



Universidad
Rey Juan Carlos

MÁSTER EN VISIÓN ARTIFICIAL

Curso Académico 2021/2022

Trabajo Fin de Máster

ODOMETRÍA VISUAL TRIDIMENSIONAL EN LA PLATAFORMA EDUCATIVA UNIBOTICS

Autor/a : Pablo Asensio Martínez
Tutor/a : Dr. JoseMaria Cañas Plaza

Trabajo Fin de Máster

Título del Trabajo con Letras Capitales para Sustantivos y Adjetivos

Autor/a : Nombre del Alumno/a

Tutor/a : Dr. Nombre del profesor/a

La defensa del presente Proyecto Fin de Grado/Máster se realizó el día 3 de
de 20XX, siendo calificada por el siguiente tribunal:

Presidente:

Secretario:

Vocal:

y habiendo obtenido la siguiente calificación:

Calificación:

Móstoles/Fuenlabrada, a de de 20XX

*Aquí normalmente
se inserta una dedicatoria corta*

Agradecimientos

Aquí vienen los agradecimientos...

Hay más espacio para explayarse y explicar a quién agradeces su apoyo o ayuda para haber acabado el proyecto: familia, pareja, amigos, compañeros de clase...

También hay quien, en algunos casos, hasta agradecer a su tutor o tutores del proyecto la ayuda prestada...

AGRADECIMIENTOS

Resumen

Aquí viene un resumen del proyecto. Ha de constar de tres o cuatro párrafos, donde se presente de manera clara y concisa de qué va el proyecto. Han de quedar respondidas las siguientes preguntas:

- ¿De qué va este proyecto? ¿Cuál es su objetivo principal?
- ¿Cómo se ha realizado? ¿Qué tecnologías están involucradas?
- ¿En qué contexto se ha realizado el proyecto? ¿Es un proyecto dentro de un marco general?

Lo mejor es escribir el resumen al final.

Summary

Here comes a translation of the “Resumen” into English. Please, double check it for correct grammar and spelling. As it is the translation of the “Resumen”, which is supposed to be written at the end, this as well should be filled out just before submitting.

Índice general

1	Introducción	1
1.1	Visión Artificial	1
1.2	Autolocalización Visual	3
1.2.1	Structure from Motion (SfM)	5
1.2.2	Visual SLAM	6
1.2.3	Odometría Visual	7
1.3	Introducción a Unibotics	9
2	Objetivos	13
2.1	Planificación temporal (Falta)	13
2.2	Estructura de la memoria (Falta y cambiar a parrafo final del capitulo 1 Intro)	13
3	Estado del arte	15
3.1	Algoritmos	15
3.1.1	Ocho Puntos	17
3.1.2	Siete Puntos	17
3.1.3	RANSAC	18
3.1.4	LMEDS	18

3.2	Trabajos de Evaluación en Visual SLAM	19
3.2.1	SLAMTestbed	19
3.3	Trabajos de Formación en Robótica	21
3.3.1	TheConstruct	21
3.3.2	Riders.ai	22
4	Algoritmo de Odometría Visual 3D	23
4.1	Incorporación de código en la memoria	23
4.1.1	Pseudo códigos / codigos de backend	23
4.1.2	Pseudo códigos / codigos de frontend	23
4.2	Test y validación	23
4.2.1	Validación Experimental	23
5	Integración en Robotics Academy	25
6	Conclusiones y trabajos futuros	27
6.1	Consecución de objetivos	27
6.2	Trabajos futuros	27
	Referencias	29

Índice de figuras

1.1	Beta de la conducción autónoma total de Tesla	2
1.2	Control de frontera en el aeropuerto	3
1.3	Realidad aumentada del videojuego para móvil Pokemon Go	4
1.4	Structure from Motion: La sagrada familia	6
1.5	PTAM: Funcionamiento sobre un escritorio	8
1.6	Odometría visual 2D. En rojo, la verdad absoluta, frente al resultado de un algoritmo sobre un escenario del dataset de KITTI	9
1.7	Logo de Unibotics	10
1.8	Ejercicio <i>Color Filter</i> en la plataforma <i>Unibotics</i>	10
3.1	Relación entre los puntos de dos imágenes de una misma escena	16
3.2	Resultado de aplicar LMEDS para generar líneas epipolares	19
3.3	SLAMTestbed: Resultados de la estimación de un cambio de escala y traslación, rotación, offset y ruido gaussiano simultáneos.	20
3.4	TheConstruct	21
3.5	Riders.ai	22

ÍNDICE DE FIGURAS

Índice de fragmentos de código

ÍNDICE DE FRAGMENTOS DE CÓDIGO

Capítulo 1

Introducción

En este primer capítulo se propone dar una visión general del contexto en que se encuadra el proyecto fin de máster, que es la visión artificial en el ámbito de las plataformas educativas en línea. Dentro de ésta se abordará el problema al que vamos a hacer frente: creación de un ejercicio de autolocalización visual, o dicho de otro modo, la estimación de la posición y orientación 3D de una cámara haciendo uso de algoritmos de odometría visual. Concretamente se implementará un algoritmo de odometría visual 3D para la plataforma educativa Unibotics.

1.1 Visión Artificial

La visión artificial es un campo de la inteligencia artificial que pretende obtener información del mundo a partir de una o varias imágenes, que normalmente vienen dadas de forma de matriz numérica. La información relevante que se puede obtener a partir de las imágenes puede ser el reconocimiento de objetos, la recreación en 3D de la escena que se observa, el seguimiento de un objeto, etc.

El inicio de la visión artificial se produjo en 1961 por parte de Larry Roberts, quien creó un programa que podía ver una estructura de bloques, analizar su contenido y reproducirla desde otra perspectiva, utilizando para ello una cámara y procesando la imagen desde un ordenador. Sin embargo, para obtener el resultado las condiciones de la prueba estaban muy controladas. Otros muchos científicos también han tratado de solucionar el problema de conectar una cámara a un ordenador y hacer que este describa lo que ve. Finalmente, los científicos de la época se dieron cuenta de que esta tarea no era sencilla de realizar, por lo que se abrió un amplio campo de investigación, que tomó el nombre de visión artificial.

En este campo se persigue, por ejemplo, que el ordenador sea capaz de reconocer en

una imagen distintos objetos al igual que los humanos lo hacemos con nuestra visión. Se ha demostrado que este problema es muy complejo y que algo que para nosotros resulta automático puede que se tarde mucho tiempo en que lo resuelva, con la misma robustez y versatilidad, una máquina. Por otra parte, a pesar del alto precio computacional que se paga por utilizar cámaras como sensor, si se consigue analizar correctamente la imagen, es posible extraer *mucha* información de ella que no podría obtenerse con otro tipo de sensores.

En los años noventa empezaron a aparecer los primeros ordenadores capaces de procesar las imágenes lo suficientemente rápido. Además se comenzó a dividir los posibles problemas de la visión artificial en otros más específicos.

