

Resumen

En la actualidad, los robots móviles utilizan mapas de ocupación para localizarse y poder navegar de forma autónoma. Estos mapas son, generalmente, generados a partir de barridos de un láser 2D que detecta los obstáculos presentes en el entorno y sus distancias al robot, permitiendo de esta manera caracterizar áreas del entorno de trabajo como ocupadas o libres. Una de las principales limitaciones de este procedimiento es la imposibilidad de distinguir un objeto estático, como el caso de una pared, de uno móvil, como el caso de una silla o un sofá, siendo ambos incluidos indistintamente en el mapa de ocupación. Cuando los objetos móviles del entorno cambian su posición, el resultado es que el mapa de ocupación originalmente creado deja de ser representativo y, por tanto, pierde su utilidad para el robot, siendo necesario regenerar dicho mapa para representar la configuración actual del entorno. El objetivo de este trabajo es aprovechar la información contenida en una secuencia de imágenes y sus correspondientes datos de profundidad proporcionados por una cámara RGBD, para generar un mapa de ocupación donde únicamente se representen los elementos estructurales del entorno. Asumiendo que dicha cámara observa en cada momento parte de las paredes que componen el entorno de trabajo, junto con una serie de objetos móviles, en este trabajo se propone un procesado de los datos de profundidad que permite eliminar los objetos que se encuentren entre la pared y el robot, no siendo estos incluidos en el mapa de ocupación. Paralelamente, se ha utilizado la imagen de color de la misma cámara para realizar un proceso de detección de objetos mediante una red neuronal convolucional ya entrenada. Los datos obtenidos tras la detección han sido procesados para así analizar los resultados de la red y su fiabilidad, permitiendo comprobar que objetos han sido descartados en el mapa de ocupación final. Los resultados obtenidos con datos reales provenientes de un robot móvil dotado con una cámara RGBD, validan la propuesta de este trabajo, verificando que el mapa de ocupación resultante solo contiene los elementos estructurales del entorno.

Palabras clave: robótica móvil, localización, mapa de ocupación, cámara RGBD, imagen de profundidad, detección de objetos.

Abstract

Currently, mobile robots use occupancy grid maps to locate themselves and navigate autonomously. These maps are generally generated from 2D laser scans that detect the obstacles present in the environment and their distances from the robot, thus making it possible to characterise areas of the working environment as occupied or free. One of the main limitations of this procedure is the impossibility of distinguishing a static object, such as a wall, from a moving object, such as a chair or a sofa, both of which are included indistinctly in the occupancy grid map. When the mobile objects in the environment change their position, the result is that the occupancy grid map originally created is no longer representative and, therefore, loses its usefulness for the robot, being necessary to regenerate the occupancy grid map to represent the current configuration of the environment. The aim of this work is to take advantage of the information contained in a sequence of images and their corresponding depth data provided by an RGBD camera, to generate an occupancy grid map where only the structural elements of the environment are represented. Assuming that the camera observes part of the walls that make up the working environment at any given moment, together with a series of moving objects, this work proposes a depth data processing that allows for the elimination of objects that are between the wall and the robot, which are not included in the occupancy grid map. At the same time, the colour image of the same camera has been used to carry out an object detection process using a convolutional neural network already trained. The data obtained after the detection has been processed in order to analyse the results of the network and its reliability, allowing to check which objects have been discarded in the final occupancy grid map. The results obtained with real data from a mobile robot equipped with an RGBD camera validate the proposal of this work, verifying that the resulting occupancy grid map only contains the structural elements of the environment.

Keywords: mobile robotic, localization, occupancy grid map, RGBD camera, depth image, object detection.

Agradecimientos

Quisiera transmitir mi más sincero agradecimiento a todos los que han ayudado a cerrar esta etapa tan bonita.

Gracias a mis tutores Javier González Monroy y Cipriano Galindo Andrades por brindarme la oportunidad de realizar este trabajo. Gracias por ayudarme en todas las etapas y por hacerlo siempre con tanta amabilidad y predisposición.

Gracias a mis padres, Manuel y Rosa, y a mi hermano Manuel por apoyarme en los momentos más difíciles durante estos cuatro años.

Gracias a mi pareja, Laura, por ese apoyo incondicional y por escuchar mis males y mis quejas sin poner ni tan siquiera alguna mala cara.

A todos ellos, muchas gracias.

Índice de Contenidos

Resumen	V
Abstract	VII
Agradecimientos	IX
Lista de Figuras	XIII
Lista de Tablas	XVII
I Introducción	XIX
1. Introducción y Visión General	1
1.1. La Localización como una Necesidad	1
1.2. Objetivo	6
1.3. Estructura del Documento	7
II Desarrollo	9
2. Herramientas Utilizadas	11
2.1. Robot Operating System	11
2.1.1. Arquitectura de la Comunicación mediante ROS	12
2.1.2. ¿Por qué ROS?	13
2.2. Mapas de Ocupación	14
2.3. Sensores RGBD	15
2.3.1. Imagen de Profundidad	15
2.4. Nubes de Puntos	16
2.5. YOLO: You only look once	17
2.5.1. Redes Neuronales Convolucionales	19
2.6. Pandas y Matplotlib	19

ÍNDICE DE CONTENIDOS

3. Diseño e Implementación	21
3.1. Vista General del Sistema	21
3.2. Fuente de Datos	22
3.3. De Imagen de Profundidad a Barrido Láser	24
3.4. Generación del Mapa de Ocupación	26
3.5. Detección de Objetos	27
3.5.1. Descompresión de la Imagen de Color	27
3.5.2. Ejecución de YOLO	28
3.6. Procesamiento de los Objetos Detectados mediante Archivos CSV . .	29
3.7. Análisis de los Objetos Detectados	30
4. Resultados Obtenidos	31
4.1. Mapa Generado a partir de la Imagen de Profundidad	32
4.1.1. Generación del Láser Artificial	34
4.1.2. Modificación de Parámetros por Defecto	37
4.2. Análisis de la Fiabilidad de las Detecciones	38
III Conclusiones	43
5. Conclusiones, Propuestas y Líneas Futuras	45
5.1. Conclusiones	45
5.2. Propuestas y Líneas Futuras	47
IV Apéndices	49
A. Otras Herramientas Utilizadas	51
A.1. Librerías y Paquetes	51
A.1.1. <i>message_filters</i>	51
A.1.2. <i>fstream</i>	51
A.1.3. <i>libfreenect</i>	52
A.2. RVIZ	52
A.3. Google Colab	52
A.4. CUDA: Compute Unified Device Architecture	52
A.5. LaTex y TexMaker	53
A.6. Draw.io	53

Listas de Figuras

1.1.	Detalle de la red de metro de la Comunidad de Madrid. [2]	2
1.2.	Fotografía del robot SHAKEY tomada en 1968. [4]	3
1.3.	Ejemplos de tipos de mapas.	4
1.4.	Comparación del resultado teórico de una medición de un láser 2D y un sensor RGBD sobre elementos presentes en la Escuela de Ingenierías Industriales.	4
1.5.	Mapa generado con el barrido de un láser. Las zonas enumeradas muestran que el mapa generado no representa fielmente la estructura del entorno debido a la presencia de elementos que pueden variar en su posición como un sofá, una silla, una cama, etc.	5
2.1.	Mecanismos de comunicación en ROS. (a) Topics, (b) Servicios y (c) Acciones. [11]	13
2.2.	Ejemplo de mapa de ocupación en formato .pgm.	14
2.3.	Fotografía del sensor Microsoft Kinect V1, lanzado como complemento de la videoconsola Xbox360 en 2010. [13]	15
2.4.	Ejemplo de imagen de profundidad [14]. Generalmente, los objetos más alejados se representan en tonos más claros, mientras que los tonos más oscuros representan las zonas más cercanas. Las zonas donde es un color negro puro son zonas de “no información” de las que el sensor no puede concretar ninguna distancia.	16
2.5.	Ejemplo de una nube de puntos. [15]	17
2.6.	Comparación de YOLO con otros modelos mostrando el tiempo de inferencia (eje X) y la métrica COCO mAP-50 (eje Y) [16]. Se puede ver como YOLO tiene unos tiempos de inferencia muchos menores que la mayoría de algoritmos, manteniendo buena puntuación en la métrica.	18
2.7.	Procedimiento llevado a cabo por YOLO [16]. Divide la imaegn en una rejilla de $S \times S$ y predice las bounding boxes junto con sus probabilidades. Las bajas probabilidades se eliminan y, posteriormente las cajas duplicadas, obteniendo así la detección final.	18
2.8.	Capas de una red neuronal artifical. Capa de entrada (izquierda), capas ocultas (centro) y capa de salida (derecha). [18]	19

LISTA DE FIGURAS

2.9. Capas de una red neuronal convolucional (CNN). <i>Convolution</i> , para la operación de convolución; <i>Pooling</i> o <i>Downsampling</i> , para la operación de submuestreo; <i>Flattening</i> , para la transformación de arrays 2D a vectores lineales y <i>Full Connection</i> , una red neuronal artificial basada en perceptrón multicapa. [21]	20
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.	22
3.2. Diagrama del nodo <code>cam_info</code> . Recibe la imagen de profundidad y publica la información de la cámara.	23
3.3. Diagrama del nodo <code>sync_info</code> . Recibe la imagen de profundidad y la información de la cámara y publica los mismo mensajes pero sincronizados.	24
3.4. Ejecución del nodo <code>darknet_ros</code> . En la parte izquierda se muestran las bounding boxes sobre la imagen a color. A la derecha, en la terminal, se muestran los FPS actuales y los objetos y sus probabilidades	28
4.1. Robot Giraff y sus componentes. [30]	31
4.2. Mapas generados con láser 2D (a) y con sensor RGBD (b). Se puede observar la diferencia de rectitud en las líneas exteriores (problema) así como lo poco ruidoso que resulta el interior (virtud).	32
4.3. Superposición de los dos mapas. Los objetos no estructurales se han eliminado correctamente.	33
4.4. Nubes de puntos generadas con la configuración de fábrica de la cámara (izq.) y con la cámara calibrada (der.). Las líneas negras representan las referencias de 1 a 4 metros. A mayores distancias, las distorsiones se hacen más evidentes. [31]	34
4.5. Muestra de medida 1: diferencia entre las medidas del láser real (blancas) y del láser artificial (rojas).	35
4.6. Muestra de medida 2: diferencia entre las medidas del láser real (blancas) y del láser artificial (rojas).	35
4.7. Muestra de medida 3: diferencia entre las medidas del láser real (blancas) y del láser artificial (rojas).	36
4.8. Muestra de medida 4: diferencia entre las medidas del láser real (blancas) y del láser artificial (rojas).	36
4.9. Resultados para diferentes parámetros de <code>linearUpdate</code> , <code>angularUpdate</code> y <code>temporalUpdate</code> , respectivamente.	37
4.10. Resultados para distintos valores de distancia máxima. A mayor distancia máxima, mayor influencia de la información percibida por ventanas y puertas abiertas.	38
4.11. Trayectoria seguida por el robot. El punto rojo indica el inicio de la trayectoria. El punto verde indica el final.	39

LISTA DE FIGURAS

4.12. Gráfica de objetos detectados. Muestra el número de veces que fue detectado cada uno de los objetos.	39
4.13. Momentos en los que se ha detectado la clase ‘oven’. Los puntos azules representan la trayectoria del robot, los rojos las veces que se detectó la clase.	40
4.14. Detección de la clase ‘oven’ en la imagen.	40
4.15. Momentos en los que se ha detectado la clase ‘refrigerator’. Los puntos azules representan la trayectoria del robot, los rojos las veces que se detectó la clase de forma incorrecta y los verdes las veces que los detectó correctamente.	41
4.16. Detecciones de la clase ‘refrigerator’ en la imagen.	41

Lista de Tablas

3.1.	Parámetros de lanzamiento del nodo <code>depthimage_to_laserscan</code>	25
3.2.	Formato de un mensaje de tipo <code>LaserScan</code>	26
3.3.	Formato del archivo CSV generado.	29
3.4.	Formato de un mensaje de tipo <code>BoundingBox</code>	29
4.1.	Comparativa de las características según datasheets del láser <i>Hokuyo URG-04LX-UG01</i> y de la cámara <i>Orbbec Astra</i> , dispositivos utilizados en el dataset. [34] [35]	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 tecnologías de localización como GPS, existen ciertas situaciones en las que es complicado aprovecharse de ellas, por ejemplo en los entornos indoor. En estos caso, los satélites no son capaces de

CAPÍTULO 1. INTRODUCCIÓN Y VISIÓN GENERAL

comunicarse con nuestros dispositivos, por lo que necesitamos otro mecanismo para localizarnos.

