UNIVERSIDAD MIGUEL HERNÁNDEZ DE ELCHE

Aplicación de la Apariencia Global de la Información Visual Omnidireccional en Color a Tareas de Navegación Robótica en Espacio $2\frac{1}{2}$D

Tesis doctoral presentada por Francisco Javier Amorós Espí

dentro del Programa de Doctorado en Tecnologías Industriales y de Telecomunicación

Dirigida por Dr. Ing. Oscar Reinoso García.

UNIVERSITAS
Miguel
Hernández

UNIVERSIDAD MIGUEL HERNÁNDEZ DE ELCHE

Aplicación de la Apariencia Global de la Información Visual Omnidireccional en Color a Tareas de Navegación Robótica en Espacio $2\frac{1}{2}$D

Tesis doctoral presentada por Francisco Javier Amorós Espí

dentro del Programa de Doctorado en Tecnologías Industriales y de Telecomunicación

Dirigida por Dr. Ing. Oscar Reinoso García.

Elche, Febrero de 2014

UNIVERSITAS
Miguel
Hernández

AUTORIZACIÓN DE PRESENTACIÓN DE TESIS DOCTORAL

Director: Oscar Reinoso García

Título de la tesis: ***Aplicación de la Apariencia Global de la Información Visual Omnidireccional en Color a Tareas de Navegación Robótica en Espacio $2\frac{1}{2}D$***

Autor: Francisco Javier Amorós Espí

Departamento de Ingeniería de Sistemas y Automática
Universidad Miguel Hernández de Elche

El director de la tesis reseñada CERTIFICO QUE HA SIDO REALIZADA BAJO MI DIRECCIÓN POR D. Francisco Javier Amorós Espí en el Departamento de Ingeniería de Sistemas y Automática de la Universidad Miguel Hernández de Elche y autorizo su presentación.

En Elche, a de del 2014.

Fdo: Dr. D. Oscar Reinoso García

UNIVERSITAS
Miguel
Hernández

DEPARTAMENTO DE CIENCIAS MATERIALES, ÓPTICA Y TECNOLOGÍA ELECTRÓNICA

Dña. Julia Arias Rodriguez, Profesora Titular de Universidad y Directora del Departamento de Ciencias Materiales, Óptica y Tecnología Electrónica de la Universidad Miguel Hernández de Elche.

Certifica

que el trabajo realizado por D. Francisco Javier Amorós Espí titulado ***Aplicación de la Apariencia Global de la Información Visual Omnidireccional en Color a Tareas de Navegación Robótica en Espacio $2\frac{1}{2}D$*** ha sido dirigido por el Dr. D. Oscar Reinoso García y se encuentra en condiciones de ser leído y defendido como Tesis Doctoral ante el correspondiente tribunal en la Universidad Miguel Hernández de Elche.

Lo que firmo para los efecto oportunos en Elche, a de del 2014.

Fdo.: Dña. Julia Arias Rodriguez
Directora del Departamento de Ciencias Materiales,
Óptica y Tecnología Electrónica

DEPARTAMENTO DE CIENCIAS MATERIALES, ÓPTICA Y TECNOLOGÍA ELECTRÓNICA
Campus de Elche. Edif. Torrevaillo, Avda. de la Universidad s/n, 03202 Elche
Telf: 96 665 8498, Fax: 96 665 8497

UNIVERSITAS
Miguel
Hernández

Abstract

The autonomous navigation of a robot requires of the knowledge of the environment it is surrounded with. The robot can use different sensors to gather the information around its position. This information must be processed and interpreted by the robot in order to perform its task, which directly depends on the kind of sensor used.

Among the multiple sensors the robot can be equipped with, visual systems stand out. These sensors are light, have reduced energy consumption and multiple configuration options, which make them suitable for almost any application and means of navigation. Moreover, images provide very rich information, which can be used in different ways. Considering visual navigation, during the last decades feature-based approximations have stood out in order to describe the visual information. These techniques use the image segmentation or the extraction of distinctive points, also known as landmarks.

This thesis suggests the application of global-appearance to the visual information in order to obtain image descriptors. Unlike feature-based methods, these techniques process the image as a whole, without considering the scene contents.

Global-appearance descriptors present an important advantage in unstructured environments, where landmark recognition might be difficult. However, as they work with the whole image, it is important to find methods that process images efficiently and describe them in few terms. These descriptors have demonstrated their utility in navigation tasks, allowing the pose estimation in visual maps. The great majority of the proposals use the visual information of a single channel, which corresponds to the grey-scale image.

For that reason, a comparison of different global-appearance techniques using the colour in different fashions is carried out. This study covers the computational requirements and the precision of the pose estimation in a dense map, also simulating different noise and occlusion situations in the test scenes.

Once the abilities of characterization and distinction of images have been proved, our intention is to use these techniques to extract information regarding the relative position of two scenes captured closely. Specifically, three different situations are suggested to apply the global appearance to visual navigation tasks.

First, we use the information provided by projective images to select and order nodes distributed along the navigation area, making up the map of the environment. To carry out this goal, the system estimates the relative displacement between two consecutive images. This is achieved by means of the Multiscale Analysis, defining the scales as artificial zooms of the original scene. With this study, we aim to demonstrate that, apart from the possibility of obtaining measures of the image displacement, the applicability and performance of the global appearance in non-omnidirectional images.

The objective of the second study is to adapt the Multiscale Analysis to omnidirectional scenes. Combining omnidirectional information and the Multiscale Analysis, an odometry visual system is proposed, using the global appearance of scenes. Afterwards, the topological visual odometry is used in visual path estimation, considering also loop closures to improve the initial estimations.

Finally, we address the problem of vertical displacement estimation between scenes using visual global appearance. The increasing interest in Unmanned Aerial Vehicles (AUV's) as a navigation platform, combined with visual sensors, encourages us to study the application of omnidirectional information to obtain a topological height estimator.

All the proposals are validated by means of experiments that use our own image database captured in real environments.

Resumen

La navegación de un robot de forma autónoma implica el conocimiento del entorno que le rodea. Existen distintos sensores que permiten al robot recoger la información de su alrededor. Esta información debe ser procesada e interpretada por el robot para el desempeño de su tarea.

De entre los múltiples sensores con los que un robot puede estar equipado, los sistemas de visión pueden ser destacados. Son sensores con un peso reducido, un consumo energético limitado y unas opciones de configuración muy extensas, lo cual los convierten en aptos para casi cualquier tipo de aplicación y medio de navegación. Además, las imágenes proporcionan información de gran riqueza, que puede ser aprovechada de distintas formas.

Dentro de la visión aplicada a navegación, en las últimas décadas han predominado las aproximaciones basadas en extracción de características para describir la información visual. Estas técnicas utilizan la segmentación de la imagen o la extracción de puntos significativos de las escenas.

En esta tesis se plantea el uso de la apariencia global de la información visual para llevar a cabo la descripción de las imágenes y sus aplicaciones en tareas de navegación. A diferencia de los métodos basados en características, las técnicas de apariencia global procesan la imagen en su conjunto, sin tener en cuenta el contenido de la escena. Estos descriptores presentan una clara ventaja en entornos desestructurados, donde es difícil extraer puntos significativos. Sin embargo, al trabajar con la imagen completa, es necesario buscar técnicas que permitan el procesamiento de la escena de forma rápida y que describan la escena en pocos términos.

En este trabajo se emplean dos sistemas visuales distintos: una cámara con lente de ojo de pez, que recogen un amplio campo de visión, y un conjunto catadióptrico, con el que se capturan imágenes omnidireccionales. Se incluye un estudio de las múltiples posibilidades de representación de las escenas que la información omnidireccional proporciona.

Los descriptores de apariencia global han demostrado su utilidad en tareas de navegación, permitiendo la estimación de la pose del robot dentro de mapas visuales. La mayoría de las técnicas propuestas utilizan la información visual de un sólo canal, es decir, de la imagen en escala de grises.

Por ello, se realiza una comparación de diferentes algoritmos basados en la apariencia global utilizando la información de color de distintas formas. El estudio tiene en consideración los requerimientos computacionales y la precisión de estimación de la pose dentro de un mapa denso, simulando además distintas situaciones de ruido y oclusiones en las escenas de test.

Demostrada la capacidad de caracterización y distinción que este tipo de descriptores poseen, nuestra intención es utilizar estas técnicas para extraer información sobre la posición relativa de escenas capturadas en un entorno cercano. Específicamente, se plantean tres situaciones distintas para aplicar la apariencia global a tareas de navegación visual.

El primero de ellos utiliza la información proporcionada por rutas de imágenes proyectivas para seleccionar y ordenar nodos de imágenes distribuidos por el área de navegación, formando el mapa del entorno. Para llevar a cabo esta tarea, el sistema estima el desplazamiento relativo entre dos imágenes consecutivas. Esto se consigue mediante el Análisis Multiescala, entendiendo las escalas como ampliaciones artificiales de las escenas. Con este estudio se pretende demostrar, además de la posibilidad de obtener una medida del desplazamiento entre dos imágenes, la aplicabilidad y desempeño de la apariencia global sobre información no omnidireccional.

El segundo caso a estudio tiene como objetivo la adaptación del Análisis Multiescala a las escenas omnidireccionales. Aprovechando las propiedades de la información omnidireccional, y el análisis multiescala, se propone un sistema de odometría visual utilizando la apariencia global de las escenas. Este sistema de odometría visual topológico es utilizado posteriormente en un proceso de estimación de rutas visuales, considerando además cierres de bucle para mejorar las estimaciones.

Por último, se aborda el problema de estimación de desplazamientos verticales entre dos escenas usando de nuevo la apariencia global visual. El creciente interés en los vehículos aéreos como plataformas de navegación y el uso combinado con sensores visuales, nos llevan a plantear la utilización de la información omnidireccional junto con los descriptores de apariencia global para obtener estimadores de altura topológica.

Todas las técnicas propuestas se validan a través de experimentos que usan distintas bases de imágenes propias capturadas en entornos reales.

Acknowledgements

First, I would like to thank my supervisor, Oscar Reinoso, for supporting and guiding me during this project. I owe him a debt of gratitude for his advice and unconditional help.

I also want to thank Luis Payá. First as a teacher during my degree and later as a colleague, he has offered his suggestions and guidance on every doubt and problem.

My thanks to Andrew Calway for accepting me in his research group at Bristol University during my stay. His ideas mean a new line of investigation.

My gratitude to my colleagues of the Automation, Robotics and Computer Vision group: Arturo Gil, David Úbeda, Luis Miguel Jiménez and Jose María Marín, and specially to those whom I have spend almost the whole of the daytime with: Mónica Ballesta, Lorenzo Fernández, Miguel Juliá and David Valiente.

I also want to sincerely thank Maria José Velacoracho, whose help has gone beyond the administrative processes, becoming an important support. Also to all my other colleagues, Nuria, Javi, Manuel, Ramón, Fran,... for making such a warm and friendly work environment.

Last but not least, I want to specially express my immense gratitude to my family: to my parents, who have guided me all my life, the encouragement of my sister and brother-in-law, the energy of my nephews, and the attention of my grandparents.

UNIVERSITAS
Miguel
Hernández

Agradecimientos

Primero, me gustaría agradecer a mi director de tesis, Oscar Reinoso, todo el apoyo que me ha ofrecido durante este tiempo. Le debo mi más profundo agradecimiento por sus consejos y ayuda incondicional, por encontrar siempre un hueco para atenderme.

