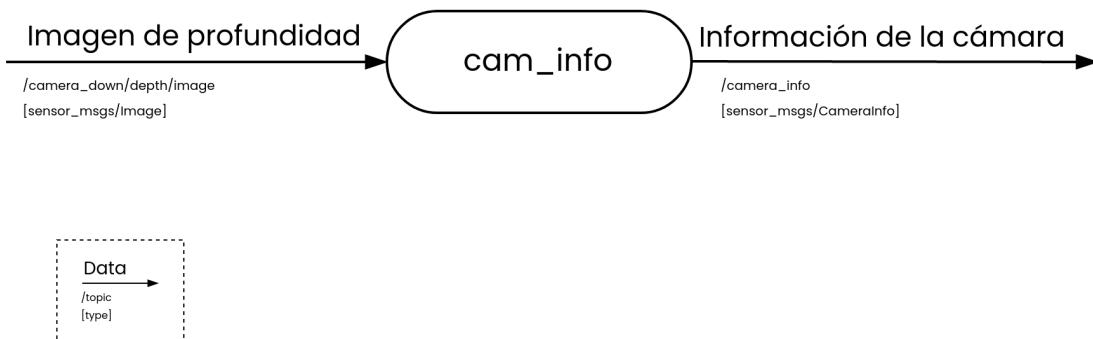
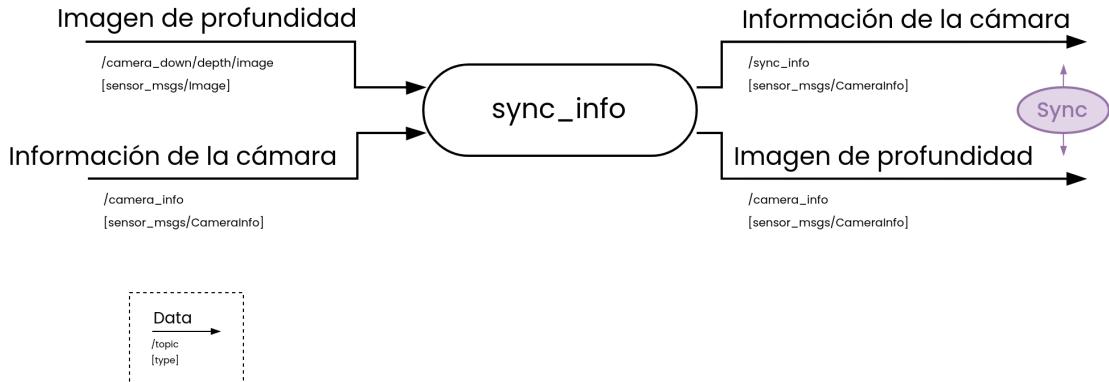


Figura 3.2: Diagrama del nodo `cam_info`

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 CameraInfo con los parámetros de la cámara y lo envía por el topic `/camera_info`. Estos parámetros han sido dados por la persona que generó el rosbag. [\*Indicar los parámetros utilizados].

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++.

Con la adición de estos nodos ya se tendría la información correcta para utilizar los otros paquetes sin ningún problema.

Figura 3.3: Diagrama del nodo `sync_info`

### 3.2.2. Otras fuentes tomadas en cuenta

Inicialmente, se propuso trabajar con un rosbag obtenido a partir de la ejecución de un robot virtual en un entorno simulado en Unity. Este entorno estaba ambientando en una vivienda común, con mobiliario típico y sin demasiados objetos en escena. Esto probablemente hubiera dado un mejor resultado que el rosbag que finalmente se utilizó ya que este último fue ejecutado en un laboratorio con muchos elementos que dificultan la labor del algoritmo (ventanas, entrantes y salientes, gran cantidad de objetos).

Sin embargo, la decodificación de la imagen de profundidad ocasionó muchos problemas que se escapaban del ámbito de este proyecto por lo que se decidió tomar otra vía.

## 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.

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.

[Añadir imagen de comparación del laser]

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[\*Explicar los de los encodings]), procede 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`). Además tiene varios parámetros.

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 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 señalizan 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. 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, de 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 de la fiabilidad de la red neuronal así como la calidad de los datos proporcionado por 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