En la figura 1.1 se muestra una parte del 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 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: Detalle de la red de metro de la Comunidad 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 de disponer de una ruta fijada previamente. Para poder llevarlo a cabo, los robots

1.1. LA LOCALIZACIÓN COMO UNA NECESIDAD



Figura 1.2: Fotografía del robot SHAKEY tomada en 1968. [4]

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. De forma general, podemos distinguir 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 un tercer tipo de mapa que es un híbrido entre los dos tipos 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.

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

CAPÍTULO 1. INTRODUCCIÓN Y VISIÓN GENERAL

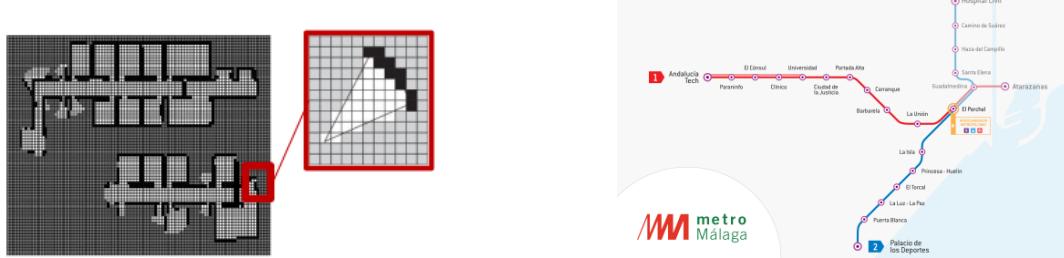


Figura 1.3: Ejemplos de tipos de mapas.

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 1.4 se muestra un ejemplo de lo que se quiere conseguir aprovechando las ventajas de los sensores RGBD frente a los láseres 2D aplicado a la Escuela de Ingenierías.

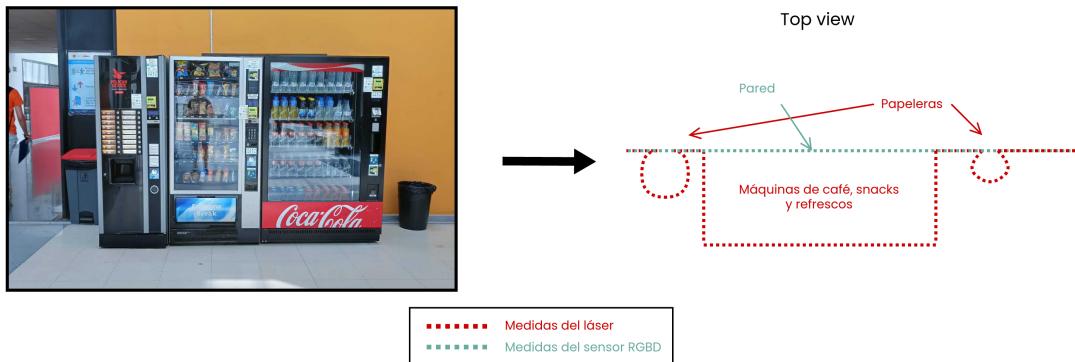


Figura 1.4: Comparación del resultado teórico de una medición de un láser 2D y un sensor RGBD sobre elementos presentes en la Escuela de Ingenierías Industriales.

1.1. LA LOCALIZACIÓN COMO UNA NECESIDAD

Otro ejemplo se puede ver en la figura 1.5 donde se muestra 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 generado no podría ser utilizado por el robot para la navegación.



Figura 1.5: Mapa generado con el barrido de un láser. Las zonas enumeradas muestran que el mapa generado no representa fielmente la estructura del entorno debido a la presencia de elementos que pueden variar en su posición como un sofá, una silla, una cama, etc.

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:

CAPÍTULO 1. INTRODUCCIÓN Y VISIÓN GENERAL

- 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, este sistema permitiría obtener las dimensiones reales de la vivienda y así tener una idea en 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 éstos 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 a 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 dividido en 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 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.

En una cuarta y última parte se sitúan los apéndices del proyecto. En el primero y único se explica el software que se ha utilizado de apoyo pero que no se ha mencionado en los capítulos anteriores.

Parte II

Desarrollo

Capítulo 2

Herramientas Utilizadas

En este capítulo se describen brevemente las herramientas que se han utilizado en este trabajo. Se hará una introducción a ROS y sus mecanismos para la comunicación. Se explicarán los conceptos de “sensor RGBD”, “mapa de ocupación” y “nube de puntos”. Seguidamente se introducirá el concepto de “red neuronal convolucional” y de “YOLO”, la red utilizada en este proyecto. Por último, se expondrán muy brevemente las herramientas de Python utilizadas para el análisis de los datos.

2.1. Robot 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 [8]. 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.

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. [7]

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. [9]

CAPÍTULO 2. HERRAMIENTAS UTILIZADAS

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 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 en qué topic 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 como 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 pero, a diferencia de los servicios, no son bloqueantes y permiten ser canceladas mientras se están ejecutando.

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. [10]

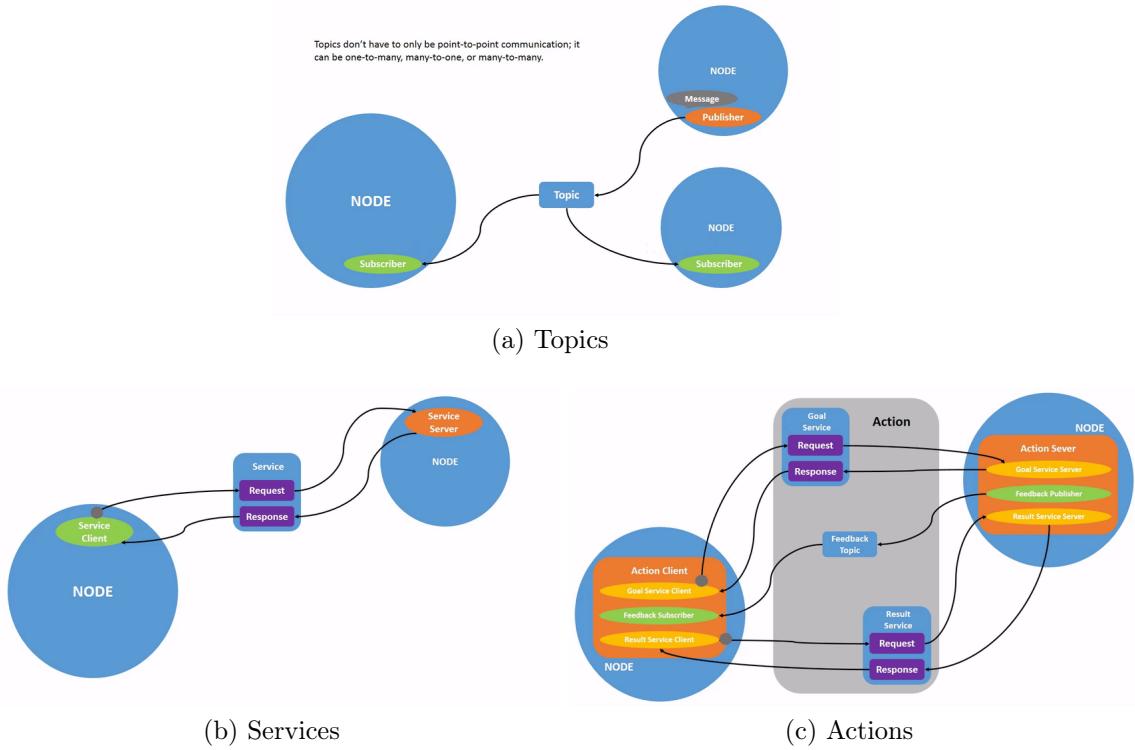


Figura 2.1: Mecanismos de comunicación en ROS. (a) Topics, (b) Servicios y (c) Acciones. [11]

ROS también implementa un mecanismo muy útil a la hora de probar los programas que se desarrollan: el *rosbag*. 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 se requieran, 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*. [11]

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 desarro-

CAPÍTULO 2. HERRAMIENTAS UTILIZADAS

lladores infinitas de posibilidades.

Pero la que es quizás la mayor ventaja de ROS frente a otros frameworks y software especializados en robótica es la gran comunidad que le ampara. 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, las celdas libres en color blanco y en gris se representa el espacio no explorado aún por el robot.

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) [12].

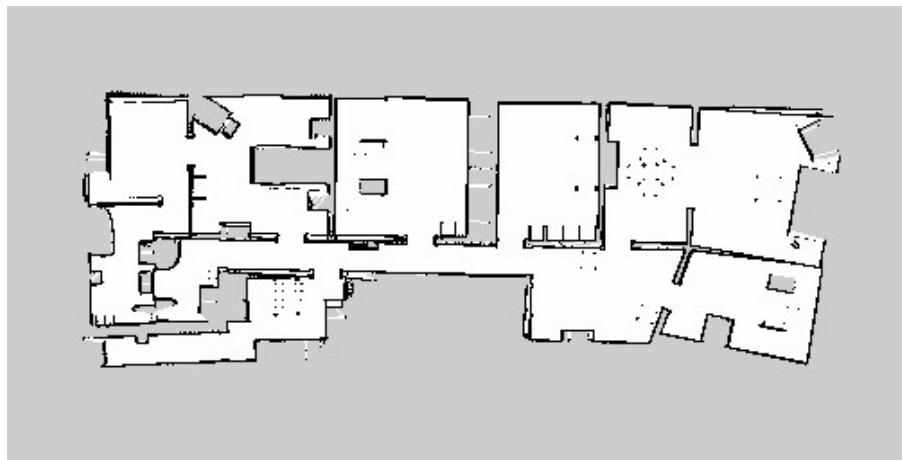


Figura 2.2: Ejemplo de mapa de ocupación en formato .pgm.

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



Figura 2.3: Fotografía del sensor Microsoft Kinect V1, lanzado como complemento de la videoconsola Xbox360 en 2010. [13]

La cámara Microsoft Kinect V1 cuenta con un emisor infrarrojos capaz de proyectar patrones de luz infrarroja para después recibirlos por la cámara de profundidad y analizar las distorsiones sufridas en el patrón emitido debido a la colisión con los elementos del entorno. Con esta información el propio sensor es capaz de construir un mapa 3D o nube de puntos del entorno que está capturando.

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. En la figura 2.4 se muestra un ejemplo de una 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.

CAPÍTULO 2. HERRAMIENTAS UTILIZADAS

En muchas ocasiones, los sensores RGBD no consiguen obtener la distancia de los objetos capturados debido a algunos factores como el tipo de material o el ruido presente en la escena. En estos casos, el sensor colorea las zonas de las que desconoce su distancia de color negro puro.



Figura 2.4: Ejemplo de imagen de profundidad [14]. Generalmente, los objetos más alejados se representan en tonos más claros, mientras que los tonos más oscuros representan las zonas más cercanas. Las zonas donde es un color negro puro son zonas de “no información” de las que el sensor no puede concretar ninguna distancia.

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. Dada la imagen de profundidad y la configuración intrínseca de la cámara, se puede obtener esta representación tridimensional. También se puede generar una nube de puntos a partir de láseres 3D, sensores con mayor precisión pero mucho menos económicos.

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.