También debo dar las gracias especialmente a Luis Payá. Primero como profesor durante la carrera, y luego dentro del grupo de investigación, siempre ha estado ahí para ofrecer su experiencia en todas las cuestiones que han surgido en el día a día de esta investigación.

Mi agradecimiento a Andrew Calway, por aceptarme en su equipo en la Universidad de Bristol durante el periodo de estancia. Sus ideas significaron un nuevo enfoque en el transcurso de la tesis.

Por otro lado, quiero dar las gracias a mis compañeros del grupo de ARVC: Arturo Gil, David Úbeda, Luis Miguel Jiménez y Jose María Marín, por el apoyo mostrado y su colaboración. Y especialmente a mis compañeros del laboratorio, con quienes he pasado la mayor parte de este tiempo: Mónica Ballesta, Lorenzo Fernández, Miguel Juliá y David Valiente. El haberme permitido compartir con ellos momentos especiales de sus vidas, y las muchas horas de trabajo juntos, me hacen considerarlos como amigos más que como compañeros de trabajo. Junto a ellos he crecido tanto académica como personalmente. Sin olvidarme de la última incorporación al grupo, Adrián Peidró.

Quiero expresar mi gratitud a María José Velacoracho, cuya ayuda va más allá de las gestiones administrativas, convirtiéndose en un importante apoyo. También al resto de mis compañeros, Nuria, Javi, Manuel, Ramón, Fran,... Todos ellos consiguen crear un ambiente de trabajo inigualable.

Doy las gracias a mis amigos, quienes han sabido estar en los momentos buenos, y en los que necesitaba un apoyo extra, tratando de echar siempre una mano.

Por último, debo expresar mi inmensa gratitud a mi familia. A mis padres, Paco y Rita, quienes han sido guía y modelo durante toda mi vida. A mi hermana, María José, y a mi cuñado, Carlos, por tener siempre alguna palabra de ánimo en los momentos bajos. A mis sobrinos, Guillermo y Gonzalo, por ser la energía que necesitaba muchas tardes para poder continuar. Y a mis abuelos, que tantos esfuerzos, tiempo e interés me han dedicado a lo largo de toda mi vida.

UNIVERSITAS
Miguel
Hernández

A mi familia.

UNIVERSITAS
Miguel
Hernández

Índice general

Índice de figuras **xix**

Índice de tablas **xxvii**

1 Introducción **1**

1.1 Motivación 1
1.2 Objetivos 3
1.3 Marco de la Tesis 4
1.4 Publicaciones 6
1.5 Estructura 8

2 Estado de la Técnica **11**

2.1 Entornos de Navegación 11
2.1.1 Robots Submarinos 11
2.1.2 Robots Aéreos 13
2.1.3 Robots Terrestres 15
2.2 Sensores visuales 18
2.3 Descripción de la información visual 21
2.4 Navegación y Creación de Mapas 26

3 Visión Omnidireccional en Navegación Robótica **31**

3.1 Sistemas de Visión Catadióptricos 33
3.1.1 Sistema Catadióptrico usado en este trabajo 38
3.2 Representación de la Información Omnidireccional 43
3.2.1 Imagen Esférica 44
3.2.2 Imagen Panorámica 45
3.2.3 Imagen Perspectiva 48
3.2.4 Imagen Ortográfica 50

3.3 Sistemas de Visión con Lente Ojo de Pez. 53
3.3.1 Sistema con Lente Ojo de Pez usado en este trabajo 53
3.4 Plataformas de Adquisición de Imágenes 56
3.4.1 Robot Pioneer P3-AT 56
3.4.2 Trípode 58

4 Apariencia Global de Información Visual: Descriptores 61
4.1 Técnicas basadas en la Transformada de Fourier 63
4.1.1 Transformada 1D de la Imagen 65
4.1.2 Firma de Fourier 68
4.1.3 Transformada 2D de Fourier 68
4.1.4 Transformada Esférica de Fourier (SFT) 70
4.2 Técnicas basadas en el Análisis de Componentes Principales (PCA) 77
4.2.1 Análisis de Componentes Principales (PCA) 77
4.2.2 Variante al desarrollo matemático de PCA 81
4.2.3 Análisis de Componentes Principales de Modo Incremental (PCAI) 82
4.2.4 PCA Rotacional 83
4.2.5 PCA sobre la Firma de Fourier 86
4.3 Histogramas de Orientación del Gradiente (HOG) 87
4.3.1 Implementación del Algoritmo 88
4.3.2 Aplicación a tareas de Navegación 91
4.4 GIST 95
4.4.1 Filtros de Gabor 96
4.4.2 Gist-Gabor 100
4.4.3 Gist-Color 105

5 Análisis Comparativo de Técnicas de Apariencia Global sobre Escenas Panorámicas en Color. 113
5.1 Consideración de la Información de Color. 115
5.2 Base de Imágenes. 119
5.3 Experimentos y Resultados 128
5.3.1 Variables de los Descriptores 128
5.3.2 Análisis Comparativo 130
5.3.2.1 Selección de Parámetros 130
5.3.2.2 Espacios de Color 141
5.3.2.3 Comportamiento ante Ruido y Oclusiones 146
5.4 Conclusiones 152

6 Análisis Multiescala en Tareas de Navegación Topológica **155**
6.1 Construcción de Mapas y Localización usando el Análisis Multiescala sobre Imágenes Proyectivas . . . 157
6.1.1 Análisis Multiescala . . . 158
6.1.1.1 Estimación de la posición relativa entre dos escenas usando Análisis Multiescala . . . 161
6.1.2 Construcción del Mapa . . . 162
6.1.2.1 Asociación de las Imágenes de Nodos y Rutas . . . 162
6.1.2.2 Construcción del Grafo . . . 164
6.1.3 Estimación de las Rutas sobre el Mapa . . . 168
6.1.3.1 Función de Ponderación . . . 168
6.1.3.2 Localización dentro del Mapa . . . 169
6.1.4 Experimentos y resultados . . . 170
6.1.4.1 Selección del Descriptor y Resolución de Imagen . . . 170
6.1.4.2 Bases de Imágenes . . . 172
6.1.4.3 Resultados de la Construcción del Mapa . . . 177
6.1.4.4 Resultados de Localización de Rutas en el Mapa . . . 179
6.2 Aplicación del Análisis Multiescala a Imágenes Omnidireccionales . . . 182
6.2.1 Extracción de Posición Relativa entre dos Imágenes. . . . 182
6.2.1.1 Aplicación a tareas de navegación . . . 184
6.2.2 Mejora de la localización dentro de mapa topológico . . . 186
6.2.3 Experimentos y resultados . . . 191
6.2.3.1 Selección del Descriptor y parámetros del Análisis Multiescala . . . 191
6.2.3.2 Base de Imágenes . . . 192
6.2.3.3 Resultados de estimación de la ruta . . . 192
6.3 Conclusiones . . . 197

7 Estimación Topológica de Altura. **199**
7.1 Estimación Topológica Altura. . . . 201
7.1.1 Correlación de la Celda Central en Imágenes Panorámicas . . . 201
7.1.2 Desfase Vertical usando FFT2D . . . 202
7.1.3 Análisis Multiescala sobre la Vista Ortográfica . . . 204
7.1.4 Cambio de Coordenadas del Sistema de Referencia de la Cámara (SRC) . . . 205
7.2 Base de Imágenes . . . 208
7.3 Experimentos y Resultados . . . 212

7.4 Conclusiones 221

8 Conclusiones **223**

8.1 Aportaciones 223

8.2 Líneas Futuras de Investigación 225

Bibliografía **229**

Índice de figuras

2.1 Ejemplos de Vehículos Autónomos Submarinos: (a) USAL y (b) Puma. . . 12
2.2 Ejemplos de Vehículos Autónomos Aéreos: (a) Helicóptero, (b) Cuadricóptero y (c) Octocóptero. 14
2.3 Ejemplos de distintas plataformas robóticas terrestres: (a) plataforma bípeda, (b) robot con dos ruedas y (c) coche con sistema de conducción autónoma. 15
2.4 Imágenes de sensores (a) Laser y (b) Sónar. 16
2.5 Imágenes de distintos sistemas visuales. (a) Robot equipado con una sola cámara, (b) Robot con un par estéreo, (c) Sistema de visión trinocular, (d) imagen de coche equipado con array de cámaras para obtener imagen omnidireccional y (e) robot con sistema de visión catadióptrico. 19
2.6 Esquema del proceso de extracción de lineas verticales sobre imagen omnidireccional. 25
2.7 Esquema de tareas de navegación y sus relaciones. 28

3.1 Representación de sistemas cadadióptricos centrales. Sistemas formados por (a) espejo hiperbólico con lente perspectiva, y (b) espejo parabólico con lente ortográfica. 34
3.2 Modelo de proyección de un punto **P** del mundo real en el plano de proyección del sistema catadióptrico (**p**). 36
3.3 Ejemplo de dos imágenes capturadas en un mismo entorno por sistemas catadióptricos distintos. 37
3.4 Cámara CCD DFK-21BF04 38
3.5 Imagen omnidireccional con patrón usado para la calibración del conjunto catadióptrico. (a) Imagen original y (b) imagen con esquinas del patrón marcadas por el algoritmo de calibración. 40

3.6 Calibración del conjunto catadióptrico formado por cámara DFK-41BF02 y espejo Eizo Wide70. (a) Estimación de la posición de los patrones de calibración de las distintas imágenes respecto al sistema de referencia del espejo, y (b) representación de la función de proyección y del ángulo del rayo óptico con respecto a ρ. 41
3.7 Calibración del conjunto catadióptrico formado por cámara DFK-41BF02 y el espejo Accowle SuperWide Large. (a) Estimación de la posición de los patrones de calibración de las distintas imágenes respecto al sistema de referencia del espejo, y (b) representación de la función de proyección y del ángulo del rayo óptico con respecto a ρ. 41
3.8 Imagen Omnidireccional 44
3.9 Modelo de proyección de la imagen esférica. 45
3.10 Imagen Esférica 46
3.11 Modelo de proyección de la imagen panorámica. 46
3.12 Imágenes panorámicas obtenidas mediante cambio de sistema de coordenadas a partir del sistema catadióptrico que utiliza el espejo (a) Eizoh Wide70 y el (b) Accowle SuperWide. 47
3.13 Imagen Panorámica. 48
3.14 Modelo de proyección de una imagen perspectiva. 49
3.15 Imágenes Perspectivas usando distintas distancias entre el foco del espejo y el plano de proyección. 50
3.16 Modelo de proyección de la vista ortográfica. 51
3.17 Imagen Ortográfica 52
3.18 Imagen de la cámara GoPro Hero. 53
3.19 Calibración del sistema dióptrico GoPro Hero 960. (a) Estimación de la posición de los patrones de calibración de las distintas imágenes respecto al sistema de referencia del espejo, y (b) representación de la función de proyección y del ángulo del rayo óptico con respecto a ρ. 54
3.20 (a) Imagen capturada con cámara GoPro Hero y (b) imagen con distorsión corregida. 55
3.21 Imagen del robot P3-AT utilizado para la captura de imágenes. 58
3.22 Imagen del trípode K&M 20811 y detalle de la adaptación de sistema catadióptrico. 59