A continuación se presentan algunas aplicaciones y escenarios de aplicación, no solo en el área académica sino también en la vida cotidiana, que hacen uso de los descubrimientos y avances logrados por diversos investigadores en visión artificial.

- Robótica. Unos de los procesos más complejos que tiene que realizar un robot es interpretar el mundo a su alrededor a través de la información que adquiere mediante los sensores, como puede ser una cámara de vídeo. Esta información puede usarse para la localización o reconocimiento de objetos o incluso el seguimiento de los mismos. Los coches autónomos de Tesla o las aspiradoras robóticas de gama alta son ejemplos de robots que usan visión para comprender la escena, detectar objetos de interés o autolocalizarse.

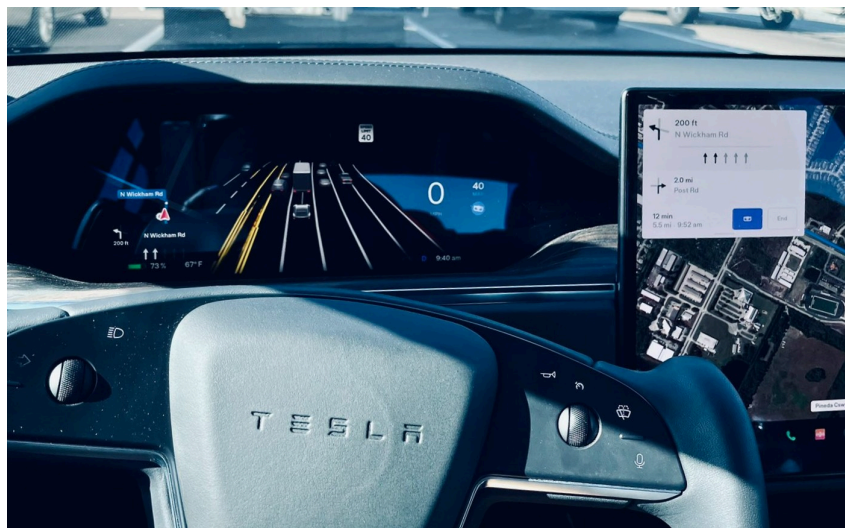


Figura 1.1: Beta de la conducción autónoma total de Tesla

- Videojuegos. Este sector ha contribuido notablemente al desarrollo de la visión artificial. Ejemplos claros de su uso son el dispositivo Kinect desarrollado por Microsoft (cámara RGBD), y Eye Toy, desarrollado por PlayStation para el reconocimiento de los gestos realizados por los jugadores. En la nueva generación de consolas de videojuegos se está avanzando en mejorar la interacción jugador-consola.

- Medicina. Uno de los objetivos de la visión artificial, en este contexto, es el tratamiento y análisis de imágenes, detección de patrones y reconstrucción 3D para ayudar al correcto diagnóstico por parte del especialista clínico. Aplicaciones comunes aquí son la detección y caracterización automática de tumores, mejora de la imagen de microscopio análisis hiperespectral para extraer la composición de los tejidos orgánicos contenidos en una imagen.
- Biometría. La biometría estudia los métodos automáticos para el reconocimiento único de humanos basado en uno o más rasgos conductuales o rasgos físicos intrínsecos. Dentro de estos sistemas se encuentran los sistemas de reconocimiento facial. Estos son muy usados en seguridad, como ejemplo de ello se encuentra el control de fronteras en aeropuertos, el sistema de vigilancia de la ciudad de Londres que realiza un reconocimiento facial mediante cámaras distribuidas por la ciudad; o el reconocimiento de caras y posterior emborronado automático de caras de Google Street View.



Figura 1.2: Control de frontera en el aeropuerto

1.2 Autolocalización Visual

Dentro de la visión artificial se encuentra el problema de la autolocalización visual que consiste en conocer la localización 3D de la cámara en todo momento solamente con las imágenes capturadas y sin disponer de ninguna información extra. Debido al gran abanico de posibilidades que abre resolver este problema, es uno de los retos más importantes dentro del campo de la robótica.

Esta técnica se plantea en los sistemas de navegación automáticos náuticos, terrestres y aéreos. Actualmente numerosas empresas están invirtiendo en este tipo de sistemas en el que apuestan por una navegación total o parcialmente autónoma, un ejemplo puede ser Roomba.

La autolocalización visual es una técnica que permite a aplicaciones de realidad aumentada, que es el término que se usa para definir una visión directa o indirecta de un entorno físico del mundo real, combinar el entorno real con elementos virtuales generados por ordenador para la creación de una realidad mixta en tiempo real. En la imagen 1.3 se puede observar un pokemon que parece que se encuentra en la calle, de frente de quien toma la fotografía, pero realmente ha sido el uso de la realidad aumentada lo que ha permitido una integración fidedigna del pokemon con el entorno en el contexto de la imagen.



Figura 1.3: Realidad aumentada del videojuego para móvil Pokemon Go

Las técnicas de autolocalización han suscitado gran interés por los investigadores en los últimos años. El problema ha sido abordado por dos comunidades distintas. Por un lado la de visión artificial que denominó al problema como *structure from motion* (SfM), donde la información es procesada por lotes, capaz de representar un objeto 2D a 3D con solo unas cuantas imágenes desde diferentes puntos de vista. Y por otro lado la comunidad robótica denominó al problema SLAM (*Simultaneous Localization and Mapping*) que trata de resolver el problema de una manera más ágil adaptando el funcionamiento de los sistemas en tiempo real.

Algunos de los conceptos más interesantes de conocer en la autolocalización visual son:

- Localización absoluta. Esta técnica pretende situar la cámara en una posición cuyo sistema de referencia sea común para todos. Podría usarse el sistema de referencia GPS, o el inicio de la escena, por ejemplo.
- Localización incremental. A diferencia de la localización absoluta, esta técnica hace

uso de un sistema de referencia horizonte local, siendo origen de referencia el *frame* inmediatamente anterior. Haciendo la suma de estos incrementos se podría calcular la posición absoluta.

- Localización desde un mapa conocido. Una técnica que permite, conociendo el escenario de trabajo, estimar de una mejor forma la posición de la cámara.
- Localización sin mapa. Aquella técnica de localización que no hace uso de información conocida previa del escenario.
- Error acumulativo. Es un tipo de error que está en la naturaleza de este problema, ya que al haber pequeños errores de cálculo numérico al computar las matrices de rotación en cada iteración, a la hora de actualizar la pose, se va acumulando este pequeño error.
- Cierre de bucle. Es un mecanismo por el cual se podría utilizar para reajustar los cálculos y reducir en mayor medida el error acumulativo. Por ejemplo, cuando se pasa por una zona ya conocida, poder comparar el resultado de la estimación 3D las dos veces, sabiendo que es la misma zona, y por ello deberían ser las mismas coordenadas.

1.2.1 Structure from Motion (SfM)

Dentro de la visión artificial el Structure from Motion (SfM) es la línea de investigación que toma como entrada únicamente un conjunto de imágenes, y pretende conocer de manera totalmente automática la estructura 3D de la escena vista y las ubicaciones de las cámaras desde donde las imágenes fueron captadas. El SfM es una de las áreas más atractivas de investigación en la última década. Ha llegado a un estado de madurez donde alguno de los algoritmos tienen aplicación comercial.

