



Universidad de Alcalá

Escuela Politécnica Superior
Máster Universitario en Ingeniería de
Telecomunicación

Tesis Fin de Máster

DESARROLLO DE SOLUCIONES PARA NAVEGACIÓN ROBÓTICA BASADAS EN SEGMENTACIÓN SEMÁNTICA Y APRENDIZAJE POR IMITACIÓN.

Autor: Rafael Flor Rodríguez-Rabadán.

Tutor: Roberto Javier López Sastre.

Cotutor: Carlos Gutiérrez Álvarez.

TRIBUNAL:

Presidente: D. Nombre y apellidos del presidente

Vocal 1º: D. Nombre y apellidos del vocal 1

Vocal 2º: D. Nombre y apellidos del vocal 2

FECHA: indicar la fecha de depósito

Desarrollo de soluciones para navegación robótica
basadas en segmentación semántica y aprendizaje por
imitación.

Rafael Flor Rodríguez-Rabadán

29 de octubre de 2024

Aquí escribes las dedicatorias. Por ejemplo: A mis padres.

Agradecimientos

Aquí debes incluir los agradecimientos.

Índice general

Agradecimientos	V
Resumen	XIII
Abstract	XV
Resumen Extendido	XVII
Glosario	XIX
1. Introducción	1
1.1. ¿Cómo escribir tu TFG/TFM en L ^A T _E X ?	1
1.1.1. La mejor manera para aprender a programar es programando . . .	1
1.1.2. Algunos detalles más	3
1.2. Recomendaciones importantes finales	3
1.3. Para terminar	4
A. CÓDIGO FUENTE.	5
A.1. Primera sección del apéndice.	5
A.1.1. Segunda sección del apéndice.	5
A.1.1.1. Tercera sección del apéndice.	5

Lista de figuras

1.1. El título corto de la gráfica.	3
---	---

Lista de tablas

1.1. El título corto de la tabla.	3
---	---

Resumen

Resumen de 100 palabras MÁXIMO

Palabras clave: cinco palabras como máximo, separadas por comas.

Abstract

Resumen de 100 palabras MÁXIMO - Es OBLIGATORIO

Keywords: cinco palabras como máximo, separadas por comas.

Resumen Extendido

Debe incluir un resumen del trabajo de un máximo de 5 páginas. Resaltar aspectos fundamentales del desarrollo, los resultados más relevantes y las conclusiones.

0.1. Glosario

AI	Artificial Intelligence
CNN	Convolutional Neural Networks
IL	Imitation Learning
ML	Machine Learning
LiDAR	Light Imaging Detection and Ranging
LLM	Large Language Models
RL	Reinforcement Learning
RNN	Recurrent Neural Networks
SLAM	Simultaneous Localization and Mapping
VSN	Visual Semantic Navigation

Capítulo 1

Introducción

La habilidad del ser humano para navegar en entornos desconocidos se fundamenta en su capacidad para identificar patrones comunes entre diversas estancias. Esto le permite, mediante un proceso de exploración, localizar su objetivo en un espacio nunca antes visitado.

Esta capacidad para recorrer entornos desconocidos ha demostrado ser de gran interés para su integración en plataformas robóticas reales, debido a su versatilidad tanto funcional como comercial. Históricamente, en el ámbito de la robótica, la navegación ha estado estrechamente vinculada a la sensorización y al uso de tecnologías como **LiDAR!** (**LiDAR!**), empleando técnicas como **SLAM!** (**SLAM!**). No obstante, estas aproximaciones presentan limitaciones a la hora de generalizar el problema de la navegación en entornos no explorados, además de mayores desafíos en escenarios abiertos o simétricos.

Como alternativa, actualmente se aborda este reto mediante el uso de redes neuronales entrenadas con diversos algoritmos de *Deep Learning*. Es lo que se conoce como las soluciones de navegación basadas en aprendizaje, que emplean, principalmente, sensores visuales. Es por ello, que el problema recibe el nombre de Navegación Visual Semántica. Algunos de los trabajos más recientes en esta dirección son los siguientes[? ? ?].

El entrenamiento de estas soluciones se lleva a cabo en entornos virtuales, donde un agente tiene la capacidad de desplazarse por espacios reales renderizados. Esta metodología permite acelerar el proceso de aprendizaje al aprovechar máquinas especializadas en computación paralela, y algoritmos diversos como el aprendizaje por refuerzo o el aprendizaje por imitación.