4.1 (a) Transformada de Fourier de una serie numérica de 10 elementos (Secuencia A) y la de la misma serie rotando un elemento (Secuencia A'); (b) Módulo de los coeficientes de Fourier; (c) Fase de los coeficientes de ambas series, y (d) Diferencia de fases calculadas entre 0^o y 360^o. 66
4.2 Imágenes panorámicas rotadas 68^o entre sí. 67
4.3 Construcción de la Transformada de Fourier 1D de una imagen panorámica. 67
4.4 Módulos de la Firma de Fourier. 69
4.5 Transformada de Fourier 2D y representación en el dominio espacial de una imagen. Escena (a) original, (b) eliminando las bajas frecuencias y (c) eliminando las altas frecuencias. 71
4.6 Sistema de referencia esférico representado en $\mathbb{R}^3$. 72
4.7 Cuadrícula de puntos equiangular sobre esfera unitaria. (a) 16×16 elementos, (b) 32×32 elementos. 73
4.8 Matriz de covarianza de un conjunto de (a) 32 rotaciones y (b) 128 rotaciones de una misma imagen. 84
4.9 Producto interior de una matriz que incluye P=5 localizaciones y N=128 rotaciones. 85
4.10 Proyecciones de 2 componentes del conjunto de rotaciones de una imagen utilizando PCA Rotacional. 85
4.11 (a) Imagen original, (b) Derivada respecto al eje x, (c) Derivada respecto al eje y, y (d) Magnitud del gradiente. 89
4.12 (a) Imagen original, (b) División en celdas de la magnitud del gradiente de la imagen, y (c) representación de la dirección ponderada del gradiente de la imagen. 90
4.13 Dirección ponderada del gradiente e Histograma de Orientación asociado. . 90
4.14 Descriptor HOG para localización sobre imagen panorámica. 92
4.15 Descriptor HOG para estimación de la orientación sobre imagen panorámica, con ejemplo de rotación circular de los histogramas para el cálculo del desfase entre imágenes. 94
4.16 (a) Sinusoide compleja, (b) Envoltura Gaussiana y (c) Filtro de Gabor resultante de la convolución de ambas funciones. 96
4.17 Envolturas espaciales con $x_0 = y_0 = 0$, $a = 1/2$, $b = 1/4$ con ángulo (a) $\theta = 0^o$ y (b) $\theta = 45^o$ 99
4.18 Representación bidimensional de un filtro de Gabor en el espacio de la frecuencia. 99

4.19 Máscaras de Gabor en espacio frecuencia con distintas escalas espaciales y orientaciones. (a) Representación de los límites de las máscaras en 2D, y (b) representación de los valores de las máscaras en 3D. 101
4.20 Filtrado de Gabor de una imagen con (a) diferentes orientaciones (0°, 45° , 90°, 135°) y (b) distintas escalas espaciales. 102
4.21 Obtención del descriptor GIST-Gabor a partir de una imagen filtrada por una máscara de Gabor. 104
4.22 Pirámide Gaussiana de una imagen formada por 8 escalas 106
4.23 Resultado de las operaciones de comparación *center-surround* para distintas escalas sobre los canales RG, BY e I de una imagen. 109
4.24 Esquema de características y operaciones para la obtención de GIST-Color. 110

5.1 Descriptor HOG de una misma imagen en escala de gris, y sobre los canales R, G y B de la imagen en color para (a) Localización y (b) Orientación. . . 115
5.2 Descriptor HOG para una imagen en escala de gris, y cálculo de los histogramas de intensidad de los canales H, S, V con las mismas ventanas y de divisiones por histograma. 118
5.3 Plano de distribución de las estancias incluidas en la base de imágenes. . . 120
5.4 Detalle de la posición de las imágenes del mapa (rojo) y de test (verde) para las zonas (a) 1, (b) 5 y (c) 4 de la base de imágenes. 122
5.5 Detalle de la posición de las imágenes del mapa (rojo) y de test (verde) para las zonas (a) 6, (b) 2 y (c) 3 de la base de imágenes. 123
5.6 Imágenes de ejemplo de cada estancia incluida en la base. 125
5.7 Ejemplo de *aliasing* visual. 126
5.8 Imágenes de ejemplo incluyendo oclusiones y ruido Gaussiano. 127
5.9 Comparación de dos curvas Recall-Precision. 134
5.10 Gráficas *Recall-Precision* incluyendo el vecino más cercano (N.N.), los dos más cercanos (S.N.N.) o los tres vecinos más cercanos (T.N.N.). 135
5.11 Distancia métrica entre el vecino más cercano del mapa y la posición estimada de las imágenes de test. 137
5.12 Desfase entre la la orientación real del robot y la obtenida experimentalmente. 138
5.13 Memoria necesaria para almacenar el mapa. 139
5.14 Tiempo para (a) creación del mapa y (b) estimación de la pose. 140
5.15 Precisión de localización usando la información de color. 143
5.16 Memoria necesaria para almacenar el mapa usando la información de color 144

5.17 Tiempo usando la información de color para (a) creación del mapa y (b) estimación de la pose. 145
5.18 Precisión de localización usando la información de color ante oclusiones en las imágenes de test. 147
5.19 Precisión de localización usando la información de color ante ruido Gaussiano en las imagenes de test. 148
5.20 Imágenes de ejemplo incluyendo oclusiones y ruido Gaussiano. 149
5.21 Error en la estimación de fase ante (a) oclusiones y (b) ruido Gaussiano en las imágenes de test. 151

6.1 Representación de escena capturada por un sistema de visión considerando un desplazamiento perpendicular al plano de proyección. 158
6.2 Comparación Recall-Precision en precisión de asociación de imágenes al introducir análisis multiescala, utilizando HOG sobres escenas de resolución 32×64 píxeles. 159
6.3 Asociación entre imágenes usando el Análisis Multiescala. (a) Imagen del Nodo1, (a') ampliación de imagen (a), y (b) imagen de ruta localizada delante del Nodo1. (c) Imagen de ruta localizada tras el Nodo2, (c') ampliación de imagen (c), y (d) imagen del Nodo2. l_1 y l_2 representan distancias topológicas entre las imágenes de ruta y el nodo más cercano, y c la distancia entre nodos. 160
6.4 Escenas consecutivas de una ruta, y distancia imagen de las escenas respecto a distintas escalas de la Escena 1. 161
6.5 Ejemplo de asociación de imágenes usando dos imágenes de nodo y nueve escenas de ruta que conectan ambos nodos. En la parte derecha, los resultados de asociación incluyen el nodo más cercano (n), la escala de la escena de nodo (s^n), la escala de la escena de ruta (s^r) y la distancia topológica (l). 163
6.6 Estimación del cambio de fase en un nodo. 167
6.7 (a) Tiempo y (b) Memoria usando distintos tamaños de imágenes y descriptores. 171
6.8 Resultados Recall-Precision en precisión de Localización considerando el Tercer Vecino más Cercano (T.N.N.) para distintas resoluciones de imagen usando (a) la Firma de Fourier, (b) HOG, (c) GIST-Gabor y (d) GIST-Color. 173
6.9 Representación de los distintos grafos sobre el plano de cada área de navegación. 175
6.10 Representación de los mapas y las rutas de las distintas áreas. 176

6.11 (a), (b), (c) Escenas correspondientes a Nodos del Área 1. (d), (e), (f) Escenas correspondientes a Rutas del Área 1. 176
6.12 Grafos con las distribuciones de los nodos obtenidos experimentalmente en la construcción del mapa del (a) Área 1 y (b) Área 2. 177
6.13 Precisión de asociación de imágenes respecto a las constantes de ponderación w_1 y w_2. (a) Precisión variando la constante relativa a la distancia topológica (w_1) y (b) precisión variando la constante relativa al cambio de fase (w_2). 179
6.14 Resultados de estimación del camino de las diferentes rutas del Área 1 y 2. . 181
6.15 Extracción de escenas proyectivas en la dirección de avance y de la opuesta a partir de una imagen omnidireccional. 183
6.16 Variación de focales de las escenas proyectivas en la dirección de avance y la opuesta para su aplicación en el Análisis Multiescala. 185
6.17 Ejemplo de distintas trayectorias seguidas por el robot. (a) Trayectoria rectilínea, y (b) Trayectoria con cambio de dirección. 185
6.18 Ejemplo de corrección de ruta por cierre de bucle. (a) Detección de escena por la que se ha pasado anteriormente, (b) corrección de los desfases de la odometría, y (c) corrección de las posiciones de la trayectoria. 190
6.19 Esquema del algoritmo de estimación de la ruta usando la odometría visual topológica. 191
6.20 Ruta seguida por el robot, junto con ejemplos de escenas de los distintos entornos. 193
6.21 Representación gráfica de la ruta de referencia y la ruta estimada usando (a) Fourier HSV y (b) HOG con HC. 194
6.22 Tiempo de comparación entre descriptores para localización variando el número de imágenes incluidas en la base. 195

7.1 Selección de celdas sobre una escena panorámica para estimación de altura usando la Correlación de la Celda Central. 202
7.2 Ejemplo de proyección ortográfica utilizando distancias focales distintas. . 204
7.3 Esquema de proyección de un punto al variar el sistema de referencia de la cámara (SRC) usando la geometría epipolar, junto con ejemplo de imagen omnidireccional aplicando desplazamiento vertical, y su transformada panorámica. 206
7.4 Plano de localizaciones de las imágenes de (a) exterior y (b) interior capturadas a distintas alturas. 209

7.5 Ejemplos de imágenes omnidireccionales capturadas en el entorno de exterior en tres localizaciones distintas variando la posición relativa con los edificios y las condiciones de iluminación 210
7.6 Ejemplos de imágenes omnidireccionales capturadas en el entorno de exterior en la misma localización variando su altura. (a) Altura 125 cm ($h = 1$), (b) altura 200 cm ($h = 6$) y (c) altura 290 cm ($h = 12$). 210
7.7 Ejemplos de imágenes omnidireccionales capturadas en el entorno de interior con distintas estancias. 211
7.8 Ejemplos de imágenes omnidireccionales capturadas en el entorno de interior en la misma localización variando su altura. (a) Altura 125 cm ($h = 1$), (b) altura 185 cm ($h = 5$) y (c) altura 275 cm ($h = 11$). 211
7.9 Variación del ángulo de incidencia de los rayos provenientes de dos puntos a distinta distancia cuando se produce una variación de la altura del sistema visual. 215
7.10 Desplazamiento de elementos de la escena panorámica situados a distintas distancias respecto del sistema visual al variar la altura de captura. 216
7.11 Estimación del desplazamiento vertical de las distintas escenas de exterior tomando como referencia la imagen a altura $h = 1$ y $h = 5$. 217
7.12 Estimación del desplazamiento vertical de las distintas escenas de interior tomando como referencia la imagen a altura $h = 1$ y $h = 5$. 218
7.13 Estimación de distintos gradientes de desplazamiento vertical positivos para imágenes de exterior e interior. 219
7.14 Estimación de distintos gradientes de desplazamiento vertical negativos para imágenes de exterior e interior. 220

UNIVERSITAS
Miguel
Hernández

Índice de tablas

3.1 Especificaciones de los sensores CCD usados en este trabajo. 39
3.2 Especificaciones técnicas de los espejos Eizoh Wide 70 y Accowle Super-Wide Large. 39
3.3 Especificaciones Técnicas de la cámara GoPro Hero 960. 55
3.4 Especificaciones Técnicas del Robot Pioneer P3-AT 57
3.5 Especificaciones del Trípode K&M 20811. 59

4.1 Escalas de la pirámide Gaussiana de la imagen comparadas en las operaciones *center-surround* 108

5.1 Número de imágenes de la base de imágenes por área. 119
5.2 Número de imágenes de test por área de la base experimental. 121
5.3 Parámetros de cada descriptor. 131
5.4 Parámetros seleccionados para cada descriptor. 136

6.1 Resoluciones de Imagen usadas en los experimentos 171
6.2 Número de imágenes por área de la base experimental. 174
6.3 Resultados del análisis Procrustes de los grafos de los distintos mapas. . . . 178
6.4 Error en la estimación de la ruta medido con análisis Procrustes (μ) variando el umbral de cierre de bucle (th_{pano}) y el parámetro de localización del descriptor (N) usando la Firma de Fourier sobre HSV. 195
6.5 Error en la estimación de la ruta medido con análisis Procrustes (μ) variando el umbral de cierre de bucle (th_{pano}) y el parámetro de localización del descriptor (C_H) usando la HOG junto con Histograma de Color (HC). . . . 195

7.1 Altura de cada imagen respecto al plano del suelo, y número de imágenes de cada base por altura. 208
7.2 Diferencia de altura para los distintos gradientes de imágenes y número de comparaciones posibles para la base de exterior e interior. 212

7.3 Resumen de Métodos de Estimación de altura junto con las distintas representaciones de la información omnidireccional, el descriptor empleado y el indicador de cambio de altura. 213

CAPÍTULO

1

Introducción

1.1 Motivación

Esta tesis estudia la aplicación de descriptores basados en apariencia global sobre imágenes a tareas de navegación robóticas. Se aborda especialmente, aunque no de forma única, su aplicación sobre escenas omnidireccional obtenida a partir de un conjunto catadióptrico, tratando de aprovechar distintas posibilidades de proyección de la información visual que este tipo de escenas permiten.