El SfM surge en la segunda mitad del siglo XIX, denominado fotogrametría, que tuvo como objetivo extraer información geométrica de las imágenes a partir de un conjunto de características manualmente identificadas por el usuario. La fotogrametría hace uso de técnicas de optimización no lineales como el ajuste de haces para reducir al mínimo error de retroproyección. El problema abordado por la comunidad de visión artificial ha sido en su mayoría lograr la completa automatización del proceso. Esto provocó avances en tres aspectos: en primer lugar, restricciones impuestas al movimiento bajo el supuesto de rigidez de la escena; en segundo lugar, la detección de características y descriptores [Canny, 1986], [Harris, Stephens y col., 1988]; y por último, la eliminación de espúreos.

Dadas múltiples vistas, un punto tridimensional puede ser reconstruido mediante triangulación. Un prerequisite importante es determinar la calibración de la cámara. Se puede representar por una matriz de proyección. La teoría geométrica de SfM permite calcular las

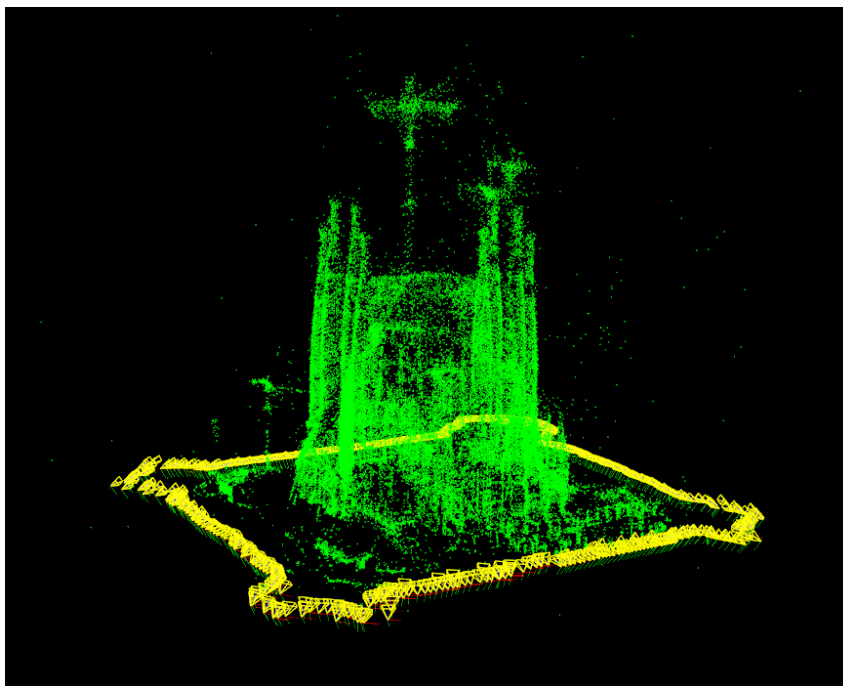


Figura 1.4: Structure from Motion: La sagrada familia

matrices de proyección y los puntos 3D utilizando solo correspondencia entre los puntos de cada imagen. Para mejorar el rendimiento del SfM se puede aprovechar el conocimiento sobre la escena, con el fin de reducir el número de grados de libertad. Por ejemplo, las restricciones que imponen el paralelismo y la coplanaridad, pueden utilizarse para reconstruir formas geométricas simples como líneas y polígonos planos, a partir de sus posiciones proyectadas en vistas simples.

1.2.2 Visual SLAM

La cuestión conocida como Simultaneous Localization and Mapping (SLAM) busca resolver los problemas que plantea colocar un robot móvil en un entorno y una posición desconocidas, y que él mismo se encuentre, sea capaz de construir incrementalmente un mapa de su entorno consistente y a la vez utilizar dicho mapa para determinar su propia localización.

La solución a este problema junto con un mecanismo de navegación haría que el sistema se encuentre con la capacidad para saber a dónde desplazarse, ser capaz de encontrar obstáculos y reaccionar ante ellos de manera inteligente. Esto conseguiría hacer sistemas de robots completamente autónomos.

La resolución al problema SLAM visual ha suscitado un gran interés en el campo de la robótica y se han propuesto muchas técnicas y algoritmos para darle solución, como es el

caso del artículo de Durrant-Whyte y Bailey [4]. Y aunque algunas de ellas han obtenido buenos resultados, en la práctica siguen surgiendo problemas a la hora de buscar el método más rápido o el que genere un mejor resultado con menos índice de fallo. La búsqueda de algoritmos y métodos que resuelvan completamente estos problemas sigue siendo una tarea pendiente.

Uno de los trabajos más importantes en el ámbito es el de monoSLAM de Davison¹ (Andrew J. Davison y Stasse, 2007) [3] que propone resolver este problema con una única cámara RGB como sensor y realizar el mapeado y la localización simultáneamente. El algoritmo propuesto por Davison utiliza un filtro extendido de Kalman para estimar la posición y la orientación de la cámara, así como la posición de una serie de puntos en el espacio 3D. Para determinar la posición inicial de la cámara es necesario a priori dotar de información sobre la posición 3D de por lo menos 3 puntos. Después el algoritmo es capaz de situar la cámara en el espacio tridimensional y de generar nuevos puntos para crear el mapa y servir como apoyo a la propia localización de la cámara. En la Figura 1.8 se pueden ver unas capturas de pantalla sobre uno de los experimentos realizados.

Es importante destacar también la trascendencia que ha tenido el trabajo PTAM (Klein y Murray, 2007) [8] que viene a solucionar uno de los principales problemas que tienen los algoritmos monoSLAM; el tiempo de cómputo, ya que aumenta exponencialmente con el número de puntos (Figura 1.9). Para ello se aborda el problema separando el mapeado de la localización, de tal modo que solo la localización deba funcionar en tiempo real, dejando así que el mapeado trabaje de una manera asíncrona en segundo plano. Este algoritmo parte de la idea de que solo la localización es necesaria que funcione en tiempo real. PTAM hace uso de *keyframes*, es decir, fotogramas clave que se utilizan tanto para la localización como para el mapeado y también de una técnica de optimización mediante ajuste de haces, como en SfM.

1.2.3 Odometría Visual

Dentro de las familias de técnicas pertenecientes a Visual SLAM se encuentra la odometría visual, que es una parte que abordaremos en este trabajo. Consiste en la estimación del movimiento 3D *incremental* de la cámara en tiempo real. Es decir, el cálculo de la rotación y traslación tridimensionales de la cámara a partir de imágenes consecutivas. Se trata de una técnica incremental ya que se basa en la posición anterior para calcular la nueva.