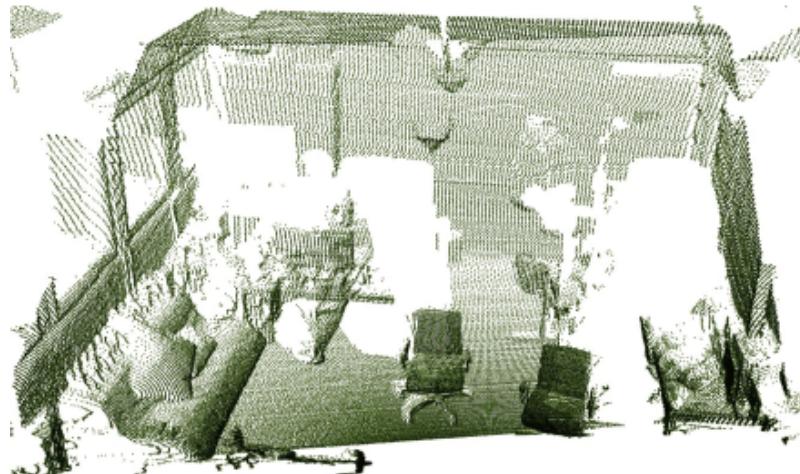


Figura 2.5: Ejemplo de una nube de puntos. [15]

2.5. YOLO: You only look once

You only look once (YOLO) es un sistema en tiempo real de detección de objetos basado en redes 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. Esta propiedad le hace ser más rápido que los modelos competidores permitiendo la detección en tiempo real en vídeos de hasta 30 FPS. Según comentan sus creadores, con respecto a otras redes, es 1000 veces más rápido que R-CNN y 100 veces más rápido que Fast R-CNN. [16]

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 predice N posibles cajas delimitadoras o *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. Más información sobre este software en el apéndice A.4.

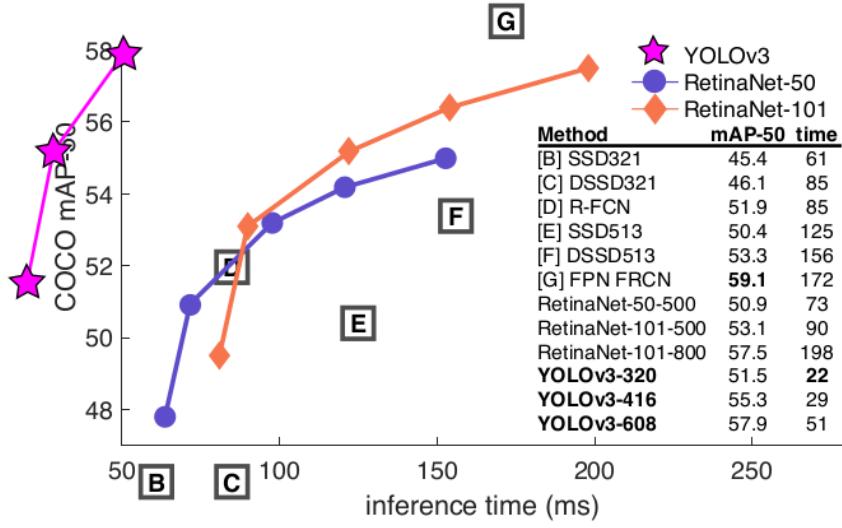


Figura 2.6: Comparación de YOLO con otros modelos mostrando el tiempo de inferencia (eje X) y la métrica COCO mAP-50 (eje Y) [16]. Se puede ver como YOLO tiene unos tiempos de inferencia muchos menores que la mayoría de algoritmos, manteniendo buena puntuación en la métrica.

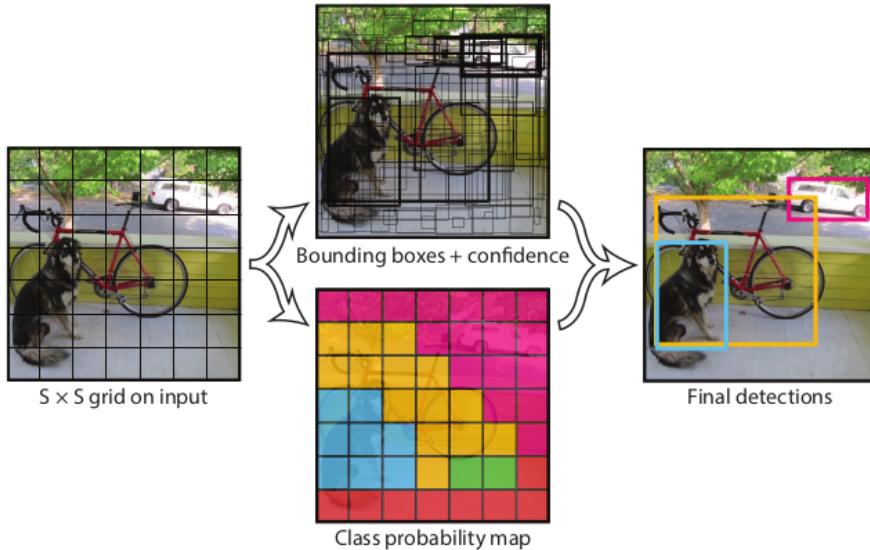


Figura 2.7: Procedimiento llevado a cabo por YOLO [16]. Divide la imagen en una rejilla de $S \times S$ y predice las bounding boxes junto con sus probabilidades. Las bajas probabilidades se eliminan y, posteriormente las cajas duplicadas, obteniendo así la detección final.

2.5.1. Redes Neuronales Convolucionales

Una red neuronal artificial o ANN (del inglés, *Artificial Neural Network*) es un modelo simplificado que emula el modo en que el cerebro humano procesa la información, simultaneando multitud de unidades de procesamiento interconectadas, emulando una versión artificial de las neuronas. Las neuronas se organizan en capas. En una red neuronal hay tres tipos de capas: la capa de entrada, unas o varias capas ocultas y la capa de salida. Las neuronas se conectan con las neuronas de la siguiente capa mediante ponderaciones, propagando los datos de entrada a la salida. La red aprende examinando los registros individuales, generando una predicción para cada registro y, en caso de haber realizado una predicción incorrecta, corrigiendo las ponderaciones. Este procedimiento, conocido como entrenamiento, se repite hasta alcanzar criterios de finalización. Una vez la red ha sido entrenada, puede aplicarse en casos en los que no conoce el resultado para obtener, en función del tipo y de la cantidad de entrenamiento, unos resultados satisfactorios o erróneos. [17]

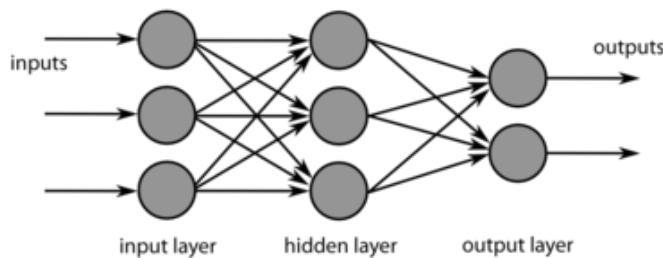


Figura 2.8: Capas de una red neuronal artificial. Capa de entrada (izquierda), capas ocultas (centro) y capa de salida (derecha). [18]

Una red neuronal convolucional o CNN (del inglés, *Convolutional Neural Network*) son un tipo de NN donde las neuronas corresponden a campos receptivos de una manera muy similar a las neuronas de la corteza visual primaria de un cerebro humano. Es una variación del perceptrón multicapa⁴, muy efectiva para tareas de visión artificial. Las capas ocultas de una CNN están especializadas en determinadas funciones de forma que las primeras capas pueden detectar líneas o curvas y se van especializando hasta llegar a capas más profundas que reconocen formas complejas como rostros o siluetas. [20]

2.6. Pandas y Matplotlib

Pandas es una librería de código abierto de Python utilizada en el ámbito del *Data Science* y *Machine Learning*. Ofrece estructuras muy poderosas y flexibles que

⁴El perceptrón multicapa es una red neuronal artificial formada por múltiples capas, de tal forma que tienen capacidad para resolver problemas que nos son linealmente separables. [19]

CAPÍTULO 2. HERRAMIENTAS UTILIZADAS

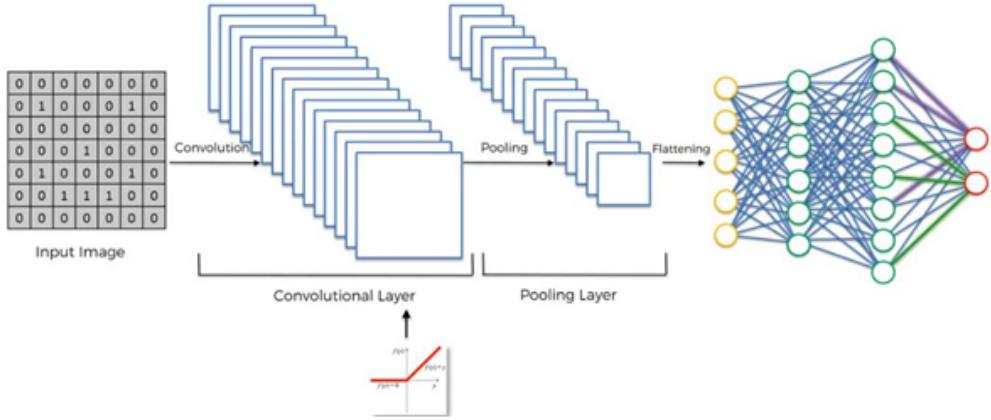


Figura 2.9: Capas de una red neuronal convolucional (CNN). *Convolution*, para la operación de convolución; *Pooling* o *Downsampling*, para la operación de submuestreo; *Flattening*, para la transformación de arrays 2D a vectores lineales y *Full Connection*, una red neuronal artificial basada en perceptrón multicapa. [21]

facilitan la manipulación y el tratamiento de los datos. [22]

Matplotlib es una librería de Python diseñada para la creación de visualizaciones estáticas, animadas e interactivas. Permite la creación y personalización de todo tipo de representaciones gráficas de datos, como gráficos de barras, sectores circulares, puntos, etc.[23]

Estas herramientas, en combinación con la librería *Numpy* (diseñada para el manejo de arreglos con mayor rendimiento), son las herramientas más utilizadas en el ámbito de la ciencia de datos, ofreciendo multitud de herramientas y funcionalidades que simplifican el manejo de los mismos.

Capítulo 3

Diseño e Implementación

Para implementar el sistema que nos permita obtener estos mapas de ocupación a partir de la imagen de profundidad capturada por una cámara RGBD, 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 e implementados 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.

Con respecto a los paquetes de terceros, se hará una breve explicación sobre su función y la funcionalidad que se le han dado en este trabajo, así como el modo de comunicación con todo el grafo 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 se generará un archivo de texto o CVS con los elementos detectados por YOLO de la imagen de color para tener un registro y poder analizarlo. En la figura 3.1 se muestra el grafo 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.

- 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.

CAPÍTULO 3. DISEÑO E IMPLEMENTACIÓN

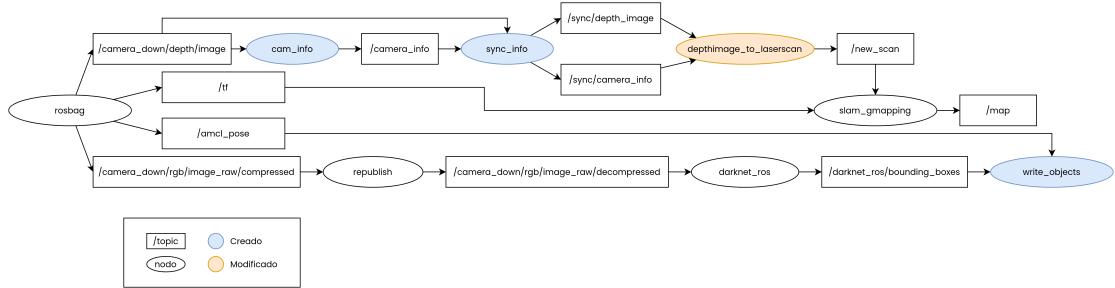


Figura 3.1: Árbol de nodos y topics. Los nodos diseñados para este trabajo son de color azul. Los nodos de terceros cuyo código ha sido modificado son de color naranja. Los nodos descargados que no se han modificado son de color blanco.

- 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.
- 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.
- `depthimage_to_laserscan` es un nodo de un paquete homónimo diseñado por Chad Rocke [24] 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 [25]. 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 [26]. 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 [27]. Implementa YOLO en ROS.