Comparados con los descriptores basados en características, los descriptores que tratan la imagen en su conjunto (sin segmentación ni extracción de marcas significativas) pueden ser considerados una línea reciente de investigación, con una tendencia creciente en su empleo en aplicaciones de navegación autónoma. Estos descriptores basados en apariencia global han demostrado ser capaces estimar la pose de un robot dentro de un mapa denso de imágenes. La mayoría de las técnicas propuestas utilizan la información visual de un sólo canal, es decir, de la imagen en escala de grises. Sin embargo, no tenemos conocimiento de ningún estudio sobre la utilización de la información de color y la apariencia global de forma comparativa con distintas técnicas.

Por ello, se realiza una comparación de diferentes algoritmo basados en la apariencia global utilizando la información de color de distintas formas. En concreto, se usan los canales RGB, HSV, y un histograma con los niveles de intensidad de los canales de color. El estudio tiene en consideración los requerimientos computacionales y la precisión de estimación de la pose dentro de un mapa denso, teniendo en cuenta además distintas situaciones de ruido y oclusiones en las escenas de test.

Demostrada la capacidad de caracterización y distinción que este tipo de descriptores poseen, nuestra intención es utilizar estas técnicas para extraer información sobre la posición relativa de dos escenas capturadas en un entorno cercano. Para ello, se van a plantear distintos experimentos de navegación visual.

El primero de ellos utiliza la información proporcionada por rutas de imágenes para seleccionar y ordenar nodos distribuidos por el área de navegación. Estos nodos están compuestos por distintas imágenes proyectivas capturadas en diferentes orientaciones respecto el plano del suelo, de manera que cubren el área visual completa alrededor de su posición. La distribución de los nodos se representa mediante un grafo, que será el mapa del entorno. Para poder posicionar los distintos nodos en el grafo, es necesario estimar el desplazamiento relativo entre las imágenes las rutas de forma consecutiva. Esto se consigue mediante el Análisis Multiescala, entendiendo las escalas como distintas ampliaciones artificiales de las escenas. Con este estudio se pretende demostrar, además de la posibilidad de obtener una medida del desplazamiento entre dos imágenes, la aplicabilidad y desempeño de la apariencia global sobre información no omnidireccional.

El segundo caso a estudio tiene como objetivo la adaptación del análisis multiescala a las escenas omnidireccionales. A partir de la información omnidireccional, se pueden extraer proyecciones sobre planos perpendiculares a la dirección de avance. Aplicando el análisis multiescala sobre esas proyecciones, es posible estimar el desplazamiento relativo entre las escenas omnidireccionales. Uniendo la capacidad de los descriptores para estimar la orientación entre dos escenas, y el análisis multiescala, se propone un sistema de odometría visual utilizando la apariencia global de las escenas. Los experimentos emplean una ruta de imágenes capturada en un entorno real de interior. El propósito es la estimación de la ruta seguida por el robot de forma topológica usando únicamente información visual. El algoritmo tendrá en cuenta cierres de bucle para mejorar la estimación de la ruta, que se realizará mediante el reconocimiento de zonas visitadas anteriormente y la estimación de la posición respecto a la escena seleccionada.

De igual forma, no hemos encontrado en la bibliografía referencias que traten la estimación de desfase vertical usando la apariencia global de las imágenes. Sin embargo, el creciente interés en los vehículos aéreos como plataformas de navegación y el uso de los sensores visuales, nos llevan a plantear la utilización de la información omnidireccional junto con los descriptores de apariencia global para obtener estimadores de altura topológica.

1.2 Objetivos

Esta tesis tiene como objetivos el análisis y comparación de técnicas basadas en apariencia global de imágenes, especialmente de información visual omnidireccional en color, y de su aplicación en tareas de navegación robóticas.

Esto objetivos pueden dividirse en los siguientes puntos:

- Análisis y Comparación de Técnicas basadas en Apariencia Global de imágenes:
  - Estudio de las principales técnicas actuales y propuesta de un nuevo descriptor.
  - Obtención de una base de imágenes extensa que permita comparar la precisión y requisitos computacionales de los distintos descriptores.
  - Análisis de los descriptores al añadir información de color de las escenas.
  - Estudio del comportamiento de las distintas técnicas ante situaciones comunes que modifican las imágenes iniciales, como son el ruido y las oclusiones

  El Capítulo 4 describe las distintas técnicas utilizadas, mientras que en el Capítulo 5 se incluyen los resultados comparativos.

- Aplicación de Descriptores de Apariencia Global a Tareas de Navegación Robótica basada en Visión:
  - Obtención de un sistema de información mínima para navegación topológica disponiendo de información previa del entorno.
  - Creación de un sistema de navegación que aproveche la apariencia global de información omnidireccional para localizar al robot, estimar el desplazamiento entre escenas consecutivas, y crear un mapa topológico que describa la ruta recorrida por el robot.
  - Estudio de distintas técnicas que permitan el cálculo de la diferencia de altura entre escenas capturadas en un mismo punto usando la apariencia global.

  Estas aplicaciones se describen en los Capítulos 6 y 7.

1.3 Marco de la Tesis

Este trabajo ha sido desarrollado con ayuda de una beca dentro del programa VALi+d para investigadores en formación de la Generalitat Valenciana, convocatoria de 2011, con referencia ACIF/2011/034, además de una beca para realizar un periodo de investigación fuera de la Comunidad Valenciana dentro del mismo programa de ayudas, con referencia BEFPI/2012/083. La estancia, de tres meses de duración, se realizó en el departamento de Computer Science de la Bristol University durante 2012.

La presente tesis se encuentra enmarcada dentro de dos proyectos del grupo de investigación ARVC del Departamento de Ingeniería de Sistemas y Automática de la Universidad Miguel Hernández de Elche. A continuación se describen brevemente dichos proyectos:

- **Proyecto:** *Sistemas de Percepción Visual Móvil y Cooperativo como Soporte para la Realización de Tareas con Redes de Robots*

  Financiado por: CICYT Ministerio de Ciencia e Innovación

  Duración: 01/01/2011 al 31/12/2013

  Descripción: La realización de tareas de forma coordinada por parte de un conjunto de robots resulta de enorme interés para lograr unos mejores resultados en la consecución de las mismas. Es en este ámbito donde se centra el presente proyecto de investigación planteando la necesidad de utilizar diferentes sistemas de visión distribuidos a lo largo de toda la red de agentes móviles de modo tal que se posibilite una adquisición más completa y detallada del entorno. Para acometer los objetivos propuestos se abordan diferentes líneas de investigación. Así, en el marco del proyecto se han desarrollado entre otros los siguientes aspectos: modelización y caracterización de las marcas visuales a partir de información adquirida y de la información del sistema de adquisición, extensión de las técnicas de filtrado Rao-blackwellized a un conjunto de robots que actúan de forma coordinada, diseño de una base de datos coordinada a partir del estudio basado en apariencia de las imágenes adquiridas por cada robot, y para finalizar, estrategias de cooperación para maximizar la calidad de la información percibida.

- **Proyecto:** *Exploración integrada de entornos mediante robots cooperativos para la creación de mapas 3D visuales y topológicos que puedan ser usados en navegación con 6 grados de libertad.*

  Financiado por: CICYT Ministerio de Ciencia e Innovación

  Duración: 01/10/2007 al 31/09/2010

Descripción: El desarrollo de tareas por parte de un conjunto de robots en entornos no determinísticos requiere contar con un mapa del entorno que permita una localización precisa. Durante los últimos años se han desarrollado diferentes métodos que posibilitan la construcción de mapas por parte de un robot móvil a medida que se desenvuelve en un entorno no determinístico y en el que se debe localizar (SLAM). Numerosos han sido los desarrollos llevados hasta el momento para la resolución de esta tarea. A menudo los sistemas sensoriales empleados han consistido en el uso de sensores de rango que posibilitan la construcción de los clásicos mapas de ocupación. En este sentido los mapas visuales constituyen hoy en día una alternativa a los mismos. Por otro lado, la realización de estos mapas puede realizarse de una forma más precisa y rápida al emplear diferentes robots que actúan de forma cooperativa para la realización de esta tarea. Es en este ámbito donde se centra el presente proyecto de investigación, que persigue como objetivo la creación de mapas visuales en entornos 3D no determinísticos por parte de un conjunto de robots móviles equipados con diferentes sistemas de visión. El desarrollo de nuevos mecanismos de exploración integrada en estos entornos donde diferentes agentes móviles desarrollan un mapa visual del entorno a la vez que se localizan y exploran el mismo resulta un reto de enorme potencial que aún no se encuentra resuelto en la actualidad. Para la consecución de este objetivo se aborda dentro de este proyecto de investigación diferentes líneas de investigación que parten de los resultados alcanzados en proyectos de investigación previos. Así se propone abordar en el presente proyecto de investigación, entre otras, las siguientes líneas: SLAM visual cooperativo 6DOF, en el que los robots se muevan con trayectorias generales en el espacio (con seis grados de libertad) en lugar de las clásicas trayectorias en las que se asume que el robot navega en un plano bidimensional; exploración integrada donde las trayectorias de exploración de los robots tengan en cuenta la incertidumbre del/os mapa/s creados por el conjunto de robots; fusión y alineamiento de mapas visuales entre varios robots; y para finalizar la creación de mapas a partir de la información visual basada en apariencia lo que posibilita la construcción de mapas topológicos de mayor nivel.

1.4 Publicaciones

Durante el transcurso de la investigación han sido publicados dos artículo en revista, un capítulo de libro, 9 publicaciones en congresos internacionales y 5 en congresos nacionales.