En este tipo de algoritmos se suelen utilizar técnicas de extracción de puntos de interés, cálculos de descriptores y algoritmos para el emparejamiento. Normalmente el proceso es: una vez calculados los puntos emparejados se calcula la matriz fundamental o esencial y descomponerlas mediante SVD para obtener la matriz de rotación y traslación (RT). Una

¹<http://www.doc.ic.ac.uk/~ajd/>

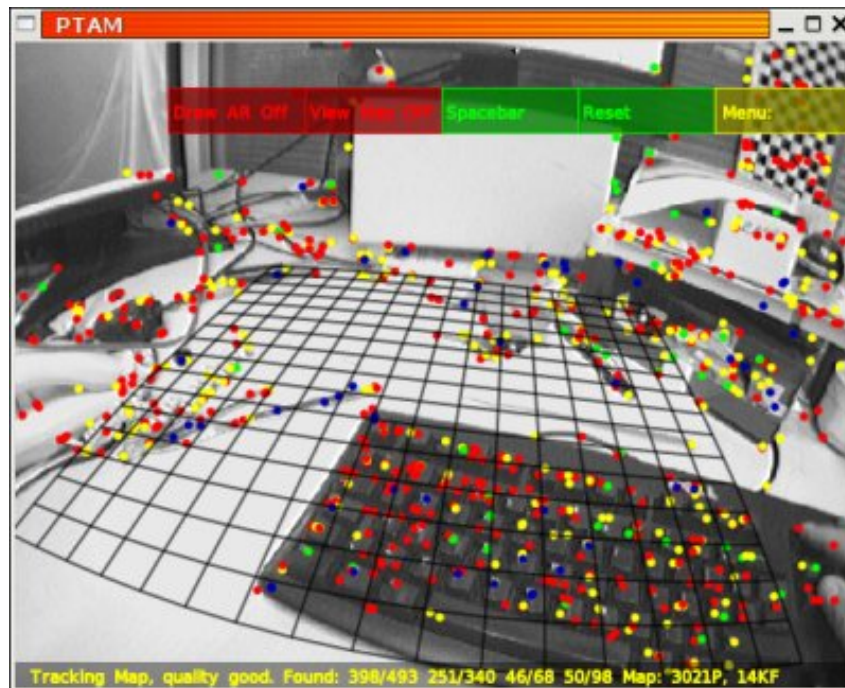


Figura 1.5: PTAM: Funcionamiento sobre un escritorio

vez que se conocen estas matrices de rotación y traslación, se puede calcular mediante un algoritmo, la posición que ocupa nuestra cámara en el espacio.

Dentro de la odometría visual podemos diferenciar diferentes tipos:

- Monocular y estéreo. Según la configuración de la cámara, la odometría visual se puede clasificar como odometría visual monocular (cámara única), odometría visual estéreo (dos cámaras en configuración estéreo).
- Método directo y basado en características. La información visual tradicional de la odometría visual se obtiene mediante el *método basado en características*, que extrae los puntos característicos de la imagen y los rastrea en la secuencia de imágenes. Los desarrollos recientes en la investigación en este campo proporcionaron una alternativa, llamada *método directo*, que utiliza la intensidad de píxeles en la secuencia de imágenes directamente como entrada visual. También hay métodos híbridos.
- Odometría inercial visual. Si se utiliza una unidad de medida inercial (IMU) dentro del sistema, esto denomina comúnmente *odometría inercial visual* (VIO).

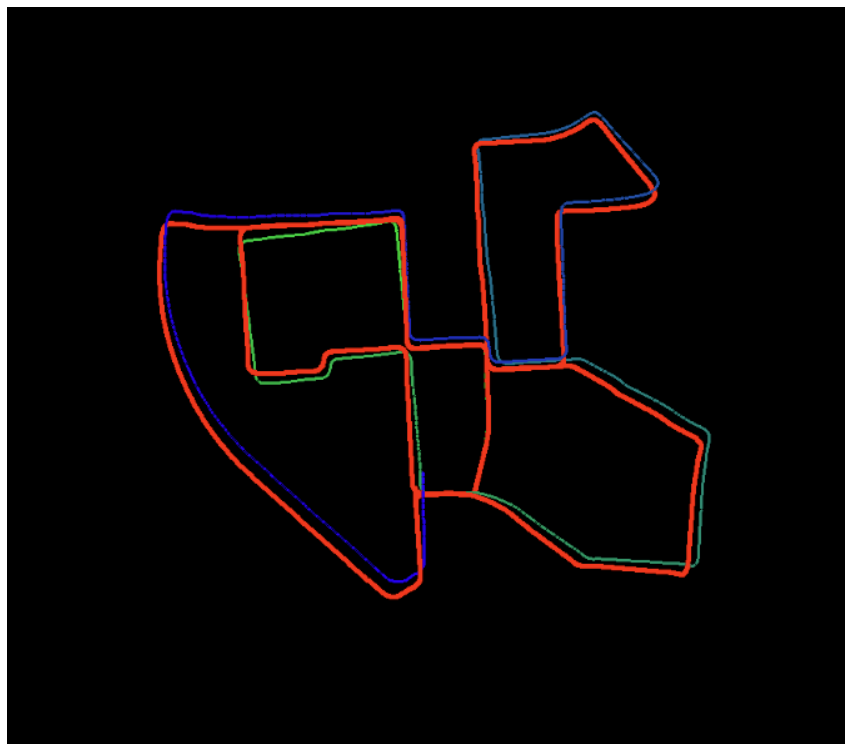


Figura 1.6: Odometría visual 2D. En rojo, la verdad absoluta, frente al resultado de un algoritmo sobre un escenario del dataset de KITTI

1.3 Introducción a Unibotics

En los últimos años el uso de plataformas web educativas ha ido incrementando debido a la pandemia del COVID-19, que ha sido un factor por el cual tanto el trabajo como en estudio a distancia ha aumentado considerablemente. Y, como consecuencia directa el uso de estas plataformas *online* es más demandado. En el área robótica esto no es una excepción. Se está haciendo especial hincapié en el desarrollo de plataformas online que permitan el aprendizaje sobre la programación de robots. Y aunque Unibotics no nació a causa de la pandemia, sí es una de las pioneras.

Unibotics nace como extensión natural de *Robotics Academy*[2]. Es un entorno docente de robótica universitaria. Este entorno tiene una orientación muy práctica en cuanto al aprendizaje de la programación de la inteligencia de los robots, ya que se utiliza como editor del código fuente el navegador web, el cual también es el interfaz gráfico de la ejecución. La plataforma posee una gran colección de ejercicios muy variados que abarcan muchas de las aplicaciones robóticas que han surgido recientemente: drones, robots aspiradores, coches autónomos, asistente de aparcamiento, control de robots móviles, etc. La diferencia entre *Robotics Academy*² y *Unibotics*³ es la forma en la que se ejecutan los ejercicios. La

²<http://jderobot.github.io/RoboticsAcademy/>

³<https://unibotics.org/>

primera es completamente en local, mientras la segunda hace uso de la red para completar los ejercicios.