Con el fin de demostrar y comparar la capacidad de navegación de estos agentes, se les asigna la tarea de localizar un objeto dentro de estancias renderizadas, las cuales no han explorado previamente.

Se ha demostrado que estos agentes son capaces de alcanzar altos porcentajes de éxito en la consecución de sus objetivos dentro de entornos virtuales. No obstante, cuando se implementan estas redes en el mundo real, surgen desafíos relacionados con la **adaptación de dominio**. Estos problemas se deben, en parte, a la menor precisión de los sensores y a factores inherentes al entorno físico, como variaciones en la iluminación.

El propósito de estas redes neuronales es emular el comportamiento humano, lo cual explica que las entradas a las redes correspondan habitualmente a los mismos estímulos sensoriales que utilizan los seres humanos. Por esta razón, se emplean de manera predominante imágenes RGB, odometría y cámaras de profundidad. Sin embargo, el ser humano no se desplaza únicamente en función de las imágenes que percibe, sino que también otorga un sentido semántico a los objetos dentro de esas imágenes. De este modo, es capaz de identificar rápidamente la función de cada elemento que observa, como suelos, paredes, armarios o sillas, asignándoles una categoría semántica de manera casi instantánea.

Esta capacidad para categorizar visualmente ha sido abordada con gran éxito en el campo de la visión artificial. En la actualidad, contamos con segmentadores semánticos que han demostrado un rendimiento sobresaliente[?].

En este proyecto, pretendemos estudiar el impacto que los sensores de segmentación semántica tienen en los modelos de navegación visual semántica. Postulamos que son una fuente de información que puede aliviar el problema de la adaptación de dominio, y, por lo tanto, el desarrollo de modelos de navegación que puedan funcionar en el mundo real, y no solo en entornos virtuales.

Así, en este trabajo, proponemos desarrollar soluciones de navegación visual semántica empleando aprendizaje por imitación (**IL** (Imitation Learning)) que recibirán imágenes de segmentación semántica como entrada. Posteriormente, se comparará su rendimiento con respecto a modelos que utilizan imágenes RGB, tanto en entornos virtuales como en escenarios del mundo real.

En la figura?? proponemos una representación del tipo de soluciones que se van a desarrollar, denominadas como SemNav, puede observarse como a la entrada de la red se ingresará información de segmentación semántica, complementada con otro tipo de información, que ayudará al modelo a completar la navegación en entornos desconocidos.