Publicaciones en revistas

- F. Amorós, L. Payá, O. Reinoso, L. Jiménez, and M. Juliá. (2014). Topological height estimation using global appearance of images. In *Advances in Intelligent Systems and Computing*, Eds. Springer International Publishing, vol. 253, pp. 77-89, 2014.
- F. Amorós, L. Payá, O. Reinoso, and L. Fernández. (2012). Map building and localization using global-appearance descriptors applied to panoramic images. In *Journal of Computer and Information Technology*, vol. 2, no. 1, pp. 55-71, 2012. (SCImago-JCR Impact Factor: 0.139)

Capítulos de Libros

- L. Payá, O. Reinoso, F. Amorós, L. Fernández, and A. Gil. (2011). *Multi-Robot Systems, Trends and Development*. Ed. INTECH, ch. 11: Probabilistic Map Building, Localization and Navigation of a Team of Mobile Robots. Application to Route Following.

Publicaciones en congresos internacionales

- F. Amorós, L. Payá, O. Reinoso, L. Fernández, and D. Valiente. (2014). Towards relative altitude estimation in topological navigation tasks using the global appearance of visual information. In *VISAPP 2014, International Conference on Computer Vision Theory and Applications*. Ed. SciTePress - Science and Technology Publications ISBN: 978-989- 758-003-1 - Volume 1, pp. 194-201. Lisbon, Portugal.
- F. Amorós, L. Payá,O. Reinoso, W. Mayol-Cuevas, and A. Calway. (2013). Topological map building and path estimation using global-apperance image descriptors. In *10th International Conference on Informatics in Control, Automation and Robotics (ICINCO 2013)*. SciTePress - Science and Technology Publications, 2013, pp. 385-392. Reykjavik, Iceland.

- D.Valiente, A. Gil, F. Amorós, and O. Reinoso (2013). SLAM of view-based maps using SGD. In *10th International Conference on Informatics in Control, Automation and Robotics (ICINCO 2013)*. SciTePress - Science and Technology Publications, 2013, pp. 385-392. Reykjavik, Iceland.

- L. Payá, L. Fernández, O. Reinoso, F. Amorós, and L. Jiménez. (2013). A new resource in the teaching of a computer vision and robotics subject. In *7th International Technology, Education and Development Conference (INTED 2013)*. Ed. IATED, 2013, pp. 3074-3082. Valencia

- L. Payá, F. Amorós, O. Reinoso, L. Fernández, and A. Gil. (2013). An educational software to compare appearance image descriptors in robot localization. In *7th International Technology, Education and Development Conference (INTED 2013)*. Ed. IATED, 2013, pp. 3097-3105. Valencia.

- L. Payá, F. Amorós, O. Reinoso, and L. Jiménez. (2013). An educational software to develop robot mapping and localization practices using visual information. In *Advances in Control Education, vol. 10. Ed. International Federation of Automatic Control (IFAC)*, 2013, pp. 174-179. Sheffield, United Kingdom.

- F. Amorós, L. Payá, O. Reinoso, and L. Jiménez. (2012). Comparison of global-appearance techniques applied to visual map building and localization. In *International Conference on Computer Vision Theory and Applications (VISAPP 2012)*, vol. 2., SciTePress - Science and Technology Publications, 2012, pp. 395-398. Rome, Italy.

- L. Fernández, L. Payá, O. Reinoso, and F. Amorós. (2011). Appearance-based visual odometry with omnidirectional images. A practical application to topological mapping. In *8th Internacional Conference on Informatics, in Control, Automation and Robotics (ICINCO 2011)*. SciTePress - Science and Technology Publications, 2011. Noordwijkerhout, The Netherlands.

- F. Amorós, L. Payá, O. Reinoso, L. Fernández, and J. M. Marín. (2010). Visual map building and localization with an appearance-based approach - comparisons of techniques to extract information of panoramic images. In *7th Internacional Conference on Informatics, in Control, Automation and Robotics (ICINCO 2010)*. SciTePress - Science and Technology Publications, 2010, pp. 423-426. Funchal (Madeira).

Publicaciones en congresos nacionales

- F. Amorós, L. Payá, L. Fernández, O. Reinoso, M. Ballesta, and M. Juliá. (2013). Construcción de mapas topológicos y estimación de trayectorias usando descriptores de apariencia visual global. In *XXXIV Jornadas de Automática*, 2013, pp. 834-841. Terrassa.

- L. Fernández, L. Payá, M. Ballesta, F. Amorós, and O. Reinoso. (2012). Localización Monte Carlo a partir de la apariencia global de imágenes omnidireccionales. In *XXXIII Jornadas de Automática*. Ed. CEA-IFAC, 2012, pp. 743- 750. Vigo.

- F. Amorós, L. Payá, O. Reinoso, and L. Jiménez. (2012). Uso de descriptores de apariencia global en tareas de construcción de mapas y localización. In *XXXIII Jornadas de Automática*. Ed. CEA-IFAC, 2012, pp. 993- 1002. Vigo.

- L. Fernández, L. Payá, M. Ballesta, F. Amorós, and O. Reinoso. (2011). Odometría visual y construcción de un mapa topológico a partir de la apariencia global de imágenes omnidireccionales. In *XXXII Jornadas de Automática*. Sevilla.

- L. Fernández, L. Payá, M. Juliá, F. Amorós, and O. Reinoso. (2011).Visual odometry with an appearance-based method. In *ROBOT 2011. Robótica Experimental*. Sevilla: Ed. Universidad de Sevilla, 2011.

- F. Amorós, O. Reinoso, L. Payá, L. Fernández, and J. M. Marín. (2010). Construcción de mapas visuales y localización mediante métodos basados en apariencia global. In *XXXI Jornadas de Automática*. Ed. CEA-IFAC, pp. 31-39. Jaén

1.5 Estructura

El documento se estructura de la siguiente manera:

- El Capítulo 2 incluye un resumen del estado de la técnica actual, tratando las distintas plataformas robóticas según el medio de navegación, profundizando en aquellas cuyo área de actuación es terrestre. Además, el capítulo trata los distintos sensores que pueden ser utilizados, con interés especial en los sensores visuales, y en los trabajos que investigan la navegación visual.

- El Capítulo 3 presenta los sistemas de visión utilizados en este trabajo. Se profundiza en los conjuntos catadióptricos y en su modelo de proyección, incluyendo además distintas representaciones obtenidas a partir de la información visual mediante la proyección de los rayos en diferentes planos. A continuación, se describe un segundo sistema de visión equipado con una lente de ojo de pez, y las plataformas empleadas para la adquisición de las bases de imágenes usadas para la validación de los distintos algoritmos propuestos en la tesis.

- En el Capítulo 4 se expone una colección de técnicas orientadas a la descripción de información visual a través de su apariencia global. El análisis de los distintos métodos están orientados a la obtención de descriptores de imágenes obtenidas a partir de la información omnidireccional.

- El Capítulo 5 incluye los resultados experimentales de medición de la precisión y requisitos computacionales de los distintos descriptores para la estimación de la pose en un mapa denso de imágenes. Se describe en profundidad la base de imágenes utilizada, los resultados utilizando la distinta información de color de las escenas en los descriptores, y el comportamiento de los descriptores bajo ruido y oclusiones.

- En el Capítulo 6 se desarrolla el Análisis Multiescala, que utiliza distintas ampliaciones artificiales sobre las imágenes para extraer información sobre el desplazamiento espacial relativo entre escenas. Además, incluye un estudio sobre la utilización conjunta del Análisis Multiescala y la apariencia global sobre imágenes capturadas con una lente de ojo de pez, y sobre imágenes omnidireccionales. Por último, el capítulo presenta la aplicación de estos métodos sobre dos entornos reales distintos en tareas de navegación.

- El Capítulo 7 introduce un conjunto de técnicas para estimar la diferencia de altura topológica entre escenas usando la apariencia global a partir de imágenes omnidireccionales. En el capítulo se presenta una base de imágenes capturadas para llevar a cabo distintos experimentos que permiten comparar la precisión de las distintas técnicas de estimación de altura.

- Por último, el Capítulo 8 recoge las principales conclusiones de la tesis, detallando las contribuciones obtenidas en el desarrollo del trabajo y las futuras líneas de investigación.

UNIVERSITAS
Miguel
Hernández

CAPÍTULO

2

Estado de la Técnica

La navegación robótica es una línea de investigación en constante desarrollo y actualización. Debido al gran abanico de posibilidades que ofrece tanto por entornos de navegación, sensores y técnicas específicas para recoger e interpretar la información que llega al robot del entorno que le rodea a través de sus sensores, existen muchas posibilidades de actuación que permiten optimizar esta tarea.

A continuación se presenta un resumen que trata de recoger de forma global el estado de la técnica actual.

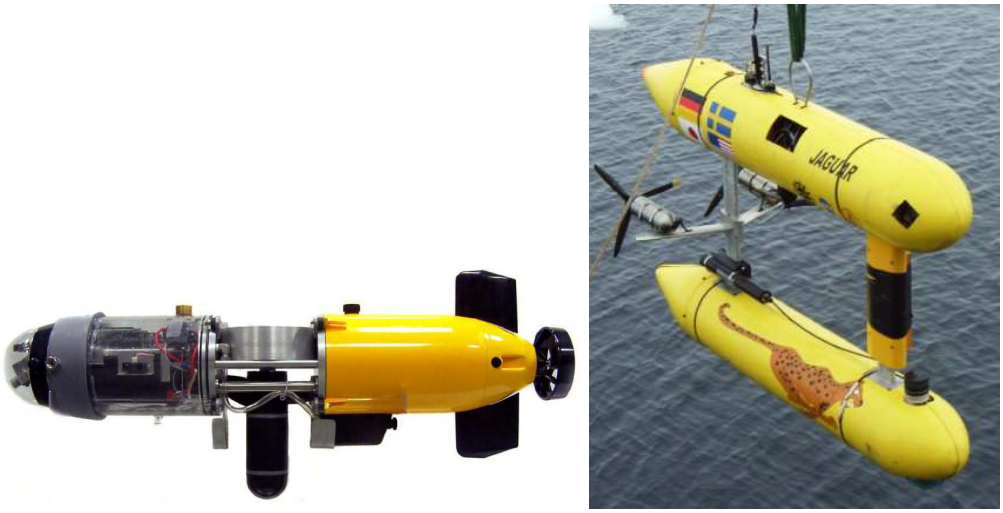
2.1 Entornos de Navegación

Son numerosos los entornos en los que un robot puede actuar. Entre las distintas clasificaciones que se pueden realizar, es posible dividir los robots según el medio por el que naveguen. Así pues, es posible encontrar plataformas de navegación submarina, aéreas o terrestres.

2.1.1 Robots Submarinos

Los robots capaces de navegar en entornos submarinos, también conocidos por sus siglas en inglés AUVs (*Autonomous Underwater Vehicles*) pueden ser empleados como plataforma de ayuda en tareas asociadas a biología [6, 150], química [64] o geología marina [92, 175].

La imposibilidad de utilizar sistemas de posicionamiento como el GPS debido a la gran atenuación del campo electromagnético en el agua provoca que, en la mayoría de aplicaciones de navegación submarinas, la estimación de la posición de los AUVs se convierta en uno de los principales problemas a resolver. Además, hay que tener en consideración que



(a) (b)

Figura 2.1: Ejemplos de Vehículos Autónomos Submarinos: (a) USAL y (b) Puma.

normalmente la escala del entorno donde se mueven los robots submarinos es grande, tanto temporal como espacialmente. En la Figura 2.1 es posible ver dos ejemplos de AUVs.