Figura 1.7: Logo de Unibotics

Toda esta colección de ejercicios son de código abierto lo que proporciona la posibilidad de compartir, modificar y estudiar el código fuente, además de colaborar entre usuarios. Unibotics web es un entorno multiplataforma pudiendo implementarse en sistemas operativos tales como Linux, Windows y MacOS. La plataforma hace uso del simulador Gazebo y el lenguaje principal con el que se programa es interpretado y multiplataforma, Python.

Algunos de los ejercicios que podemos encontrar en Unibotics relacionado con visión artificial son los siguientes:

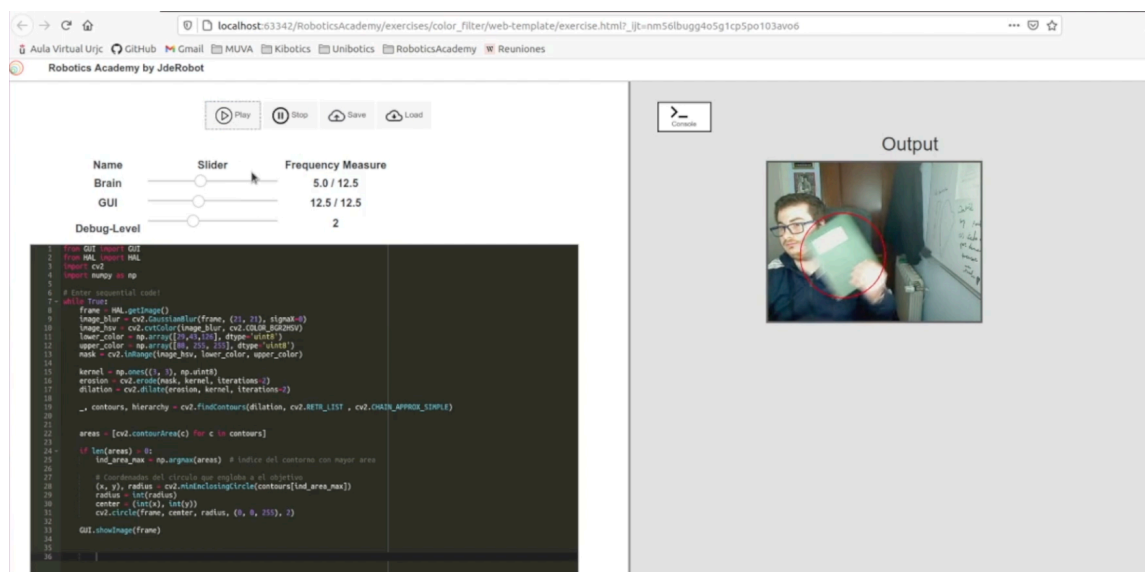


Figura 1.8: Ejercicio *Color Filter* en la plataforma *Unibotics*

- Follow line. El objetivo de *Follow line* es realizar un control reactivo PID capaz de seguir la línea pintada en el circuito de carreras. Utilizando como sensor exclusivamente la cámara. [Pincha aquí para ver el vídeo.](#)
- 3D Reconstruction. En este ejercicio se busca reconstruir una escena 3D a partir de un par estéreo.

- Color Filter. *Color Filter* es uno de los últimos ejercicios añadidos a la plataforma. Persigue el desarrollo un filtro de color para segmentar algún objeto en la imagen y rastrearlo. [Pincha aquí para ver el vídeo.](#)
- Optical Flow Teleop. En este ejercicio se pretende desarrollar un algoritmo de flujo óptico para teleoperar el robot utilizando las imágenes obtenidas de una cámara web. [Pincha aquí para ver el vídeo.](#)

Capítulo 2

Objetivos

2.1 Planificación temporal (Falta)

Es conveniente que incluyas una descripción de lo que te ha llevado realizar el trabajo. Hay gente que añade un diagrama de GANTT. Lo importante es que quede claro cuánto tiempo has consumido en realizar el TFG/TFM (tiempo natural, p.ej., 6 meses) y a qué nivel de esfuerzo (p.ej., principalmente los fines de semana).

2.2 Estructura de la memoria (Falta y cambiar a parrafo final del capitulo 1 Intro)

Por último, en esta sección se introduce a alto nivel la organización del resto del documento y qué contenidos se van a encontrar en cada capítulo.

- En el primer capítulo se hace una breve introducción al proyecto, se describen los objetivos del mismo y se refleja la planificación temporal.
- En el siguiente capítulo se describen el estado del arte así como las tecnologías utilizadas en el desarrollo de este TFM (Capítulo 3).
- En el capítulo ?? Se describe la arquitectura y el proceso de desarrollo de la herramienta.
- En el capítulo ?? Se presentan las principales pruebas realizadas para validación del ejercicio, incluyendo resultados de los experimentales efectuados.

- Por último, se presentan las conclusiones del proyecto así como los trabajos futuros que podrían derivarse de éste (Capítulo 6).

Capítulo 3

Estado del arte

En este capítulo se explicarán varios de los algoritmos de odometría visual, así como algoritmos para la evaluación de estos mismos, además de una serie de sitios web con carácter educativo sobre robótica, campo donde se usa la visión artificial habitualmente.

3.1 Algoritmos

A continuación se explicarán una serie de algoritmos utilizados en odometría visual para el cálculo de la *matriz fundamental*, F , [Luong y col., 1993], que es una matriz 3×3 que describe la relación entre los puntos de dos imágenes de una misma escena. Pero antes se explicará brevemente este problema.

$$F = \begin{pmatrix} f_{0,0} & f_{0,1} & f_{0,2} \\ f_{1,0} & f_{1,1} & f_{1,2} \\ f_{2,0} & f_{2,1} & f_{2,2} \end{pmatrix} \quad (3.1)$$

Siendo x un punto de la primera imagen, x' el mismo punto en la segunda imagen, F la matriz fundamental y lx la línea epipolar (ver Figura 3.1) que contiene el punto en la segunda imagen, se cumplen las siguientes ecuaciones.

$$lx'_i = Fx_i \quad (3.2)$$

$$lx_i = F^T x'_i \quad (3.3)$$

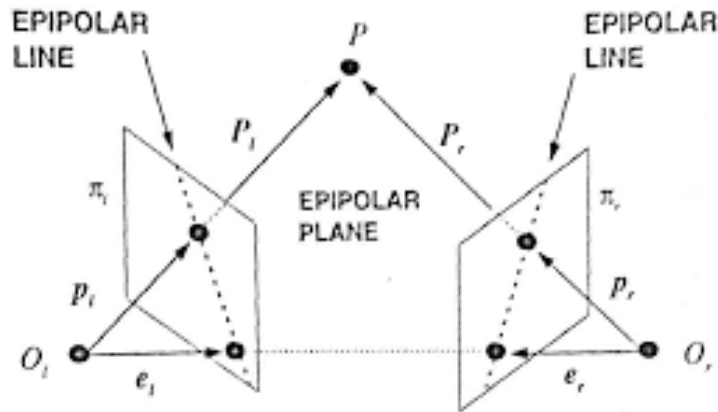


Figura 3.1: Relación entre los puntos de dos imágenes de una misma escena

Las características más importantes de la matriz fundamental son las siguientes [Faugeras, 1992] :

- La multiplicación del punto en la segunda imagen traspuesto por la matriz fundamental y por el punto en la primera imagen da como resultado 0.

$$x_i' F x_i = 0 \quad (3.4)$$

- El rango de la matriz fundamental es 2, lo que significa que tiene siete grados de libertad, pudiendo calcular dos elementos de la matriz a partir de los otros.
- El determinante es 0.

$$\det(F) = 0 \quad (3.5)$$

- La matriz fundamental es homogénea lo que quiere decir que está definida hasta un factor de escala.