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 *Giraff* 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 RGBD (color y profundidad), el barrido de un láser y otra información que típicamente se utiliza en los robots. 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 las condiciones del *rosbag* utilizado, este último mensaje no se ofrece, por lo que es necesario generarla externamente. Para ello, se ha diseñado un nodo llamado `cam_info` que se encarga de generar este mensaje.

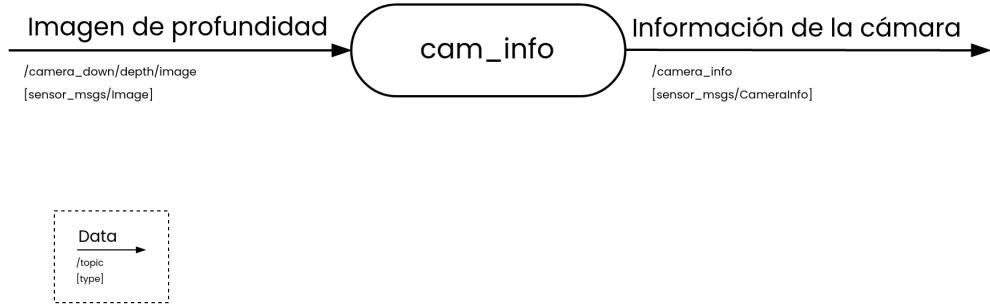


Figura 3.2: Diagrama del nodo `cam_info`. Recibe la imagen de profundidad y publica la información de la cámara.

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++.

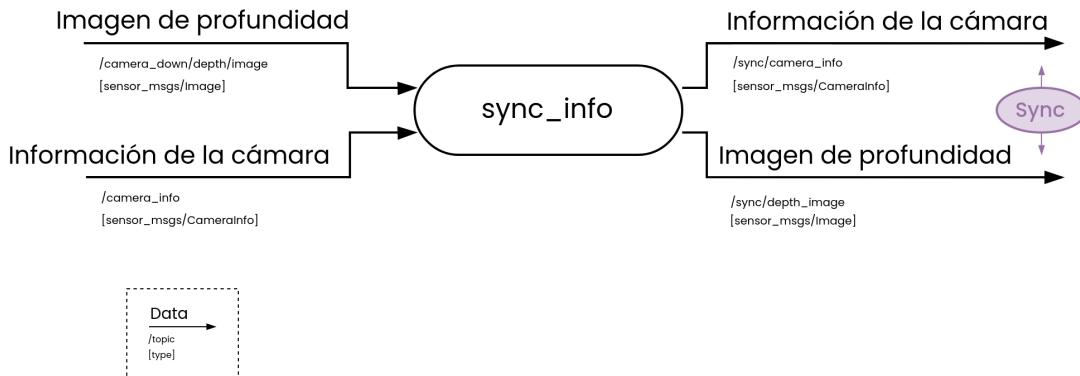


Figura 3.3: Diagrama del nodo `sync_info`. Recibe la imagen de profundidad y la información de la cámara y publica los mismo mensajes pero sincronizados.

3.3. De Imagen de Profundidad a Barrido Láser

Una vez obtenida 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 RGBD en un barrido láser equivalente, cogiendo las 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.

3.3. DE IMAGEN DE PROFUNDIDAD A BARRIDO LÁSER

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. Los parámetros se muestran en la tabla 3.1.

Parámetro	Valor por defecto	Descripción
<code>scan_height</code>	1	Establece la cantidad de filas que se quieren procesar para generar el láser
<code>scan_time</code>	0.033 (30FPS)	Establece el tiempo de actualización entre escaneos
<code>range_min</code>	0.45 metros	Rango de distancia mínima. Valores medidos menores que este valor se tomarán como -Inf
<code>range_max</code>	10 metros	Rango de distancia máxima. Valores medidos menores que este valor se tomarán como +Inf
<code>output_frame_id</code>	<code>laser</code>	Establece el id del eje de coordenadas (frame) de salida. Se indica el id del frame del láser

Tabla 3.1: Parámetros de lanzamiento del nodo `depthimage_to_laserscan`.

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, procede a convertir la imagen de profundidad a un mensaje de tipo LaserScan. El formato de un mensaje de tipo LaserScan se expone en la tabla 3.2.

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`.

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 [28]. El paquete `depthimage_to_laserscan` solo admite las dos codificaciones comentadas.

CAPÍTULO 3. DISEÑO E IMPLEMENTACIÓN

Parámetro	Descripción
<code>header</code>	Encabezado del mensaje. Propiedad que tienen todos los mensajes
<code>angle_min</code>	Ángulo mínimo que abarca el láser en radianes
<code>angle_max</code>	Ángulo máximo en radianes
<code>angle_increment</code>	Ángulo entre medidas en radianes
<code>time_increment</code>	Tiempo entre medidas en segundos
<code>scan_time</code>	Tiempo entre escaneos en segundos
<code>range_min</code>	Rango mínimo del sensor en metros
<code>range_max</code>	Rango máximo del sensor en metros
<code>ranges</code>	Vector con las medidas tomadas en metros
<code>intensities</code>	Vector con las intensidades de las medidas. Se tomará como un array vacío porque no es necesario

Tabla 3.2: Formato de un mensaje de tipo `LaserScan`.

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 más fiable. Esto se comentará más en profundidad en el capítulo 4.

El nodo `slam_gmapping` envía por el topic `/map` mensajes de tipo `nav_msgs-/OccupancyGrid` con información sobre el mapa de forma que, tras finalizar la reproducción del rosbag, el último mensaje enviado por este topic contendrá el mapa completo que se ha generado. Para guardarlo en un formato con el que poder trabajar, se utiliza el nodo `map_saver` del paquete `map_server`². Simplemente se lanza el nodo dándole como parámetro el nombre con el que se va a guardar el mapa. Finalmente se obtienen dos archivos: un `.yaml`, para la configuración, y otro en formato `.pgm` para la imagen.

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 memoria. 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 recibe un mensaje en formato `sensor_msgs-/CompressedImage` y lo convierte en un mensaje de tipo `sensor_msgs/Image` para publicarlo.

²`map_server` es un paquete diseñado por Brian Gerkey y Tony Pratkanis que ofrece dos funcionalidades: el nodo `map_server` permite leer un mapa desde un archivo para publicarlo en un topic y el nodo `map_saver` permite, mediante un comando, obtener el mapa que se está publicando. En este trabajo se utilizará este último nodo. [29]

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 los límites del 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.

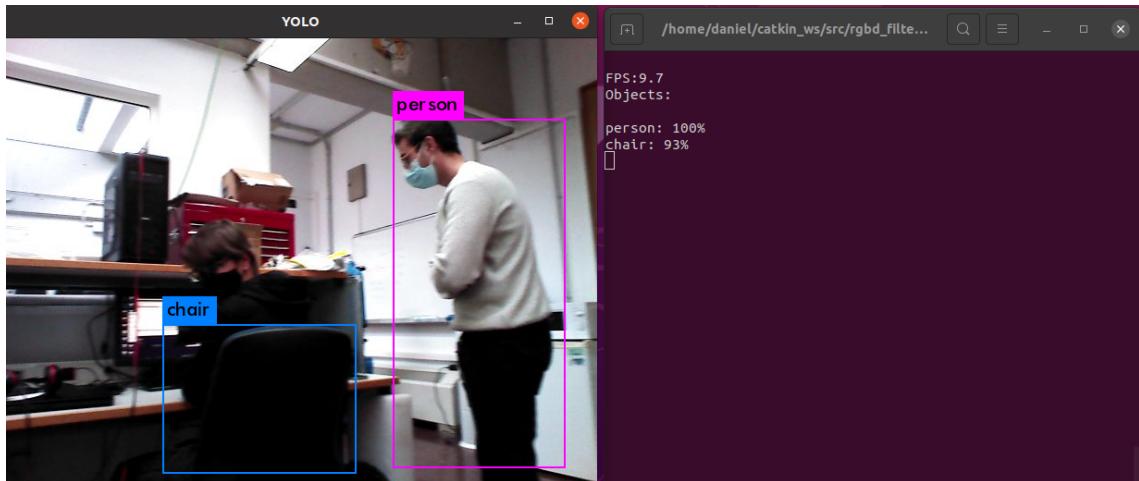


Figura 3.4: Ejecución del nodo `darknet_ros`. En la parte izquierda se muestran las bounding boxes sobre la imagen a color. A la derecha, en la terminal, se muestran los FPS actuales y los objetos y sus probabilidades

3.6. PROCESAMIENTO DE LOS OBJETOS DETECTADOS MEDIANTE ARCHIVOS CSV

3.6. Procesamiento de los Objetos Detectados mediante Archivos CSV

Como se ha comentado, el nodo que lanza YOLO publica 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 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 la tabla 3.3. 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)

Tabla 3.3: 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* (posición y orientación, en formato de quaternion). 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* se representa en la tabla 3.4.

Parámetro	Descripción
<code>probability</code>	Probabilidad de que el elemento detectado sea realmente de la clase que se le atribuye
<code>xmin</code>	Coordenada X mínima de la caja en píxeles
<code>ymin</code>	Coordenada Y mínima de la caja en píxeles
<code>xmax</code>	Coordenada X máxima de la caja en píxeles
<code>ymax</code>	Coordenada Y máxima de la caja en píxeles
<code>id</code>	Identificador único para cada bounding box
<code>Class</code>	Clase o etiqueta del elemento detectado

Tabla 3.4: Formato de un mensaje de tipo `BoundingBox`.

3.7. Análisis de los Objetos Detectados

Una vez obtenido el registro de los objetos en formato CSV es posible procesar esta información para realizar un análisis de la red y del entorno. Para ello se utilizará el entorno de programación en Python de *Google Colab*, junto con las librerías *Pandas* y *Matplotlib*. La información sobre las gráficas y los resultados obtenidos se exponen en el capítulo 4.

Capítulo 4

Resultados Obtenidos

En este capítulo se mostrarán los resultados que se han obtenido tras implementar el sistema. Primeramente se hará una comparativa con el resultado obtenido a partir del láser 2D y se expondrán los mapas generados utilizando diferentes parámetros. En una segunda parte se mostrarán algunas gráficas realizadas con el registro de objetos detectados y se expondrán los posibles errores de la red neuronal en el proceso de la detección.

Los datos con los que se ha trabajado proceden de la navegación de un robot móvil dotado de cámaras RGBD y láser 2D en un laboratorio de investigación de la Escuela Técnico Superior de Ingeniería Informática. El robot mencionado es el robot *Giraff* [30]. Entre otros elementos, este robot está dotado de dos cámaras RGBD modelo *Orbbec Astra* y un escáner láser *Hokuyo URG-04LX-UG01*. De las dos cámaras disponibles, se utilizará aquella situada a menor altura (figura 4.1).



Figura 4.1: Robot Giraff y sus componentes. [30]

4.1. Mapa Generado a partir de la Imagen de Profundidad

En la figura 4.2 se muestra el resultado de generar el mapa a partir del láser 2D (izquierda) y a partir de la imagen de profundidad captada por el sensor RGBD (derecha).

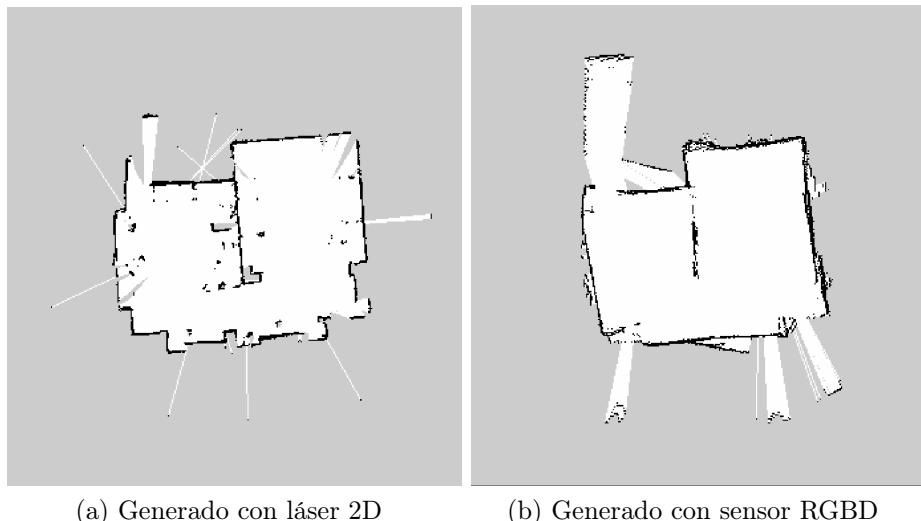


Figura 4.2: Mapas generados con láser 2D (a) y con sensor RGBD (b). Se puede observar la diferencia de rectitud en las líneas exteriores (problema) así como lo poco ruidoso que resulta el interior (virtud).