En este tipo de entornos, es común el uso de sistemas de radio balizas que emiten a determinadas frecuencias, permitiendo la estimación de la posición a través de triangulación usando distintos puntos. Los sistemas de posicionamiento LBL (*Long Baseline*) usan transponedores situados en posiciones conocidas para el robot, normalmente fijas. Tal y como exponen Kunz et al. [106], si el robot está equipado con sensor de profundidad, únicamente es necesario el uso de dos transponedores para restringir la posición del móvil totalmente. Independientemente de este hecho, en la parte experimental de su trabajo usan 4 balizas en distintas posiciones, obteniendo información redundante.

Otros sistemas de posicionamiento a través de señales acústicas son los USBL (*Ultra Short Baseline*), también conocidos como SSBL (*Super Short Baseline*), y los SBL (*Short Baseline*). Estos sistemas difieren del LBL en su forma de montaje y rango de actuación. Mientras que los sistemas LBL usan transponedores situados con una separación que varía de aproximadamente 100 metros a varios kilómetros, los sistemas SBL suelen ir colocados en plataformas o barcos, espaciados entre 10 y 50 metros entre sí, y los USBL se sitúan en el orden de las decenas de centímetros de separación [207]. A diferencia de los sistemas LBL y USBL, el propósito de los USBL es medir la distancia con el robot y su dirección, sin determinar su posición por triangulación.

Podemos encontrar trabajos que combinan los distintos sistemas. En [19], sus autores combinan un sistema LBL y USBL junto con un giróscopo para estimar la posición de un

AUV. En [98] se combina un sistema LBL y otro SSBL, consistiendo este último en una transponedor fijo situado en la superficie del fondo marino.

Entre los sensores con los que las plataformas robóticas de navegación submarina pueden ir equipadas, se encuentran medidores de temperatura, conductividad, presión, giróscopos, medidores de velocidad basados en dopplers (DVLs o *Doppler Velocity Logs*), sónar o cámaras [106].

En [97], sus autores presentan un sistema de navegación que combina la información proporcionada por varios sensores mediante el uso de un filtro EKF (*Extended Kalman Filter*) y un filtro UKF (*Unmaned Kalman Filter*). En concreto, esos sensores son un DVL y un sistema de navegación incercial (INS). Por otro lado, en [119], Mahon y Williams proponen el uso conjunto de la información proporcionada por un sónar y un sistema de seguimiento de puntos característicos de imagen, basado en el algoritmo de Lucas-Kanade, para tareas de SLAM utilizando AUVs.

El uso de sistemas de visión en estos entornos es especialmente complicado debido a las particularidades propias del medio. En concreto, aparecen distorsiones, cambios de iluminación y partículas suspendidas que afectan a las imágenes obtenidas por el sensor.

Sun et al. muestran en [180] distintos métodos para atenuar el efecto de los cambios de iluminación y para mejorar el efecto de bajo contraste de la imagen derivado de la propagación de la luz en el agua. En [44] se presenta un algoritmo de construcción de mapas topológicos y odometría visual para sistemas de navegación submarinos, que utilizan la extracción de puntos característicos con SIFT y la estimación de la pose usando la homografía de la cámara.

2.1.2 **Robots Aéreos**

En los últimos años, los robots aéreos o UAVs (*Unmanned Aerial Vehicles*) están sirviendo de plataforma para numerosas investigaciones en el campo de navegación autónoma.

Por el número de hélices con los que está equipado, es posible encontrar ejemplos de helicópteros [181], cuadricópteros [4] u octocópteros [53]. En la Figura 2.2 podemos ver algunos ejemplos.

Dependiendo de su tamaño, también podemos hablar de SAVs (Small Aerial Vehicles), que van desde las decenas de centímetros a los 200 cm, como el usado en [47], o de MAVs (Micro Aerial Vehicles). En [203] se presenta un vehículo aéreo clasificado como MAV menos de 22 gramos incluyendo las baterías, con una envergadura de alas de 7 cm.

Estos vehículos están equipados normalmente con GPS y sensores inerciales, como giróscopos o acelerómetros. También suelen incluir sensores de presión barométricos Sin embargo, tal y como exponen Conte y Doherty en [40], el uso de GPS, especialmente cuando opera en

(a) (b) (c)

Figura 2.2: Ejemplos de Vehículos Autónomos Aéreos: (a) Helicóptero, (b) Cuadricóptero y (c) Octocóptero.

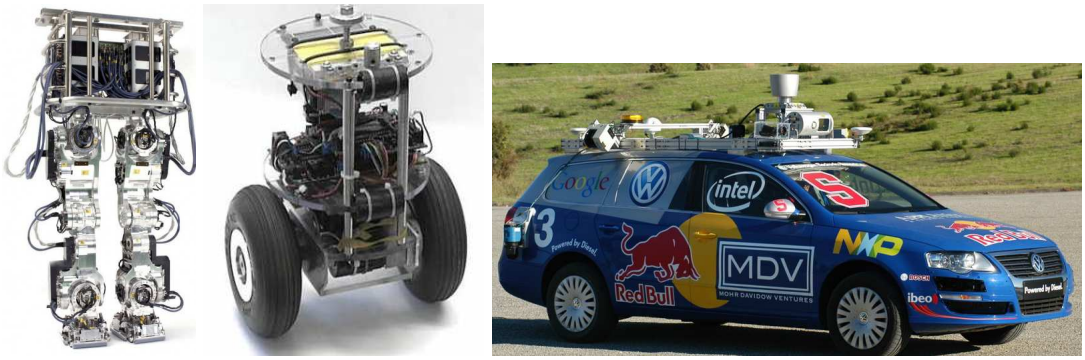
frecuencias civiles, puede volverse poco fiable al aproximarse a obstáculos o navegar entre edificios. Por otro lado, los sensores inerciales que proporcionan la odometría interna del robot pueden introducir desviaciones que, si no se corrigen, acumulan un error en la estimación de la posición que puede volverse inasumible en tareas de navegación real.

Gran parte de la investigación actual trata de evitar la dependencia en navegación de UAVs con estos sensores. Una posible solución está constituida por los sistemas de visión. Su bajo peso y reducido consumo energético los convierten en una opción relevante entre las muchas disponibles.

En el trabajo desarrollado [40] se incluye un sistema que integra la información de los sensores inerciales y odometría visual. En concreto, la información visual es utilizada con dos propósitos distintos: el primero de ellos es obtener información odométrica mediante el algoritmo de seguimiento de características KLT (Kanade-Lucas-Tomasi [185]), que se combina con un conjunto de transformaciones lineales para hallar la diferencia en la pose entre dos imágenes consecutivas. El segundo propósito es la asociación con imágenes capturadas anteriormente, cuyas posiciones son conocidas, para mejorar la localización del robot. Para ello, el algoritmo emplea el detector de bordes Sobel, realizando la asociación de imágenes a través de la medición de píxeles solapados entre las imágenes de referencia y las imágenes capturadas durante la navegación. La información visual e inercial se fusiona mediante un filtro de Kalman. Los resultados experimentales comparan el sistema propio del UAV (basado en la información inercia y el GPS) y el desarrollado por ellos cuando se produce un fallo en el GPS, demostrando que la navegación no puede ser confiada únicamente al INS.

Tal y como se remarca en este mismo trabajo, la disposición de imágenes satélites de alta resolución, como las proporcionadas por Google Earth, hacen que la idea de mejorar la localización mediante asociación con imágenes geolocalizadas previamente, sea una opción a ser considerada en el futuro próximo.

En [133], Mondragón et al. muestran la utilización de distintos sistemas visuales (cámara catadióptrica y par estéreo) para la estimación de la inclinación, altura y movimiento de un



(a) (b) (c)

Figura 2.3: Ejemplos de distintas plataformas robóticas terrestres: (a) plataforma bípeda, (b) robot con dos ruedas y (c) coche con sistema de conducción autónoma.

UAV, integrando la información en un sistema de control visual.

Por otro lado, en [101, 140] se incluyen dos trabajos distintos de SLAM visual orientados a vehículos autónomos aéreos.

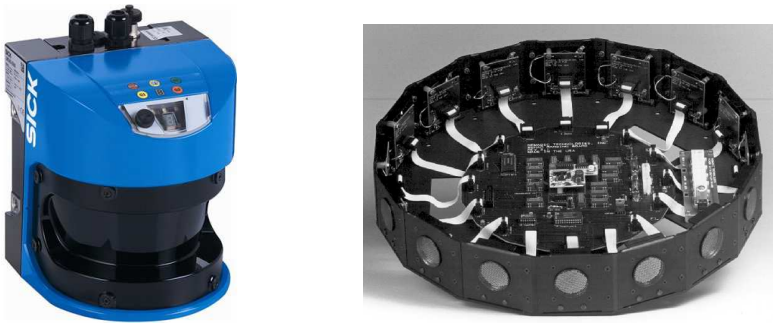
2.1.3 Robots Terrestres

Sin duda alguna, los vehículos terrestres son las plataformas robóticas que destacan por ser las más utilizadas en el campo de la navegación autónoma. Este tipo de vehículos puede ser encontrado bajos las siglas AGVs (*Autonomous Ground Vehicles*) o UGVs (*Unmanned Ground Vehicles*), aunque no es muy usual. Resulta complicado hacer una clasificación que recoja todas las implementaciones posibles que se pueden encontrar actualmente.

Dentro de su morfología, es posible encontrar robots equipados con dos ruedas [141], todo-terreno con ruedas de oruga [36], bípedos [96], con forma de serpiente que basan su movimiento en el deslizamiento [112], y un largo etcétera que trata de adaptarse a las distintas necesidades a las que se aplican los robots terrestres. En [81], sus autores muestran la adaptación de un automóvil para recorrer rutas fuera de carreteras, mientras que [107] presenta un sistema basado principalmente en visión aplicado a un vehículo para navegación en entornos ciudad.

Otra clasificación puede realizarse según el entorno de navegación, con ejemplos de interior o de exterior. Las distintas condiciones de terreno, iluminación, distancia a objetos,... condicionan la geometría y características de los robots, al igual que los sensores que se pueden utilizar en cada caso.

La odometría interna del robot depende en la mayoría de los ejemplos de encoders asociados a sus ruedas. Esta odometría es sólo precisa en desplazamientos cortos. El deslizamiento de las ruedas, por ejemplo, introduce un error que es acumulativo. Tras navegar un cierto



(a) (b)

Figura 2.4: Imágenes de sensores (a) Laser y (b) Sónar.

tiempo, el error de la odometría suele ser inadmisible en aplicaciones reales. Por ello, es común el uso combinado o sustitutivo de otros sensores que permitan medidas de mayor precisión.

Por ejemplo, los sistemas de posicionamiento global (GPS) constituyen una buena opción para entornos de exterior. Sin embargo, tal y como se ha combinado anteriormente, en aplicaciones en las que se navegue cerca de edificios o de interior, no es un sistema fiable.

Durante mediados de los años noventa y la década pasada, la mayoría de los trabajos utilizan como sensores principales en sus trabajos sónar o láser para recibir la información de su entorno.

Los sensores sónar, cuyo nombre proviene de las siglas inglesas SONAR (*Sound Navigation and Ranging*), son sensores de bajo coste que permiten mediciones de distancias con los objetos situados a su alrededor mediante la emisión de pulsos sonoros y la recepción de los ecos de dichos pulsos. En [195], Wijk y Christensen presentan una técnica llamada TBF (*Triangulation-based Fusion*) para conseguir mejorar la detección de puntos del entorno mediante la búsqueda de intersecciones entre múltiples mediciones de los sónares, adquiridas a medida que el robot va avanzando. En [182] se expone otro trabajo de construcción de mapas y localización usando los datos del sónar, que trata de hacer más robustas las asociaciones llevadas a cabo mediante la identificación de líneas rectas y esquinas en el entorno.