Para calcular la matriz fundamental lo primero que hay que hacer es relacionar de alguna forma los puntos de una imagen con los de la otra, es decir realizar un *emparejamiento*, típicamente utilizando el algoritmo de emparejamiento SURF.

Una vez que se tienen los puntos emparejados simplemente hay que desarrollar las ecuaciones y resolver el sistema. La ecuación desarrollada para un punto, siendo i este punto, es la siguiente:

$$x_i' x_i f_{0,0} + x_i' y_i f_{0,1} + x_i' f_{0,2} + y_i' x_i f_{1,0} + y_i' y_i f_{1,1} + y_i' f_{1,2} + x_i f_{2,0} + y_i f_{2,1} + f_{2,2} = 0 \quad (3.6)$$

El sistema se puede resolver de una forma mucho más sencilla si se expresa mediante matrices y se utiliza SVD (*Singular Value Decomposition* por sus siglas en inglés). De esta forma el sistema quedaría definido mediante la siguiente ecuación.

$$Af = 0 \quad (3.7)$$

siendo A una matriz de dimensiones $N \times 9$, cuyas filas se definen así:

$$(x'_i x_i, x'_i y_i, x'_i, y'_i x_i, y'_i y_i, y'_i, x_i, y_i, 1)$$

y f un vector columna definido por los coeficientes de F :

$$(f_{0,0}, f_{0,1}, f_{0,2}, f_{1,0}, f_{1,1}, f_{1,2}, f_{2,0}, f_{2,1}, f_{2,2})^T$$

Dadas las características de la matriz fundamental no es necesario desarrollar un sistema de nueve ecuaciones, sino que pueden seguirse algunas estrategias para utilizar menos puntos. A continuación se describen los dos algoritmos más utilizados para resolver el sistema de ecuaciones.

3.1.1 Ocho Puntos

Dado que la matriz F sólo está definida hasta un desconocido factor de escala, se impone una restricción adicional sobre la norma de F para eliminar el factor de escala, que consiste en que la norma cuadrática de F sea la unidad:

$$\|F\|^2 = \|f\|^2 = f^t f = 1 \quad (3.8)$$

Para que el sistema de ecuaciones (3.7) admita una solución diferente del vector nulo, la matriz A debería tener un rango máximo de 8. Por lo tanto, el sistema de ecuaciones (3.7) puede ser resuelto mediante la correspondencia entre al menos 8 puntos diferentes en ambas cámaras.

3.1.2 Siete Puntos

Este algoritmo utiliza la restricción del rango de la matriz fundamental. Dado que el rango de esta matriz es 2, se puede introducir una ecuación más que describe la relación entre dos filas de la matriz. De esta forma se pueden calcular dos términos a partir del resto, teniendo que introducir únicamente 7 puntos en las ecuaciones.

Además de estos algoritmos hay otros dos derivados de estos que utilizan sistemas de optimización para encontrar la matriz que mejor se ajusta a los puntos, eliminando a la vez los posibles outliers: RANSAC y LMedS.

3.1.3 RANSAC

RANSAC, (Fischler y Bolles, 1981)[6], es un algoritmo iterativo en el que en cada iteración se cogen 8 puntos y se calcula la matriz fundamental. Una vez calculada se comprueba qué tal se ajustan el resto de puntos a esta matriz. Si el ajuste es malo la matriz se desecha y si es buena se coge como mejor solución hasta encontrar una mejor o hasta que acabe el algoritmo. El pseudocódigo del método está representado en 1

Algorithm 1 Pseudocódigo de RANSAC

- 1: Seleccionar aleatoriamente el número mínimo de puntos necesarios para determinar los parámetros del modelo (*inliers*).
 - 2: Resolver para los parámetros del modelo.
 - 3: Determinar cuántos puntos del conjunto de todos los puntos encajan con una tolerancia predefinida E (*conjunto de consenso*).
 - 4: Si la fracción del número de valores *inliers* sobre el número total de puntos en el conjunto excede un umbral predefinido K, volver a estimar los parámetros del modelo usando todos los valores *inliers* identificados y terminar.
 - 5: De lo contrario, repita los pasos 1 a 4 (máximo de N veces).
-

3.1.4 LMEDS

Least Median of Squares (Rousseeuw, 1984)[12] es un algoritmo iterativo que trata de encontrar una matriz fundamental óptima al conjunto de datos dado. Para ello, intenta minimizar el error cuadrático medio entre el modelo (la matriz fundamental calculada en la iteración anterior) y los puntos, mediante el algoritmo de descenso del gradiente de tal forma que según pasan las iteraciones la matriz fundamental calculada se ajusta mejor a los puntos dados.

Este método es muy resistente a las coincidencias falsas, así como a los valores atípicos debido a una mala localización. Sin embargo, es muy difícil definir el algoritmo completo para la mínima mediana de los cuadrados como una fórmula matemática para un conjunto de datos completo, por lo que un algoritmo que genera una solución

$$\min_i \text{med } r_i^2 \quad (3.9)$$

de la mínima mediana de los cuadrados para un conjunto de datos que se da a continuación:

La estimación de la mínima mediana de los cuadrados es resistente a los valores atípicos debido a su alto valor de descomposición del 50%. Esta es la fracción de valores atípicos

Algorithm 2 Pseudocódigo de LMEDS

- 1: Se elije m conjuntos aleatorios de puntos con tamaño p del conjunto de datos donde p es el número de parámetros en la ecuación que se está resolviendo.
- 2: para cada subconjunto se utiliza un método como el de la mínima media de los cuadrados para encontrar una solución para los parámetros de ese conjunto de datos.
- 3: para cada conjunto de parámetros p obtenidos, se calcula la mediana de los cuadrados de los residuos M con respecto a todo el conjunto de datos:

$$M_J = \text{med}_{i=1, \dots, m} r_i^2(p_J, m_i) \quad (3.10)$$

- 4: La solución es p_J para la cual M_J es mínimo entre todos los m M_J .

que se pueden tolerar y al mismo tiempo arrojar una buena solución. Sin embargo, el alto valor de ruptura significa que LMEDS no se adapta bien al ruido gaussiano.