En el mapa generado con el láser se pueden ver los entrantes y salientes que se forman debido a la presencia de mobiliario delante de las paredes que impide obtener una medida real de la estructura. Además, en zonas interiores existen muchos grupos pequeños de celdas ocupadas causadas por la presencia de patas de mesas, sillas y otros elementos de menor tamaño. Sin olvidar que la grabación cuenta con la presencia de personas que modifican su posición, complicando aún más la generación del mapa.

Con el mapa obtenido gracias al procesamiento de la imagen de profundidad, se puede comprobar cómo los elementos situados entre el sensor y las paredes no se tienen en cuenta, quedando el interior del mapa totalmente “limpio”. Además, las zonas donde se sitúa el mobiliario quedan mucho más suavizadas, mostrándose únicamente la pared situada tras el mismo. En la figura 4.3 se muestran los dos mapas superpuestos para poder ver mejor la diferencia entre sendos mapas.

Pese a que mejora al láser en muchos aspectos, el mapa generado con la cámara tiene algunos problemas. Como se puede observar, las paredes no son totalmente

4.1. MAPA GENERADO A PARTIR DE LA IMAGEN DE PROFUNDIDAD

rectas sino que tienen curvatura, lo que provoca que las esquinas no formen 90°. Incluso hay paredes, como la que está situada más a la izquierda, que parecen desviarse con respecto al mapa global.

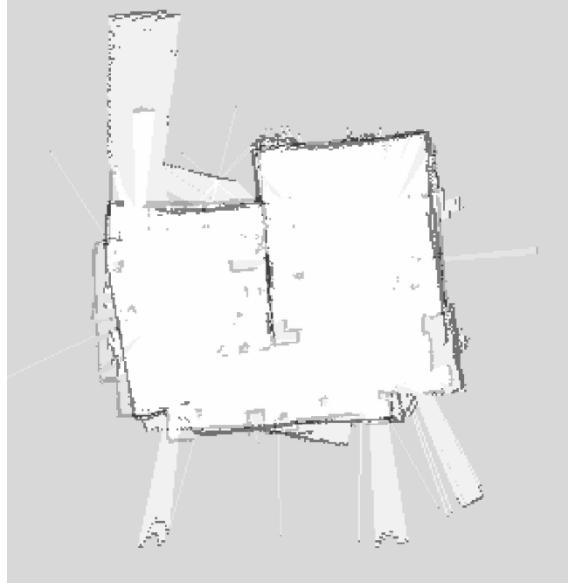


Figura 4.3: Superposición de los dos mapas. Los objetos no estructurales se han eliminado correctamente.

Estas curvaturas obtenidas en la medición puede ser causada por la calibración intrínseca de la cámara. Es posible que algunos parámetros de la cámara no estén correctamente determinados o incluso que la transformada entre la base del robot y la cámara no esté correctamente establecida. Por la curvatura en las esquinas se intuye que la causa más probable sea una mala calibración de los parámetros de distorsión de la cámara. Este resultado ya se podía percibir observando el láser generado a partir del sensor RGBD en comparación con el del láser 2D durante el proceso de construcción del mapa.

Según las características ofrecidas por los datasheets del láser y de la cámara utilizada, mostradas en la tabla 4.1, esta tiene una mejor precisión, algo incoherente con el resultado obtenido. Sin embargo, diversos estudios realizados [31] [32] [33] demuestran que los datos ofrecidos por algunos fabricantes son muy optimistas y no reflejan la realidad del sensor en determinadas situaciones. Además, destacan que el proceso de calibración de la cámara es esencial para obtener medidas fiables, ya que las estimaciones se ven afectadas por distorsiones no corregidas por sus parámetros de fábrica, como representa la figura 4.4. En este proyecto se trabaja con datos reales de una cámara que no dispone de este procedimiento, por lo que los resultados se ven inevitablemente afectados por estas distorsiones.

CAPÍTULO 4. RESULTADOS OBTENIDOS

	Láser Hokuyo URG-04LX-UG01	Cámara Orbbec Astra
Precio	1184€	209€
Precisión	$\pm 30\text{mm}$ (1m), $\pm 120\text{mm}$ (4m)	$\pm 3\text{mm}$ (1m), $\pm 36\text{mm}$ (4m)
FOV	240°	60°
Rango	60mm - 4095mm	0.6m - 8m

Tabla 4.1: Comparativa de las características según datasheets del láser *Hokuyo URG-04LX-UG01* y de la cámara *Orbbec Astra*, dispositivos utilizados en el dataset. [34] [35]

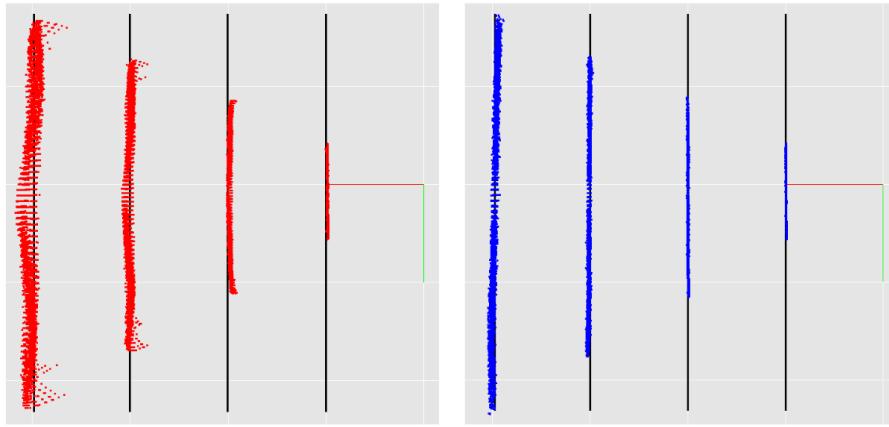


Figura 4.4: Nubes de puntos generadas con la configuración de fábrica de la cámara (izq.) y con la cámara calibrada (der.). Las líneas negras representan las referencias de 1 a 4 metros. A mayores distancias, las distorsiones se hacen más evidentes. [31]

4.1.1. Generación del Láser Artificial

Como se ha comentado previamente, el láser que se genera a partir de la imagen de profundidad tiene unas curvaturas que ya se podían percibir claramente durante el proceso de generación del mapa. En muchas ocasiones, el láser artificial generado no se superpone con el láser real en las situaciones en las que debería hacerlo, como las de las figuras 4.5 y 4.6.

En otras muchas ocasiones, como en las de las figuras 4.7 e 4.8, se puede ver como el procesamiento realiza correctamente su función, cumpliendo el objetivo inicialmente establecido de “eliminar” los objetos frente a las paredes.

4.1. MAPA GENERADO A PARTIR DE LA IMAGEN DE PROFUNDIDAD

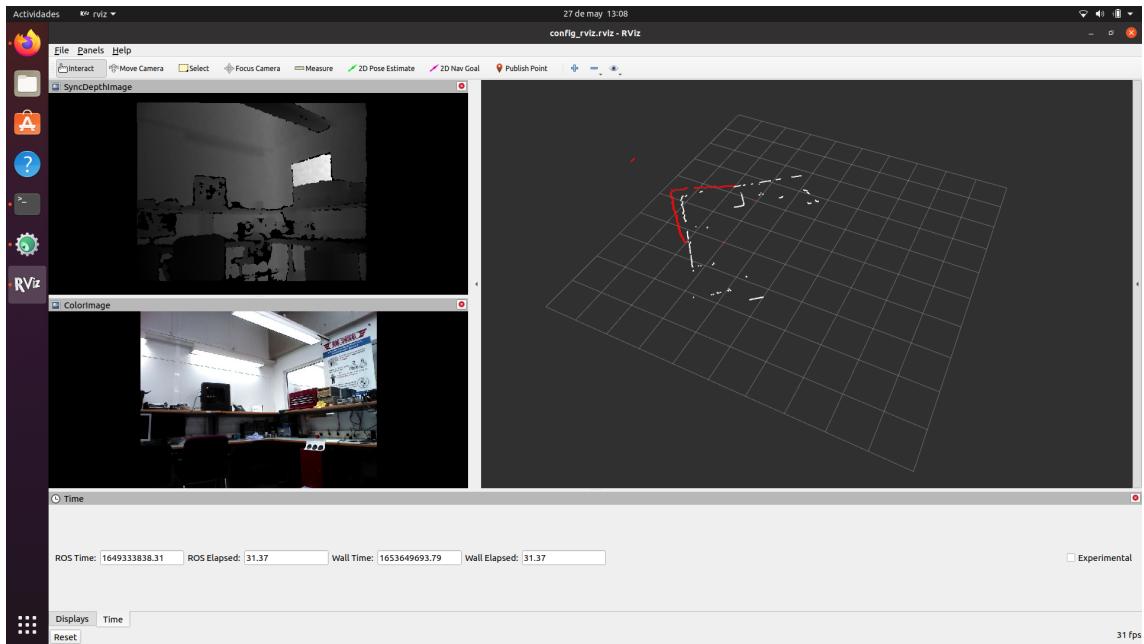


Figura 4.5: Muestra de medida 1: diferencia entre las medidas del láser real (blancas) y del láser artificial (rojas).

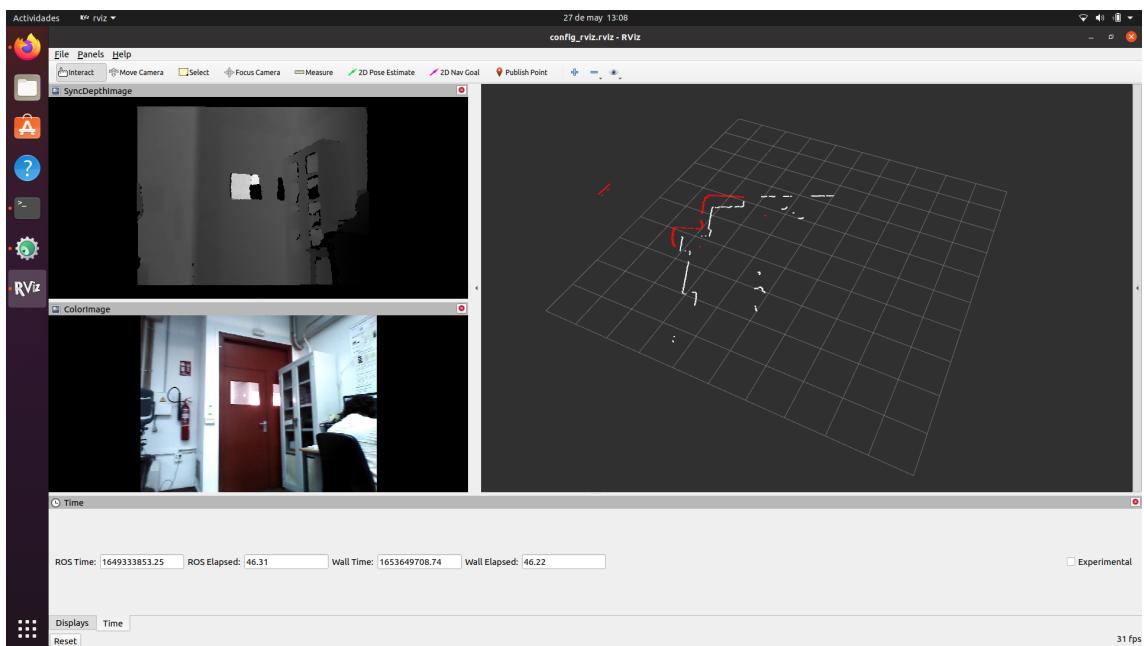


Figura 4.6: Muestra de medida 2: diferencia entre las medidas del láser real (blancas) y del láser artificial (rojas).