Sin embargo, la precisión de estos sensores es relativamente baja, debido sobre todo a la alta incertidumbre angular, al ruido introducido por reflejos de las señales acústicas (como se indica en [38]) y a la correlación de sus mediciones con la temperatura ambiente.

Por otro lado, los sensores láser presentan características distintas a los sónar. Estos equipos determinan la distancia a partir de la medición del tiempo de vuelo de un pulso láser al ser reflejado en los objetos. Tienen una precisión mucho mayor a los sónar, recogiendo la distancia de objetos que se encuentran desde las decenas de centímetros hasta los 80 metros, con una precisión que se sitúa en el orden de los milímetros. Además, la resolución angular

queda por debajo de 1º. Sin embargo, el precio y peso de estos equipos son considerables, además de tener un consumo energético elevado, aspecto crucial para la autonomía del robot durante las tareas de navegación.

Son numerosos los trabajos de robótica móvil en los que podemos encontrar la utilización de sensores laser. En [108, 208] se pueden encontrar ejemplos de creación de mapas 2D, mientras que [186, 206] incluyen ejemplos de mapas 3D. En [78] se presenta una aplicación de creación de navegación exterior que incluye la creación de mapas 3D a partir del uso de un laser con -10º de inclinación hacia el suelo con respecto a la horizontal del suelo y la odometría del robot.

También es posible encontrar trabajos que hacen uso de sensores de infrarrojo en tareas de navegación. El rango de detección de objetos con este sistema varía de 10 a 80 cm aproximadamente. Por ejemplo, en [35], Chen y Song muestran un sistema de control de movimiento para un robot aspirador. Su sistema emplea además una marca artificial basada en un emisor de infrarrojos situada en el entorno para mejorar la localización del robot y ofrecer una navegación más eficiente.

Como opción a estos sensores, en los últimos años existe un interés creciente en el uso de sensores visuales como fuente de información al robot. Las cámaras presentan características que los convierten en una buena alternativa frente a los sensores sónar y láser. Entre esas características, es posible destacar:

- Proporcionan una información tridimensional de gran riqueza.
- Son relativamente más económicos.
- Su peso y consumo energético es menor.
- Existen múltiples posibilidades de configuración que permiten, por ejemplo, capturar información omnidireccional alrededor del sensor, o extraer información de profundidad.

Los sistemas visuales pueden sustituir a otros sensores del robot, así como complementar su información. En [37], se introduce un sistema que combina el sónar con la información visual en una aplicación de navegación robótica, o [76], que emplea un láser y junto con un sistema de visión. Otro ejemplo es el introducido en [34], en cuyo trabajo aparece un sistema laser-visión compuesto por un proyector de líneas laser y una cámara CCD que ayuda a evitar obstáculos y a buscar relaciones entre los puntos extraídos por el sensor visual y el láser. La información es complementada además por una cámara situada en el techo que permite la localización global del móvil en el entorno de navegación.

La gran riqueza de la información visual permite numerosas posibilidades de descripción de las escenas. En los siguientes puntos se incluyen el estado de la técnica de algunos sensores visuales y se detallan técnicas utilizadas obtener descriptores que permiten concentrar la información recogida en la escena.

2.2 Sensores visuales

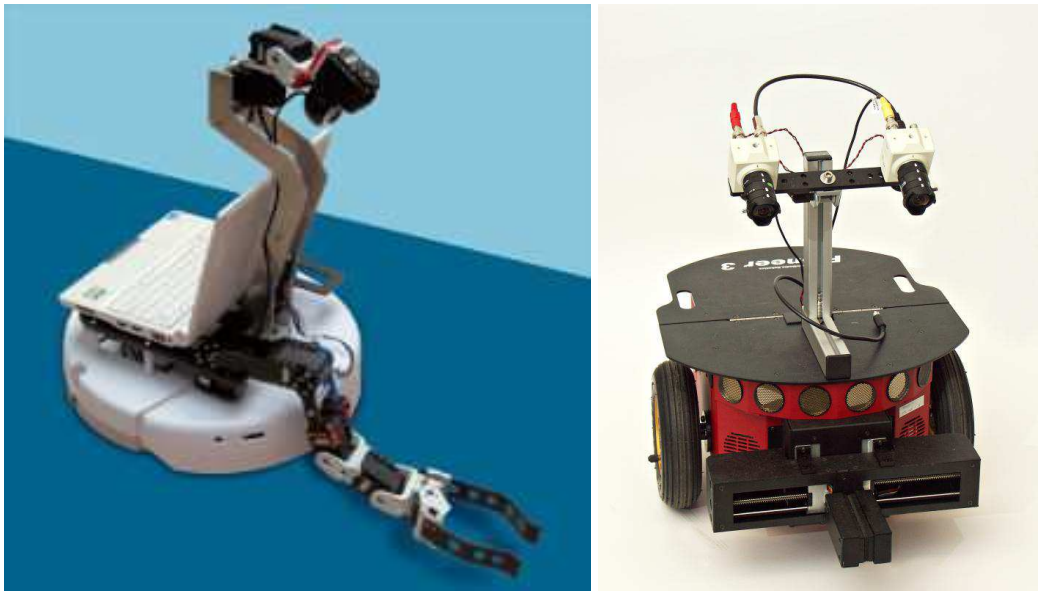
Como se ha indicado anteriormente, las múltiples posibilidades de representación de la información visual deriva en una gran cantidad de sensores visuales diferentes. Es posible clasificar los sensores entre aquellos que proporcionan información en color o en escala de grises, según la resolución proporcionada, o según la tecnología del sensor (CCD, CMOS).

Sin embargo, este punto se va a centrar en la clasificación de los sensores visuales según el número de cámaras utilizadas, y por campo visual ofrecido. En este sentido, podemos encontrar sistemas que utilizan una sola cámara [143, 210], pares estereoscópicos [29, 204], sistemas trinoculares [89], y sistemas omnidireccionales, que pueden estar formados por diversas cámaras que enfocan en distintas direcciones [24], o sistemas catadióptricos compuestos por una cámara y una superficie reflectante [138]. La Figura 2.5 recoge ejemplos de distintos sistemas visuales.

Los sistemas binoculares tratan de imitar el sistema de visión humano, que permite realizar mediciones de profundidad del entorno, sin necesidad de entrar en contacto físicamente con los objetos. Un objeto tridimensional tiene un número infinito de posibles vistas 2D resultantes de las infinitas posiciones desde las que se puede observar ese objeto. Sin embargo, si la observación se realiza por dos sensores visuales al mismo tiempo, obtenemos información de su forma tridimensional y posición. Una vez que un punto es identificado en las dos imágenes capturadas por los dos sensores visuales, se puede obtener información de profundidad de forma directa a través de triangulación usando la geometría epipolar. La distancia entre la posición de un mismo punto entre las dos imágenes capturadas por un par estéreo es conocido como disparidad, que es función directa de la posición de 3D del punto y la posición, orientación y características físicas de las cámaras.

Tal y como indican Brown et al. [29], la visión estéreo fue un área de máximo desarrollo durante los años 70 y 80. Durante los años 90, era ya un área madura, y a partir de ese momento se ha tratado de resolver problemas de mayor nivel, con algoritmos que permiten asociaciones entre escenas mucho más robustas, o aplicaciones en tiempo real.

Parte de estos nuevos avances son los sistemas trinoculares. Los sistemas trinoculares se basan también en la visión estereoscópica. El uso de una tercera cámara incrementa las res-



(a) (b)

(c)

(d) (e)

Figura 2.5: Imágenes de distintos sistemas visuales. (a) Robot equipado con una sola cámara, (b) Robot con un par estéreo, (c) Sistema de visión trinocular, (d) imagen de coche equipado con array de cámaras para obtener imagen omnidireccional y (e) robot con sistema de visión catadióptrico.

tricciones geométricas que se obtienen entre las imágenes capturadas en un mismo instante, con lo que se aumenta la precisión en la estimación de la información tridimensional.

Por otro lado, los sistemas omnidireccionales son capaces de ofrecer información de un gran campo de visión alrededor de la cámara. Dentro de este tipo de cámaras, es posible encontrar sistemas compuestos por múltiples cámaras, como el de la Figura 2.5(d), o sistemas catadióptricos (Figura 2.5(e)).

Si el sistema es capaz de recoger información en cualquier dirección o casi en su totalidad, también puede ser denominado cámara esférica. Por ejemplo, la empresa *Point Grey* comercializa las cámaras *Ladybug* [1], capaces de recoger hasta el 90 % de la esfera completa a su alrededor.

Dentro de los sistemas que recogen un campo visual amplio del entorno, es posible incluir las cámaras que emplean lentes de ojo de pez [75, 197]. Este tipo de lentes presentan un campo de visión que puede ser superior a 180º. En [111], Li et al. presentan un sistema de visión equipado con dos cámaras con lente de ojo de pez que consiguen capturar la esfera completa a su alrededor. Sin embargo, el propio Li remarca en un trabajo posterior [110] las desventajas de usar dos cámaras distintas para crear la imagen esférica, debido a que las condiciones de iluminación de cada cámara pueden ser distintas, y por lo tanto, el comportamiento del sensor puede variar. Por ello, en ese trabajo propone un nuevo sistema bajo el nombre de *EspheEye* que consigue obtener información de la esfera completa con una única cámara. Para lograr su objetivo, hace uso de un espejo que refleja la información de rayos provenientes de dos lentes de ojo de pez que se encuentran en el mismo eje pero en direcciones opuestas. El estudio incluye un método para calibrar el sistema visual. Por la utilización del espejo, podría considerarse un sistema catadióptrico. No tenemos constancia de la comercialización de este sistema visual.

No obstante, cuando hagamos mención en este trabajo a un conjunto catadióptrico, nos referiremos a un sistema formado por una cámara que enfoca a una superficie de revolución reflectante. Tal y como se detalla en el Capítulo 3, la geometría de esta superficie de revolución puede ser diversa, y condiciona las características de la escena obtenida.

Los sistemas catadióptricos pueden recoger la información visual alrededor de los 360º alrededor del eje de revolución del espejo. Sin embargo, su configuración limita el ángulo de visión en el ángulo lateral, debido a la oclusión de la cámara y del propio espejo. No obstante, estos sistemas son más sencillos de calibrar que los basados en lentes de ojo de pez presentados anteriormente ([110, 111]), y considerablemente más asequibles que las cámaras esféricas, del orden de la décima parte en precio.

Por último, cabe destacar que los sistemas omnidireccionales también pueden ser utilizados en configuraciones que permitan obtener información estereoscópica. Como ejemplo, en

[73], Goto et al. presentan un sistema de navegación que utiliza dos sistemas catadióptricos montados en el robot.

2.3 Descripción de la información visual

La gran riqueza de la información visual se traduce también en importantes requerimientos de memoria para poder almacenar las imágenes, además de la necesidad de computación derivada del procesamiento de las escenas

En tareas de navegación en tiempo real, esta cantidad de información se vuelve inmanejable si se utiliza de forma directa. Por ello, es necesario buscar descriptores que permitan reconocer las escenas y diferenciarlas de entre un conjunto de escenas, utilizando una cantidad limitada de datos. En tareas de navegación, las características de un buen descriptor son:

- Tiempo de cálculo reducido.
- Información recogida en pocos términos.
- Capacidad de distinguir la escena.
- Invariancia a rotaciones.
- Contener información relativa a orientación para la estimación del desfase respecto a otras escenas.