Figura 3.2: Resultado de aplicar LMEDS para generar líneas epipolares

3.2 Trabajos de Evaluación en Visual SLAM

3.2.1 SLAMTestbed

SlamTestbed [10] es una aplicación diseñada y creada para comparar cuantitativamente algoritmos SLAM. El diseño de esta herramienta, que por sencillez se tratará como una *caja negra* que cuenta en la entrada con dos secuencias de puntos 3D orientados. Uno de los datasets será la verdad absoluta, y el segundo dataset será la posición y orientación en 3D obtenidos tras aplicar un algoritmo Visual SLAM, correspondiendo cada registro del dataset con una posición de la cámara. Una vez procesados los dos datasets por SlamTestbed, se obtiene como salida las transformaciones estimadas por la herramientas entre la verdad absoluta y las posiciones y orientaciones calculadas por el algoritmo de SLAM. Además,

se mostrará un conjunto de estadísticas que miden el error cometido en las estimaciones de SLAM qué caracteriza la precisión de los algoritmos.

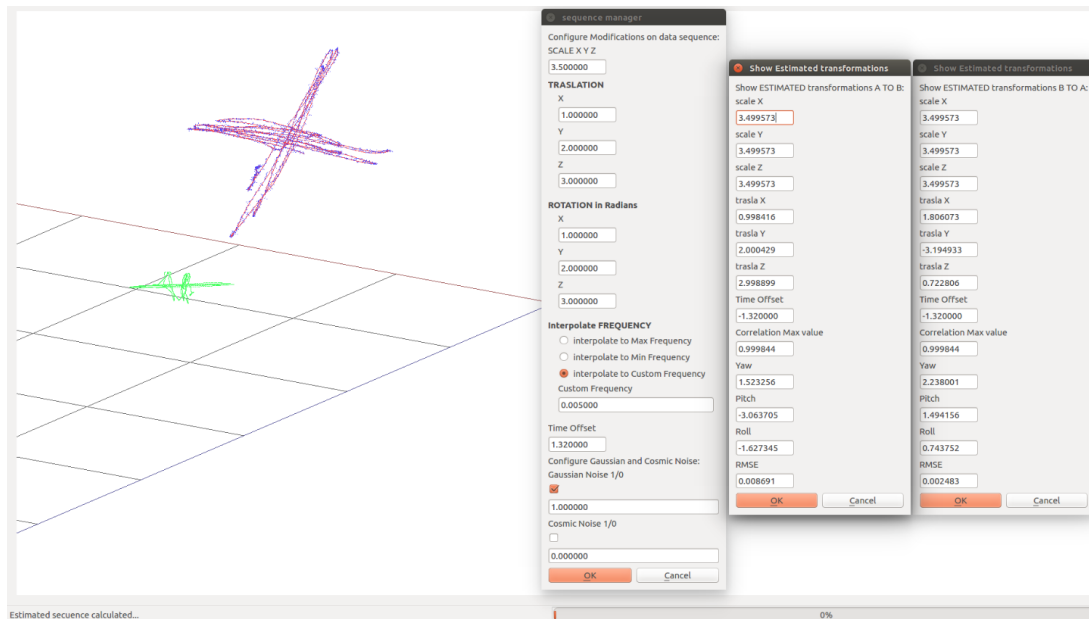


Figura 3.3: SLAMTestbed: Resultados de la estimación de un cambio de escala y traslación, rotación, offset y ruido gaussiano simultáneos.

El objetivo principal de la herramienta desarrollada es calcular el error existente entre una secuencia con las posiciones y orientaciones 3D verdaderas (dataset A) y la trayectoria calculada por el algoritmo de SLAM (dataset B). Para que esto sea posible necesitaremos antes eliminar algunas variables que no permiten comparar directamente las dos trayectorias, como son la escala, el offset temporal y la transformación en 3D entre ellas. Por ello, se necesita calcular un nuevo dataset (estimado), que sea comparable con el dataset A.

Las principales funcionalidades utilizadas para obtener el dataset estimado son:

- **Cálculo de PCA.** El análisis de componentes principales (o PCA), permite reducir los dos datasets a sus componentes principales, lo que posibilita estimar a continuación la escala y el offset existente entre ellos.
- **Estimación de escala.** Estima la diferencia de escala entre los dos datasets a partir de los datos proporcionados en el cálculo de componentes principales.
- **Estimación de offset temporal.** Con este módulo podremos hallar la diferencia entre marcas de tiempos de los 2 datasets, ya que pueden haber comenzado en periodos de tiempo distintos.
- **Interpolación para igualar frecuencias de muestreo.** Se pueden igualar en frecuencia los dos datasets, en caso de que estas sean distintas.

- **Operaciones de registro para estimar la Rotación y Traslación.** Permite estimar la traslación y rotación existentes entre el dataset A y el dataset B y llevarlas así al mismo sistema de referencia espacial, donde ya son directamente comparables.

3.3 Trabajos de Formación en Robótica

3.3.1 TheConstruct

TheConstruct ¹ es una plataforma web que enseña sobre robótica, ROS e inteligencia artificial. Tiene una versión gratuita en la que se ofrecen tres cursos: Linux para robótica, Python3 para robótica y C++ para robótica, si se quiere puedes acceder a todos los cursos con su versión de pago. Está desarrollado para que puedan utilizarlo tanto principiantes como profesionales. No requiere de la instalación de ROS. Además, una de sus ventajas es que tiene una gran comunidad donde se puede establecer contactos y aprender nuevas formas de programar robots. TheConstruct cuenta con robots reales que se pueden alquilar por un determinado tiempo, dando la posibilidad de poder conectarse a ellos y programarlos desde cualquier lugar. Una vez se selecciona un curso, se tiene en la parte izquierda la teoría para realizar el curso. También tiene un interfaz de usuario dónde se escribe el código, un terminal para escribir comandos y una simulación del robot que se va a programar.