CAPÍTULO 4. RESULTADOS OBTENIDOS

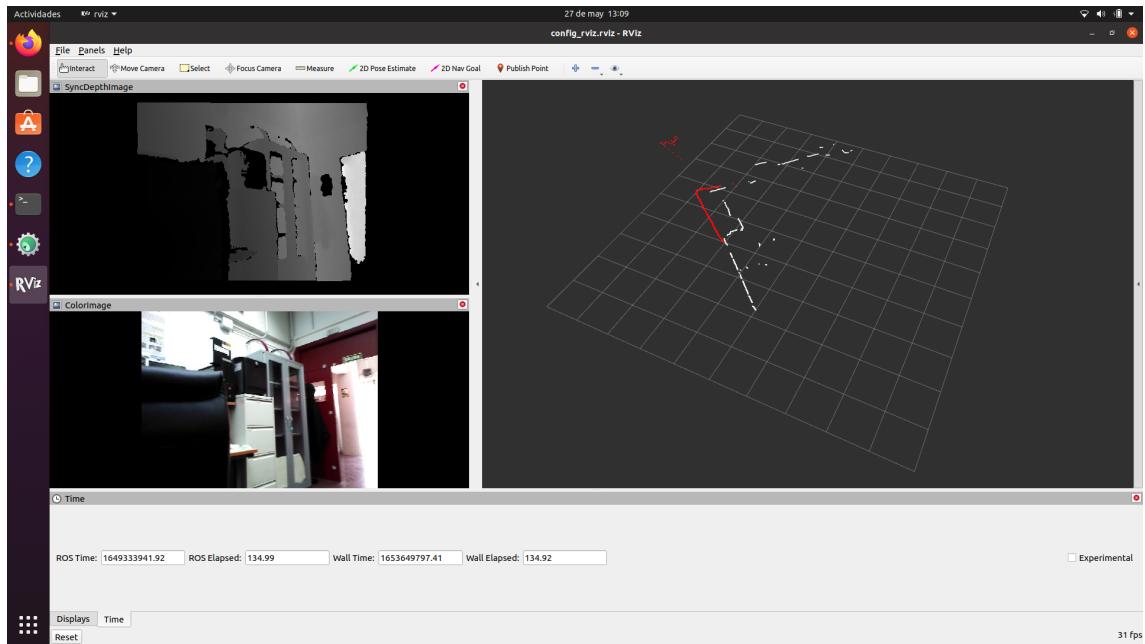


Figura 4.7: Muestra de medida 3: diferencia entre las medidas del láser real (blancas) y del láser artificial (rojas).

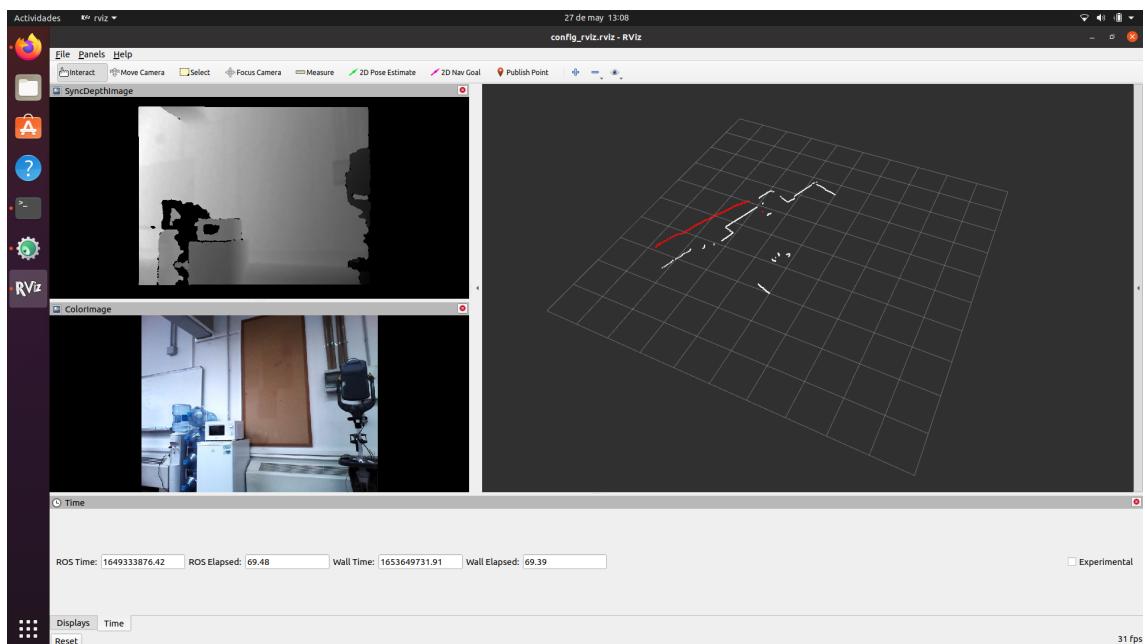


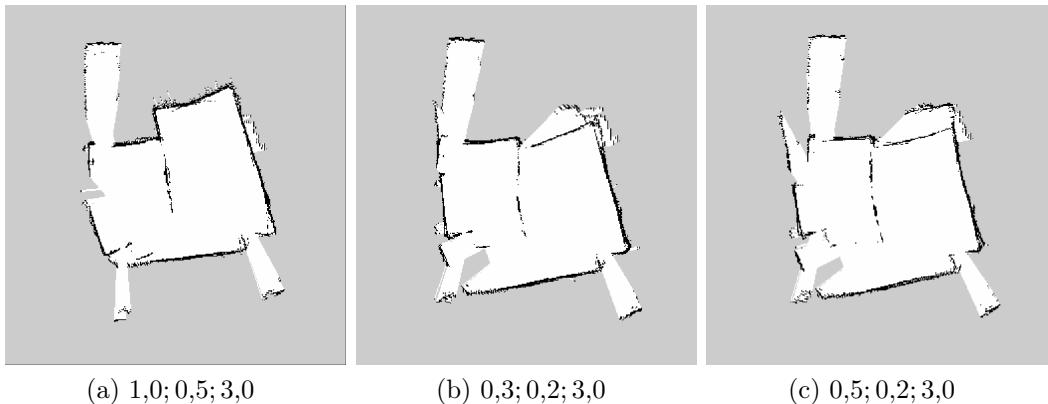
Figura 4.8: Muestra de medida 4: diferencia entre las medidas del láser real (blancas) y del láser artificial (rojas).

4.1. MAPA GENERADO A PARTIR DE LA IMAGEN DE PROFUNDIDAD

4.1.2. Modificación de Parámetros por Defecto

Como se comentó en la sección 3.4, el nodo `slam_gmapping` tiene una serie de parámetros que permiten ajustar cada cuántos metros, ángulos o segundos se procesa la información del láser.

Tras una primera ejecución del procedimiento diseñado, el resultado obtenido no fue el esperado. Es por ello que se decidió variar estos parámetros hasta obtener un resultado satisfactorio. Los mapas generados para diferentes valores de estos tres parámetros se muestran en la figura 4.9.



(a) 1,0; 0,5; 3,0

(b) 0,3; 0,2; 3,0

(c) 0,5; 0,2; 3,0

Figura 4.9: Resultados para diferentes parámetros de `linearUpdate`, `angularUpdate` y `temporalUpdate`, respectivamente.

Otro parámetro que se ha cambiado para obtener mejores resultados es la distancia máxima de profundidad, `range_max`, del nodo `depthimage_to_laserscan` comentada en el apartado 3.3. Este valor está por defecto a 10 metros. Sin embargo, debido a la existencia de amplias ventanas y puertas, el resultado se veía muy influenciado por las distancias alejadas captadas a través de estas. El valor que ha originado mejores resultados es 7 metros.

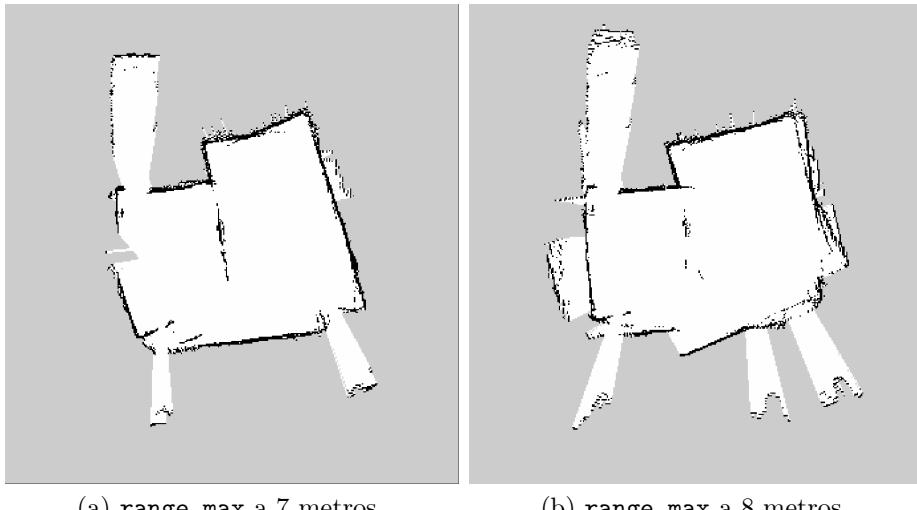


Figura 4.10: Resultados para distintos valores de distancia máxima. A mayor distancia máxima, mayor influencia de la información percibida por ventanas y puertas abiertas.

4.2. Análisis de la Fiabilidad de las Detecciones

YOLO es una de las redes neuronales más rápidas en el ámbito de la detección y localización de objetos en una imagen. Sin embargo, esta velocidad tan característica le hace sacrificar algo de exactitud en sus medidas, convirtiéndola en una red no del todo fiable. Aún así, continúa siendo de las mejores en este aspecto.

El nodo `write_objects` genera un archivo de texto con los datos en formato CSV, un formato que permite procesar los datos de manera sencilla. Este archivo contiene información sobre los objetos o clases detectadas, sus probabilidades, la pose del robot (posición y orientación) y el tiempo en el instante en que se detectaron dichos objetos. Para entender la trayectoria seguida por el robot, se propone la gráfica de la figura 4.11. Destacar que no comienza en el origen de coordenadas porque existe cierto desfase entre la ejecución de YOLO y los demás nodos, debido a la potencia necesaria para lanzarlo.

En la figura 4.12 se muestra una gráfica de barras con el número de ocasiones que fueron detectadas cada una de las clases. Las clases ‘tvmonitor’ (monitor de TV) y ‘chair’ (silla) son las que más han sido detectadas, algo coherente con la distribución del laboratorio. La tercera clase más detectada, ‘person’ (persona), tampoco se escapa de la normalidad ya que durante la grabación había personas en el laboratorio. Las clases ‘mouse’ (ratón) y ‘keyboard’ (teclado), en principio, deberían haber sido tan frecuentes como las otras. Sin embargo, la cámara está prácticamente al mismo nivel que las mesas, lo que dificulta el reconocimiento de estos objetos.

4.2. ANÁLISIS DE LA FIABILIDAD DE LAS DETECCIONES

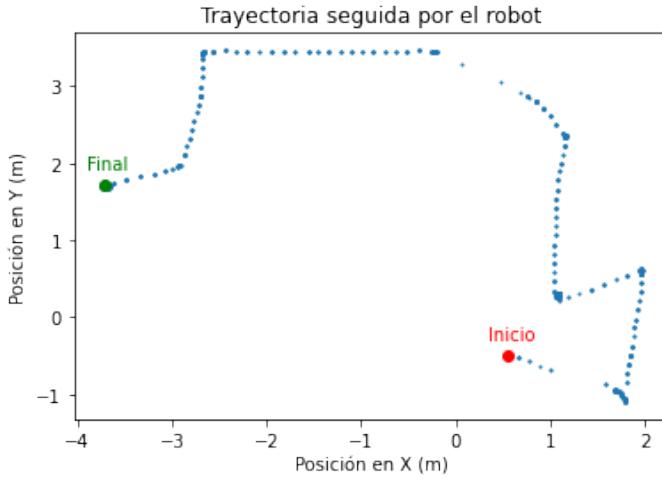


Figura 4.11: Trayectoria seguida por el robot. El punto rojo indica el inicio de la trayectoria. El punto verde indica el final.