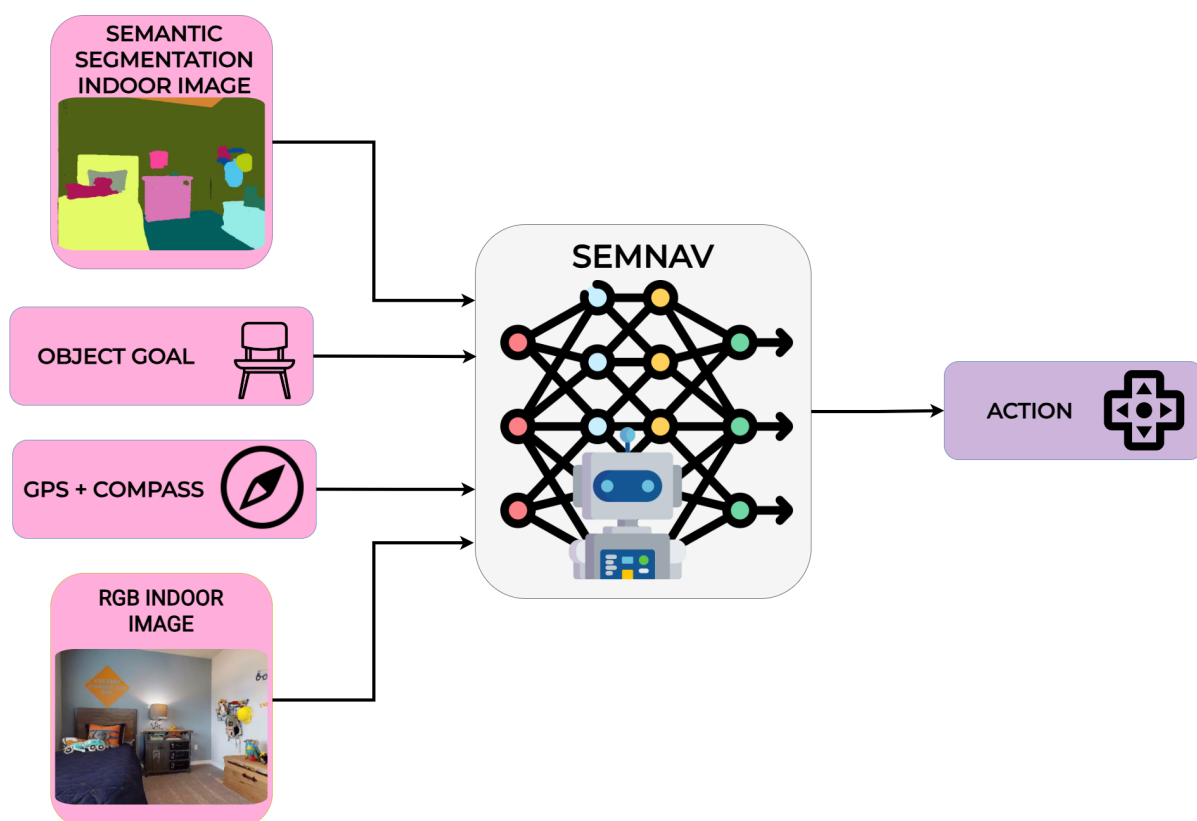


Figura 1.1: Modelo SEMNav para la navegación basada en información de segmentación semántica del entorno.

Capítulo 2

Desarrollo del modelo y dataset para la navegación mediante segmentación semántica

Como ya ha sido tratado en los capítulos anteriores de este documento, el modelo propuesto de navegación mediante segmentación semántica estará basado en el modelo propuesto por PirlNav[?] con las modificaciones indicadas en la figura

$$\text{floor}\left(\frac{255 * 255 * 255}{\text{num_categories}}\right) \quad (2.1)$$

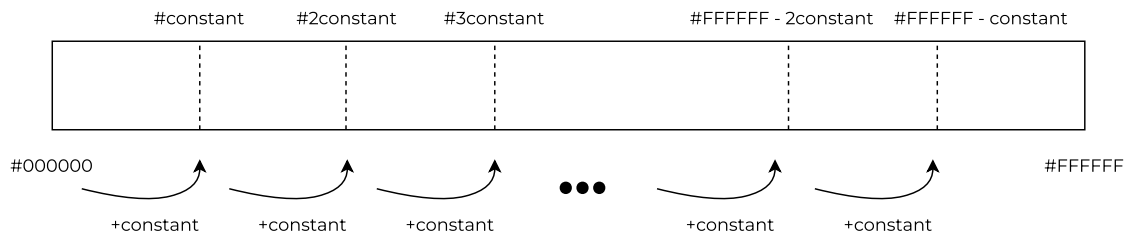


Figura 2.1: Esquema de generación de colores mediante la suma de la constante producida por la ecuación??.

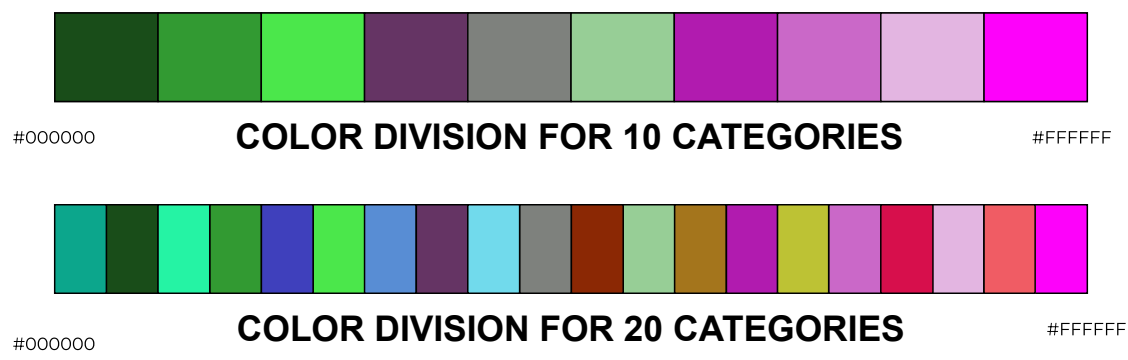


Figura 2.2: Ejemplo de generación de colores mediante la estrategia de suma de constante para 10 y 20 categorías.