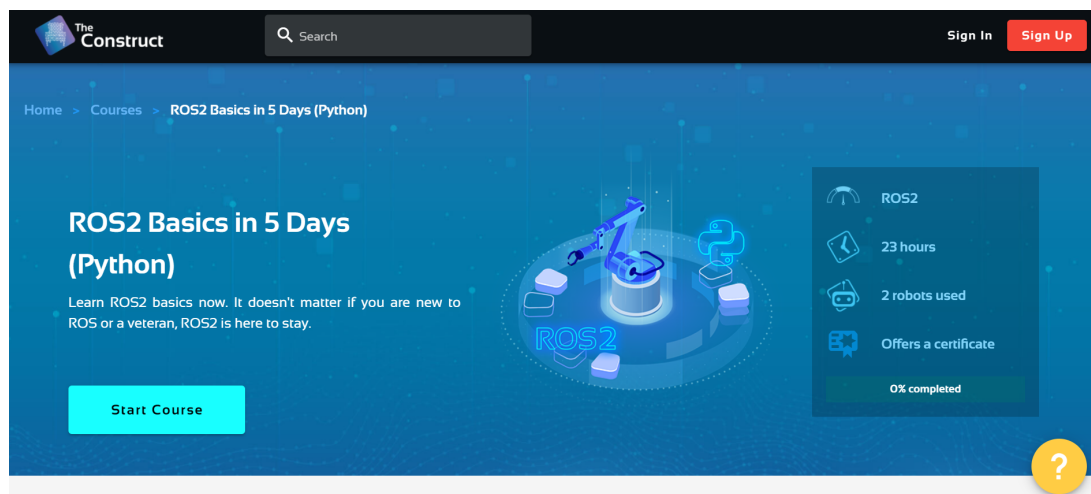


Figura 3.4: TheConstruct

¹<https://www.theconstructsim.com/>

3.3.2 Riders.ai

Riders.ai² es una plataforma sobre robótica de simulación, educación y competiciones basada en la nube, desarrollada por Acrome Robotic Systems. Es una plataforma de pago, solamente es gratis la primera lección del curso. Consta de otros dos cursos, los cuales son la continuación del curso mencionado anteriormente. Además, una vez finalizados los cursos se obtiene un certificado. Se enseña a programar en Python o C++ los robots. Las lecciones están formadas por la teoría, el interfaz de usuario y el simulador Gazebo, todo ello en la misma pestaña, como se muestra en la figura 3.5, Riders dispone de dos ligas para competir con otros programadores. Una liga consiste en programar drones voladores y otra en programar robots móviles.

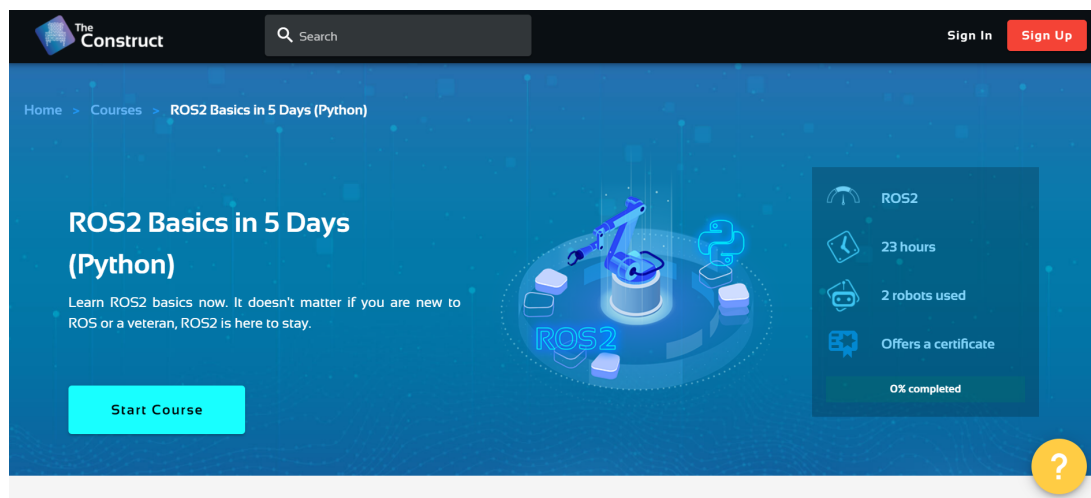


Figura 3.5: Riders.ai

²<https://riders.ai/>

Capítulo 4

Algoritmo de Odometría Visual 3D

<https://gsyc.urjc.es/jmplaza/students/tfm-visualodometry-isanroman-2015.pdf>

Atención: Este capítulo se introdujo como requisito en 2019.

Describe los experimentos y casos de test que tuviste que implementar para validar tus resultados. Incluye también los resultados de validación que permiten afirmar que tus resultados son correctos.

4.1 Incorporación de código en la memoria

4.1.1 Pseudo códigos / codigos de backend

4.1.2 Pseudo códigos / codigos de frontend

4.2 Test y validación

4.2.1 Validación Experimental

Capítulo 5

Integración en Robotics Academy

Capítulo 6

Conclusiones y trabajos futuros

6.1 Consecución de objetivos

Esta sección es la sección espejo de las dos primeras del capítulo de objetivos, donde se planteaba el objetivo general y se elaboraban los específicos.

Es aquí donde hay que debatir qué se ha conseguido y qué no. Cuando algo no se ha conseguido, se ha de justificar, en términos de qué problemas se han encontrado y qué medidas se han tomado para mitigar esos problemas.

Y si has llegado hasta aquí, siempre es bueno pasarle el corrector ortográfico, que las erratas quedan fatal en la memoria final. Para eso, en Linux tenemos `aspell`, que se ejecuta de la siguiente manera desde la línea de *shell*:

```
aspell --lang=es_ES -c memoria.tex
```

6.2 Trabajos futuros

Ningún proyecto ni software se termina, así que aquí vienen ideas y funcionalidades que estaría bien tener implementadas en el futuro.

Es un apartado que sirve para dar ideas de cara a futuros TFGs/TFMs.

Referencias

- [1] John Canny. «A computational approach to edge detection». En: *IEEE Transactions on pattern analysis and machine intelligence* 6 (1986), págs. 679-698.
- [2] José M Cañas y col. «A ROS-based open tool for intelligent robotics education». En: *Applied Sciences* 10.21 (2020), pág. 7419.
- [3] Andrew J Davison y col. «MonoSLAM: Real-time single camera SLAM». En: *IEEE transactions on pattern analysis and machine intelligence* 29.6 (2007), págs. 1052-1067.
- [4] Hugh Durrant-Whyte y Tim Bailey. «Simultaneous localization and mapping: part I». En: *IEEE robotics & automation magazine* 13.2 (2006), págs. 99-110.
- [5] Olivier D Faugeras. «What can be seen in three dimensions with an uncalibrated stereo rig?». En: *European conference on computer vision*. Springer. 1992, págs. 563-578.
- [6] Martin A Fischler y Robert C Bolles. «Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography». En: *Communications of the ACM* 24.6 (1981), págs. 381-395.
- [7] Chris Harris, Mike Stephens y col. «A combined corner and edge detector». En: *Alvey vision conference*. Vol. 15. 50. Citeseer. 1988, págs. 10-5244.
- [8] Georg Klein y David Murray. «Parallel tracking and mapping for small AR workspaces». En: *2007 6th IEEE and ACM international symposium on mixed and augmented reality*. IEEE. 2007, págs. 225-234.
- [9] Quang-Tuan Luong y col. «On determining the fundamental matrix: Analysis of different methods and experimental results». Tesis doct. Inria, 1993.
- [10] Elías Barcia Mejias. «Herramienta de evaluación cuantitativa de algoritmos Visual SLAM». https://gsyc.urjc.es/jmplaza/students/tfm-visualslam_slamtestbed-elias_barcia-2019.pdf. Tesis de mtría. Universidad Rey Juan Carlos, 2019.
- [11] Victor Arribas Raigadas. «Análisis de algoritmos de VisualSLAM: un entorno integral para su evaluación». https://gsyc.urjc.es/jmplaza/students/tfm-visualslam_evaluation-victor_arribas-2016.pdf. Tesis de mtría. Universidad Rey Juan Carlos, 2016.
- [12] Peter J Rousseeuw. «Least median of squares regression». En: *Journal of the American statistical association* 79.388 (1984), págs. 871-880.