De todos los objetos que se han detectado, hay tres que no parecen pertenecer al entorno habitual de un laboratorio. Es el caso de ‘oven’ (hornos), ‘refrigerator’ (frigorífico) y ‘microwave’ (microondas). Este último es coherente ya que en el laboratorio hay uno y las detecciones las hace correctamente. Los otros son casos especiales que se explicarán a continuación.

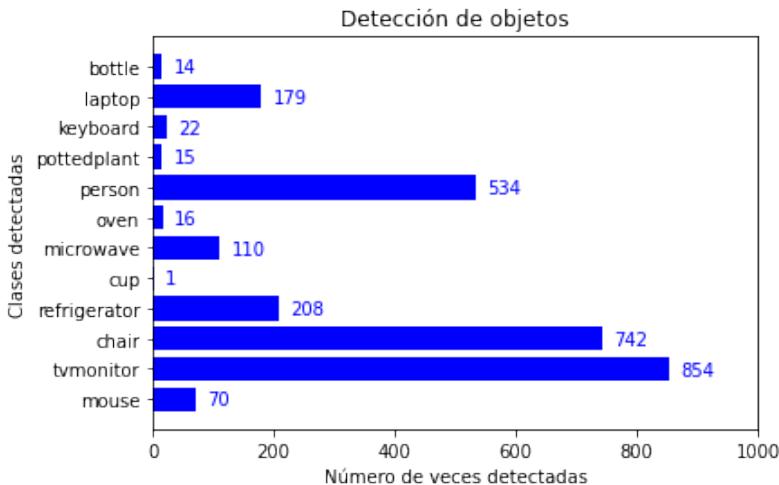


Figura 4.12: Gráfica de objetos detectados. Muestra el número de veces que fue detectado cada uno de los objetos.

En este laboratorio no hay ningún horno, por lo que la detección de esta clase es errónea. Como se muestra en la figura 4.13, ‘oven’ solo se detecta en dos instantes durante la trayectoria, haciendo referencia al mismo objeto. Revisando las imágenes

CAPÍTULO 4. RESULTADOS OBTENIDOS

del dataset, se puede ver lo que YOLO confunde con un horno, mostrado en la figura 4.14. Ciertamente, el objeto que detecta es muy parecido a los diales de control de un horno, por lo que es razonable la confusión.

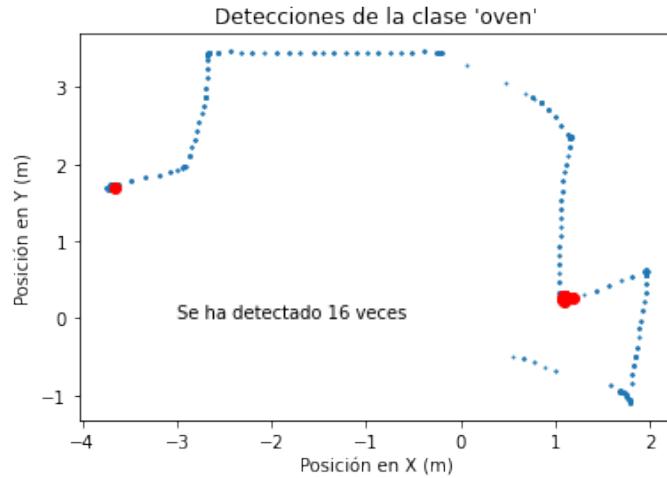


Figura 4.13: Momentos en los que se ha detectado la clase ‘oven’. Los puntos azules representan la trayectoria del robot, los rojos las veces que se detectó la clase.

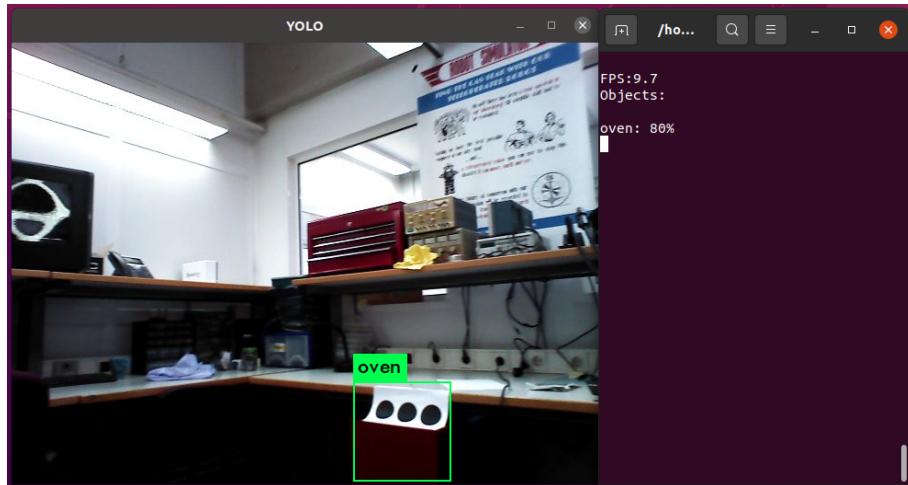


Figura 4.14: Detección de la clase ‘oven’ en la imagen.

En cambio, en el laboratorio sí que hay un frigorífico. Concretamente, un frigorífico compacto de una puerta. Sin embargo, pese a solamente haber uno, aparece en muchas ocasiones. Si obtenemos las posiciones en las que se detecta (figura 4.15), se puede ver como lo hace en posiciones muy distintas. Analizando las imágenes del dataset, se puede ver como, efectivamente, YOLO hace detecciones erróneas de la clase ‘refrigerator’. En la figura 4.16 se muestran dos detecciones erróneas y una

4.2. ANÁLISIS DE LA FIABILIDAD DE LAS DETECCIONES

correcta. Se puede ver como, en las detecciones incorrectas, los objetos detectados son similares a un frigorífico convencional.

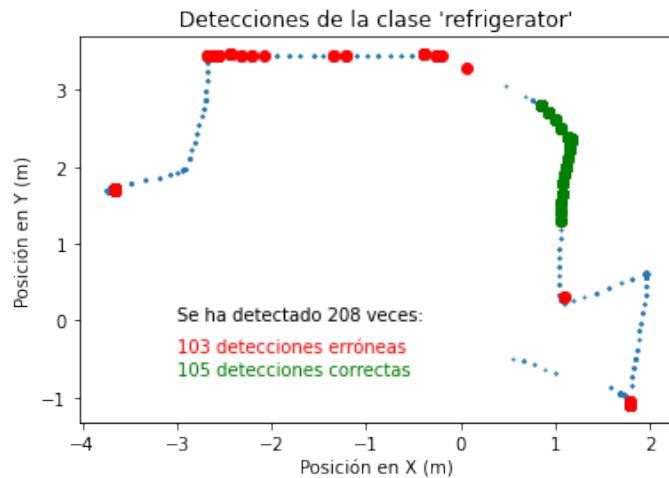
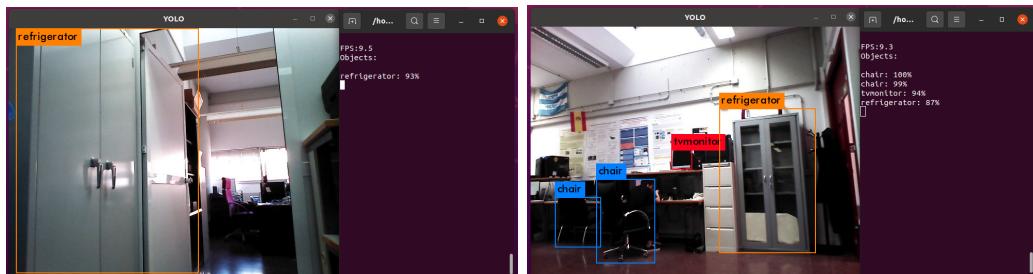
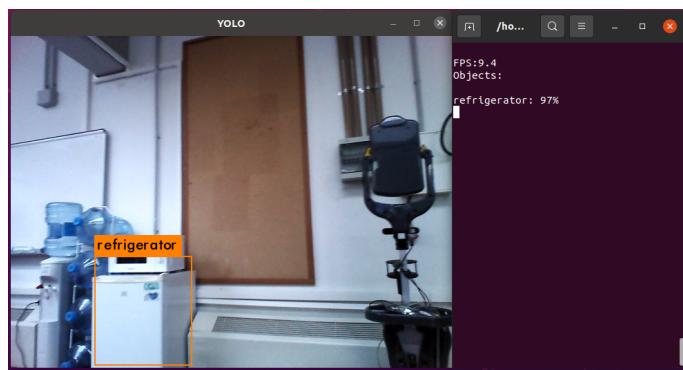


Figura 4.15: Momentos en los que se ha detectado la clase ‘refrigerator’. Los puntos azules representan la trayectoria del robot, los rojos las veces que se detectó la clase de forma incorrecta y los verdes las veces que los detectó correctamente.



(a) Detección errónea 1

(b) Detección errónea 2



(c) Detección correcta

Figura 4.16: Detecciones de la clase ‘refrigerator’ en la imagen.

Parte III

Conclusiones

Capítulo 5

Conclusiones, Propuestas y Líneas Futuras

En este capítulo se pretende, a partir de los resultados obtenidos y de todo el proceso llevado a cabo para conseguirlos, exponer las conclusiones a las que se han llegado. Además se expondrán algunas propuestas para mejorar el procedimiento. Finalmente se trazarán posibles líneas de trabajo de cara al desarrollo futuro de este proyecto.

5.1. Conclusiones

Los láseres 2D son unos sensores que se caracterizan principalmente por su gran precisión y amplio campo de visión. Sin embargo, son elementos que encarecen en gran medida el presupuesto a dedicar para el diseño de un robot. Las cámaras RGBD, en cambio, son sensores mucho más asequibles, por lo que se deberían de tener en cuenta a la hora de generar *floorplans* de la estructura del edificio o vivienda.

Como se ha mostrado en el capítulo 4, el resultado obtenido es ciertamente satisfactorio. El mapa obtenido con la cámara consigue eliminar los elementos no estructurales generando un *floorplan* acorde a la estructura fija del entorno. Sin embargo, tiene algunos problemas, principalmente influenciados por una mala calibración intrínseca de la cámara, perjudicando la rectitud de las líneas. Estos problemas también se deben a que los sensores RGBD son mucho menos precisos que los láseres y empeoran aún más cuando aumentan las distancias. Además, el entorno guardado en el dataset utilizado para comprobar el funcionamiento (el laboratorio) no es el más amigable para este procedimiento debido a la gran cantidad de elementos que perjudican la medición a través de la cámara. El resultado deja entrever que, con una correcta configuración de los parámetros de la cámara y unas mejores condiciones del entorno, sería posible obtener un *floorplan* fiel a la estructura fija del entorno.

CAPÍTULO 5. CONCLUSIONES, PROPUESTAS Y LÍNEAS FUTURAS

Las conclusiones extraídas del resultado obtenido tras la implementación de este proyecto son las siguientes:

- Un láser 2D no permite obtener un *floorplan* estructural del entorno inmune a elementos que pueden variar su posición. Los sensores RGBD son una solución frente a este problema, permitiendo generar mapas sin elementos cambiantes como personas, muebles, etc.
- Las cámaras RGBD son mucho menos precisas que los láseres 2D, de ahí su gran diferencia de precio en el mercado. Sin embargo, permiten obtener información 3D del entorno, algo que los láseres no pueden conseguir y puede resultar muy útil en determinadas situaciones.
- El entorno es uno de los factores más determinantes de cara a la efectividad del método. Un entorno con muchos elementos transparentes (como ventanas) o con entrantes y salientes perjudican en gran medida los resultados obtenidos.
- Una mala calibración de la cámara contribuye muy negativamente a la hora de obtener resultados fieles a la realidad debido a las distorsiones que se generan en sus mediciones. Un correcto proceso de calibración del sensor generaría mejores resultados.
- La red neuronal YOLO, entrenada para la detección de objetos, cumple su función. Sin embargo, muchos de los elementos del dataset no ha sido capaz de detectarlos e incluso ha detectado objetos erróneamente por lo que no hay que confiar completamente en sus mediciones.