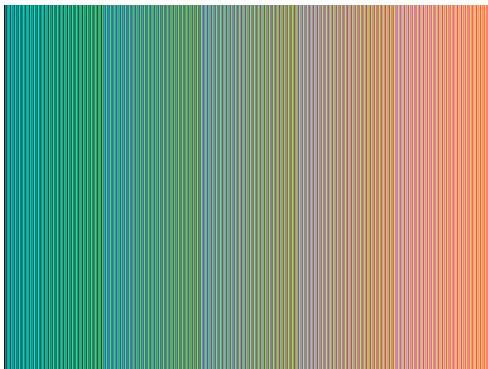


Figura 2.3: Fotografía generada mediante estrategia de división del espectro del color.

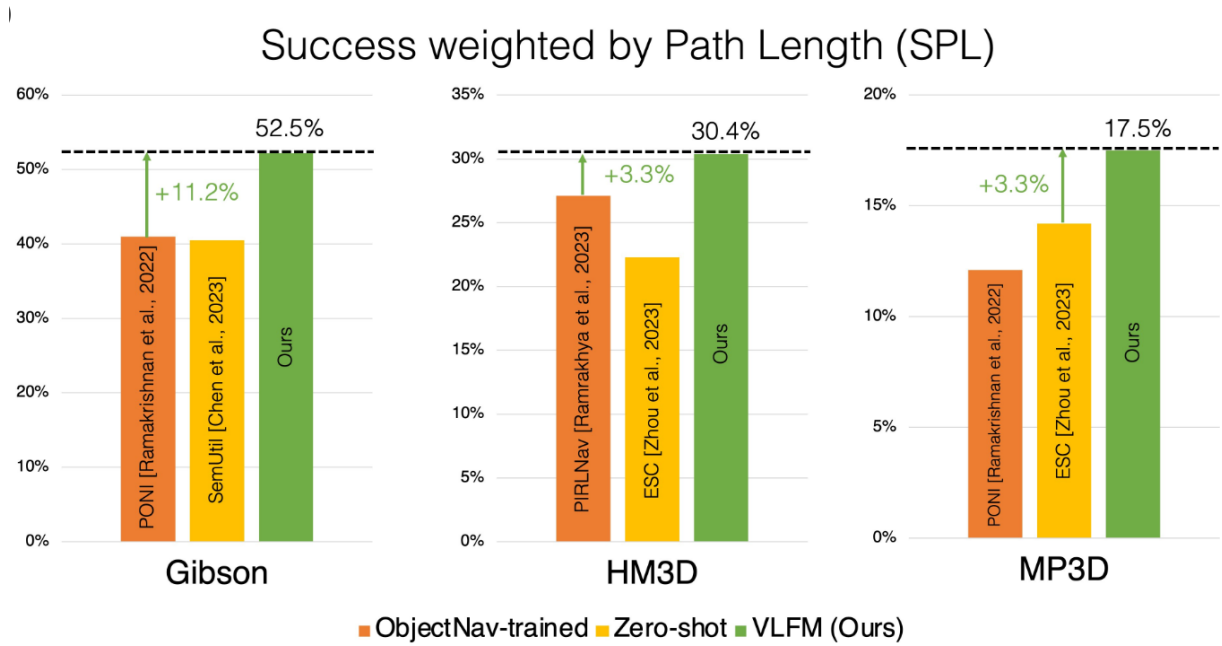


Figura 2.4: Esquema de generación de colores mediante la suma de la constante producida por la ecuación??.

2.1. Resultados

	SR					
	SEMANTIC + RGB		ONLY SEMANTIC		PIRLNAV	
	DINO	NO DINO	DINO	NO DINO	DINO	NO DINO
1630 cat	70.0 %	69.2 %	I.P(14daysstopped)	65.1 %	60.8 %	39.5 %
40 cat	74.95 %	74.3 %	72.3 %	72.9 %		