Además, el desarrollo de este proyecto ha requerido experimentar con herramientas con las que yo, autor de este trabajo, tenía poca o nula experiencia. Personalmente me he visto sorprendido con los siguientes hechos:

- La comunidad de ROS es una de las más activas que se pueden encontrar. Existen miles de usuarios que aportan paquetes con funciones que abarcan prácticamente todos los ámbitos de la robótica. Por lo tanto, antes de comenzar a desarrollar alguna función, mejor comprobar si alguien ya lo ha hecho, algo que, vista la inmensidad de su comunidad, será lo más probable. Además, ROS cuenta con su propio foro donde compartir problemas y sus soluciones, ayudando en gran medida a usuarios novatos.
- El manejo de datos con *Python* y *Pandas* resulta realmente sencillo. Complementándolo con la librería *Matplotlib* se convierte en una poderosa herramienta para el análisis y la representación de los datos.
- *Linux*, pese a resultar inicialmente más complicado de comprender que otros sistemas operativos como *Windows* o *MacOS*, tiene numerosas características que lo convierten en una herramienta muy potente para el desarrollo, como la instalación por paquetes o el debugging.

5.2. Propuestas y Líneas Futuras

Hay aspectos de este trabajo que, una vez analizado de principio a fin, pudieran haberse hecho de otra forma y así, quizás, haber obtenido mejor resultado.

El dataset (*rosbag* del laboratorio) no es el idóneo para esta tarea. Debido a todos sus entrantes y salientes y a las ventanas en la pared divisoria entre las dos estancias, el resultado obtenido no es del todo correcto. Lo mejor hubiera sido grabar al robot en un ambiente de vivienda convencional, con habitaciones, baños y cocinas. Además, la posición y orientación de la cámara no es la idónea. Dado que los elementos se encuentran mayoritariamente en la parte inferior de las paredes, lo ideal sería situar la cámara en una posición más alta que la del dataset, con una orientación paralela al plano del suelo.

YOLO utiliza una serie de pesos para detectar los objetos. Una interesante línea futura para este proyecto sería investigar los distintos datasets de pesos disponibles para YOLO y cuáles convendría utilizar en un entorno de laboratorio de ingenierías como con el que se ha trabajado.

Dado que el principal problema de los resultados obtenidos son las distorsiones surgida en las mediciones, sería conveniente buscar los parámetros que se ajustan concretamente a la cámara utilizada para generar el dataset, es decir, realizar el proceso de calibración de la cámara antes de implementar este trabajo. Posiblemente, una correcta configuración de los parámetros intrínsecos de la cámara, generará una mejoría en los resultados.

Parte IV

Apéndices

Apéndice A

Otras Herramientas Utilizadas

En este apéndice se expondrán las herramientas que se han utilizado para desarrollar el trabajo que no se explican en los capítulos anteriores pero que han sido necesarias poder alcanzar el objetivo. Principalmente se explicarán, en pocas palabras, las librerías que se han utilizado en C++ y Python.

A.1. Librerías y Paquetes

A.1.1. *message_filters*

Es un paquete de ROS que proporciona una serie de filtros capaces de recibir mensajes y enviarlos pasado un tiempo determinado en función de la necesidad dictaminada por el filtro. Está diseñado por Josh Faust, Vijay Pradeep y Dirk Thomas. [36]

Para este proyecto se ha utilizado el filtro de sincronización o *time synchronizer*. Este filtro se encarga de recibir mensajes de distintos topics y enviarlos sólo cuando se haya recibido un mensaje de cada uno de los topics con la misma marca temporal (*timestamp*). Concretamente se utiliza en el nodo `sync_info`, para sincronizar los mensajes de la imagen de profundidad y de los parámetros de la cámara.

A.1.2. *fstream*

Es una librería de C++ que permite la escritura en archivos. Se ha utilizado en el nodo `write_objects` para escribir la información recibida de la *pose* y de la detección de objetos. [37]

A.1.3. *libfreenect*

libfreenect es una librería que permite el acceso USB a la cámara Kinect de Microsoft. Es parte del proyecto *OpenKinect*, una comunidad abierta que trabaja para implementar la posibilidad de uso de este dispositivo en todos los ordenadores. Concretamente, se ha utilizado el paquete `freenect_launch` que permite lanzar la cámara desde ROS. [38]

Este paquete sirvió para tener una primera toma de contacto con los sensores RGBD, las imágenes de profundidad y las nubes de puntos.

A.2. RVIZ

RVIZ (ROS Visualization) es un visualizador 3D para ROS. Es un software que se instala junto con el framework. Su uso está prácticamente estandarizado en toda la comunidad de ROS.

Se utiliza para visualizar los mensajes que llegan a los topics. Permite visualizar poses, imágenes (en diferentes codificaciones), mapas de ocupación, orientaciones, etc. e incluso permite incluir modelos 3D personalizados.

A.3. Google Colab

Colab, también conocido como *Colaboratory*, es una herramienta creada por Google que permite a los usuarios programar y ejecutar código Python en el navegador, utilizando servicios alojados en la nube. Esta herramienta se ha utilizado para el análisis de los objetos detectados con la red neuronal. Las ventajas de utilizar esta plataforma son: no requiere configuración, da acceso gratuito a GPUs y permite compartir contenido fácilmente.

A.4. 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 [39].

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.

A.5. LaTex y TexMaker

L^AT_EX es un sistema de composición de textos, orientado a la creación de documentos escritos de alta calidad tipográfica. Por sus características, es ampliamente utilizado para la generación de documentos científicos. Este software se ha utilizado con la distro *TexLive* y el software *TexMaker* para la generación de este documento.

A.6. Draw.io

Draw.io es un recurso web gratuito que sirve para realizar grafos. Todos los diagramas de este trabajo se han creado con esta herramienta. Se ha utilizado por su facilidad de uso y por la oportunidad de hacerlo online sin necesidad de instalar ningún software adicional.

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] Sebastian Thrun, Wolfram Burgard y Dieter Fox. *Probabilistic Robotics*. MIT Press, 2010.
- [6] EMT. *Metro de Málaga*. URL: <https://metromalaga.es/>. (acc: 25.05.2022).
- [7] Universidad de Stanford. *STAIR*. URL: <http://stair.stanford.edu/index.php>. (acc: 25.05.2022).
- [8] Morgan Quigley, Eric Berger y Andrew Y. Ng. “STAIR: hardware and software architecture”. En: *In AAAI 2007 Robotics Workshop*. 2007.
- [9] Wikipedia. *Licencia permisiva*. Marzo de 2022. URL: https://es.wikipedia.org/wiki/Licencia_permisiva. (acc: 25.05.2022).
- [10] Github. *Maximum subscribers of a topic*. URL: https://github.com/ros2/rmw%5C_fastrtps/issues/249. (acc: 25.05.2022).
- [11] *ROS WIKI*. URL: <http://wiki.ros.org/>. (acc: 28.05.2022).
- [12] 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.
- [13] *Microsoft Kinect V1*. URL: <https://www.microsoft.com/en-us/download/details.aspx?id=40278>. (acc: 02.06.2022).

BIBLIOGRAFÍA

- [14] *Asus Xtion depth image*. URL: <https://answers.ros.org/question/212747/cvbridge-conversion-problem-asus-xtion-depth-image-to-opencv/>. (acc: 28.05.2022).
- [15] Radu Bogdan Rusu y Steve Cousins. “3D is here: Point Cloud Library (PCL)”. En:
- [16] Joseph Redmon y col. “You Only Look Once: Unified, Real-Time Object Detection”. En: 2015. eprint: arXiv:1506.02640.
- [17] *El modelo de redes neuronales*. Agosto de 2021. URL: <https://www.ibm.com/docs/es/spss-modeler/SaaS?topic=networks-neural-model>. (acc: 28.05.2022).
- [18] Wikipedia. *Red neuronal prealimentada*. Diciembre de 2021. URL: https://es.wikipedia.org/wiki/Red_neuronal_prealimentada. (acc: 28.05.2022).
- [19] Wikipedia. *Perceptrón multicapa*. Enero de 2022. URL: https://es.wikipedia.org/wiki/Perceptr%C3%B3n_multicapa. (acc: 28.05.2022).
- [20] Andrés Díaz-Pinto. *¿Qué son las Redes Neuronales Convolucionales?* Enero de 2022. URL: <https://keepcoding.io/blog/redes-neuronales-convolucionales/>. (acc: 28.05.2022).
- [21] Amir Ali. *Convolutional Neural Network (CNN) with Practical Implementation*. Mayo de 2019. URL: <https://medium.com/machine-learning-researcher/convlutional-neural-network-cnn-2fc4faa7bb63>. (acc: 31.05.2022).
- [22] Pandas. URL: <https://pandas.pydata.org/>. (acc: 28.05.2022).
- [23] Matplotlib. URL: <https://matplotlib.org/>. (acc: 28.05.2022).
- [24] Chad Rockey. *ROS WIKI: depthimage to laserscan*. URL: http://wiki.ros.org/depthimage_to_laserscan. (acc: 26.05.2022).
- [25] Brian Gerkey. *ROS WIKI: gmapping*. URL: <http://wiki.ros.org/gmapping?distro=noetic>. (acc: 26.05.2022).
- [26] Patrick Mihelich. *ROS WIKI: image transport*. URL: http://wiki.ros.org/image_transport. (acc: 26.05.2022).
- [27] Marko Bjelonic. *ROS WIKI: darknet ros*. URL: http://wiki.ros.org/darknet_ros. (acc: 26.05.2022).
- [28] 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).
- [29] Brian Gerkey y Tony Pratkanis. *ROS WIKI: map server*. URL: http://wiki.ros.org/map_server. (acc: 29.05.2022).
- [30] M. Luperto y col. “Integrating Social Assistive Robots, IoT, Virtual Communities and Smart Objects to Assist at-Home Independently Living Elders: the MoveCare Project.” En: 2022.

BIBLIOGRAFÍA

- [31] David Zuñiga-Noël, Jose-Raul Ruiz-Sarmiento y Javier Gonzalez-Jimenez. “Intrinsic Calibration of Depth Cameras for Mobile Robots using a Radial Laser Scanner”. En: 2019.
- [32] A. Wilkowski y D. Mańkowski. “RGB-D and Lidar Calibration Supported by GPU”. En: *Computer Analysis of Images and Patterns. Lecture Notes in Computer Science. ICCVG 2020*. Vol. 12334. 2020, págs. 659-671. DOI: https://doi.org/10.1007/978-3-030-59006-2_19.
- [33] David Zúñiga Nöel y col. “Calibración extrínseca de un conjunto de cámaras RGB-D sobre un robot móvil”. En: *Actas de las XXXVIII Jornadas de Automática*. 2017.
- [34] *Hokuyo URG-04LX-UG01*. URL: <https://www.hokuyo-aut.jp/search/single.php?serial=166#spec>. (acc: 2.06.2022).
- [35] *Orbbec Astra*. URL: <https://shop.orbbec3d.com/Astra>. (acc: 2.06.2022).
- [36] Josh Faust, Vijay Pradeep y Dirk Thomas. *ROS WIKI: message filters*. URL: http://wiki.ros.org/message_filters. (acc: 02.06.2022).
- [37] *fstream reference*. URL: <https://www.cplusplus.com/reference/fstream/fstream/>. (acc: 02.06.2022).
- [38] Hector Martin, Josh Blake y Kyle Machulis. *ROS WIKI: libfreenect*. URL: <http://wiki.ros.org/libfreenect>. (acc: 02.06.2022).
- [39] *Nvidia: CUDA*. URL: <https://developer.nvidia.com/cuda-zone>. (acc: 02.06.2022).