Es posible dividir los descriptores en dos grandes categorías: aquellos basados en segmentación o extracción de características, y los basados en la apariencia global de la escena.

En los descriptores basados en características, la información visual se extrae en forma de marcas (o *landmarks*) visuales. Estas marcas son características del entorno fácilmente detectables y reconocibles por el robot. Dependiendo de la naturaleza de las marcas, podemos dividirlas en naturales o artificiales.

Las marcas artificiales, como su nombre indica, son añadidas por el hombre en el entorno de navegación a fin de ser reconocidas por el robot. Estas marcas pueden tomar forma, por ejemplo, de etiquetas, códigos de barras, o códigos QR, que el robot puede reconocer e interpretar. Con ellas, la localización del robot se simplifica considerablemente. En [144], se introduce una aplicación que hace uso de marcas visuales artificiales para corregir la localización del robot durante la navegación en entornos de interior. Las *landmarks* están constituidas por códigos QR.

Por otro lado, la categoría de marcas naturales está formada por todas aquellas que no han sido puestas en el entorno específicamente para la tarea de navegación. Por ejemplo, en entornos de interior, estas marcas suelen corresponder a zonas naturalmente identificables, tales como esquinas, puertas, o ventanas.

Todos los descriptores basados en extracción de características siguen dos pasos básicos:

1. Detección de los puntos o marcas destacables de la imagen que serán utilizados como referencia.

2. Asociación de un vector de características que se calculará usando información local, usado como descriptor de la marca visual.

Existen numerosos estudios que tratan sobre este tipo de marcas, tratando de hallar formas cada vez más eficientes de extraer *landmarks* naturales de mayor calidad para aplicaciones de navegación robóticas. Las características deseables de estas marcas son:

- Ser fácilmente identificables, lo que significa que pueda ser visible desde distintas posiciones en el entorno.

- Ser estables, es decir, visible en distintas capturas de imagen consecutivas.

- Ser fácilmente reconocibles, o dicho de otra forma, deben ser descritas de forma que sea distinguible entre otras características. El descriptor debe permitir reconocer la marca desde los distintos puntos de vista.

- La marca debe ubicarse con precisión en el entorno.

A continuación se describen brevemente algunos de los detectores de puntos más utilizados:

- **Detector de esquinas de Harris**: Este detector es uno de los más usados para la extracción de puntos en imágenes. Aprovecha el cambio en múltiples direcciones que se dan en las esquinas dentro de una escena, frente a la ausencia de variación en las zonas lisas [77]. El algoritmo busca los puntos de la imagen donde se cruzan distintos gradientes en direcciones perpendiculares.

- **Harris-Laplace**: Este método es una extensión del detector de Harris, que trabaja con transformaciones afines de la imagen, permitiendo la detección de los puntos de las esquinas de una imagen a pesar de cambios en la escala de la escena [128].

- **SUSAN**: SUSAN corresponden a las siglas inglesas de *Smallest Univalue Segmente Assimilating Nucleus* [178]. El detector genera una máscara circular alrededor de cada punto de la imagen, y compara la intensidad de los píxeles vecinos con centro en el pixel central, que es el núcleo de la máscara. El pixel central será considerado como esquina dependiendo del número de puntos con intensidad similar al núcleo.

- **SIFT**: El algoritmo SIFT (o *Scale-Invariant Feature Transform*) [115] detecta puntos característicos en las escenas usando diferencias de Gaussianas (DoG) en distintas escalas espaciales. Los puntos seleccionados se corresponden con los extremos locales de la función DoG. Posteriormente, se calcula un descriptor para cada punto, basado en información local de la imagen.

- **SURF**: El detector SURF, siglas de *Speeded Up Robust Features* [20], está inspirado en SIFT, aunque mejora los requerimientos computacionales de este descriptor. SURF utiliza el determinante de la matriz Hessiana. Las características obtenidas son invariantes a escala y rotación.

- **MSER**: Este algoritmo fue introducido por Matas et al. [124] bajo el nombre *Maximally Stable Extremal Regions*. Extrae regiones de la escena con un método similar al de segmentación de Watershed [137].

Como se ha comentado anteriormente, tras detectar los puntos hay que asociarles un descriptor que nos permita reconocerlos. Estos descriptores son los que caracterizan e identifican a los puntos extraídos en las imágenes. A continuación, se incluyen algunos de los descriptores más utilizados:

- **SIFT**: La transformada SIFT asigna una orientación global a cada punto basado en el gradiente de direcciones de una región local de la imagen. Después, el descriptor se calcula utilizando los histogramas de orientación en subregiones alrededor del punto de interés. Para obtener invariancia ante iluminación, el descriptor se puede normalizar dividiendo por la raíz cuadrada de la suma de sus componentes al cuadrado.

- **GLOH**: Las siglas GLOH significan *Gradient location-orientation histogram*, y es una extensión del descriptor SIFT, diseñado para incrementar la robustez y distinción de los puntos con respecto a este algoritmo. El algoritmo aplica el análisis PCA para reducir la dimensionalidad del descriptor resultante. En [129] se puede ver una comparación entre diferentes descriptores, incluyendo GLOH.

- **Subventana de niveles de gris**: Para describir los puntos característicos, trabajos como [46] utilizan los valores de gris de una región limitada alrededor del punto de interés.

- **Histogramas de Orientación del Gradiente**: Como su nombre indica, estos histogramas recogen información relativa a la orientación del gradiente de la escena. Para cada pixel, se calcula el módulo y la orientación del gradiente. Posteriormente, se crea un histograma en el cual la orientación es el eje de abscisas, y en el de ordenadas se computa el número de píxeles de cada orientación ponderado por el módulo del gradiente.

En trabajos como [70, 130], sus autores incluyen comparación de estos detectores y descriptores entre otros.

Pero no sólo se pueden encontrar técnicas basadas en características relacionadas con la extracción de puntos. Son numerosas los trabajos que tratan de reconocer líneas y contornos en una imagen.

Por ejemplo, la transformada de Hough es un algoritmo de extracción de características muy usado en el análisis de imágenes para encontrar ciertas clases de objetos. La idea básica es encontrar geometrías que puedan ser parametrizadas, tales como líneas rectas, círculos, o polinomios de distintos grados. El algoritmo realiza un mapeado de la imagen y contabiliza los píxeles que pueden pertenecer a una cierta geometría con un sistema de votación. Este método de detección de geometrías es independiente a cambios de escala, rotación o traslación. En [18, 51] se muestran dos trabajos que aplican la transformada de Hough para detectar distintas formas en la escena (Transformada de Hough generalizada).

Otras soluciones posibilitan encontrar distintas geometrías en una escena son, por ejemplo, Canny [31] o RANSAC (*RANdom SAmple Consensus*) [59].

En [167], Scaramuzza et al. presentan un algoritmo que constituye un método robusto de extracción y caracterización de líneas verticales en imágenes omnidireccionales. Este algoritmo parte del cálculo de las derivadas en el eje horizontal y vertical de la imagen para, a través de la dirección del gradiente, detectar las líneas radiales de la imagen omnidireccional, que equivale a la detección de las líneas verticales en la imagen panorámica.

Posteriormente, asocia a cada línea un descriptor basado en el histograma de orientación de los píxeles adyacentes de las líneas en áreas circulares. En concreto, se divide la longitud de cada línea radial en tres partes, y se superponen tres círculos sin solapamiento a lo largo de la línea, cuyo centro queda situado en la línea identificada. En la Figura 2.6 se muestra un esquema del proceso realizado por el algoritmo para la extracción de las líneas verticales sobre la imagen omnidireccional.

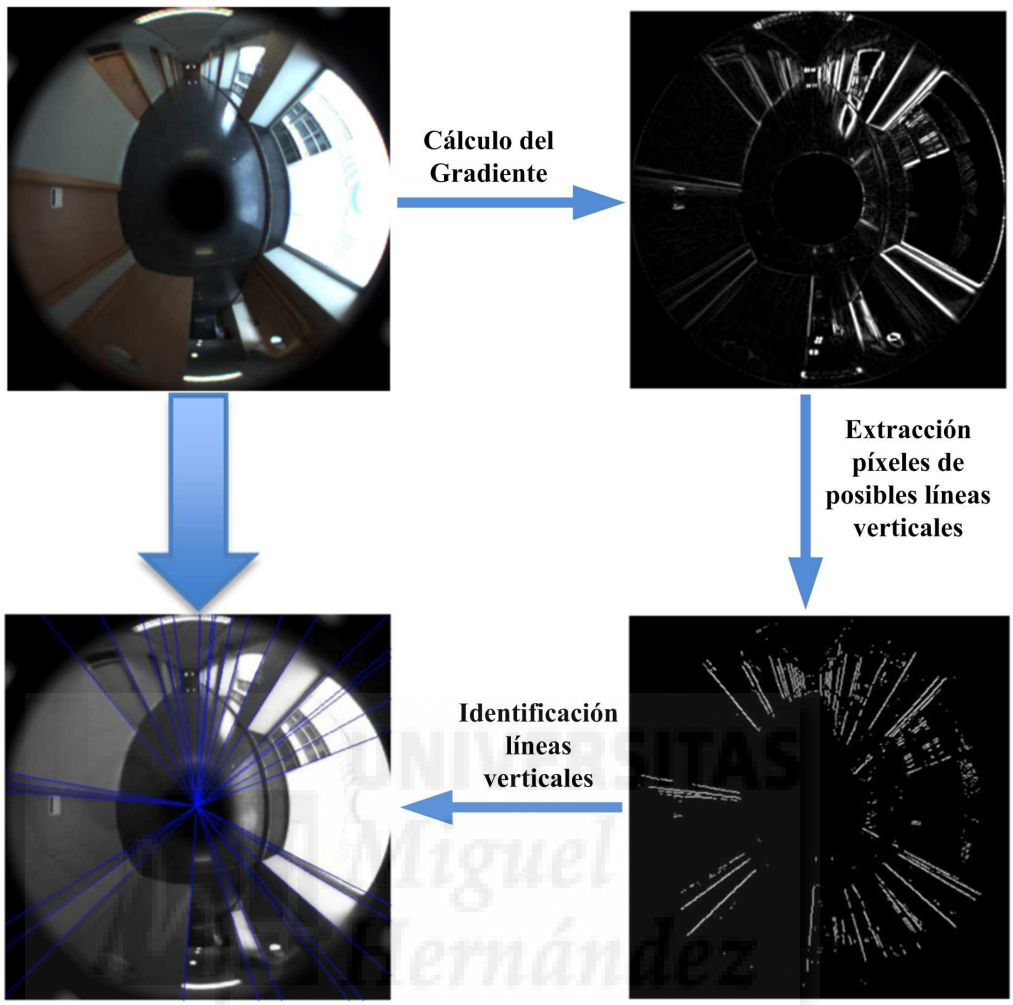


Figura 2.6: Esquema del proceso de extracción de lineas verticales sobre imagen omnidireccional.

El inconveniente de estas técnicas basadas en extracción de puntos notables o segmentación recae en su gran complejidad, debido a la dificultad en la extracción de características estables y comparación de patrones en entornos reales y cambiantes. Además, pueden surgir problemas adicionales en entornos no estructurados, en los cuales no se pueden colocar marcas artificiales y no existen marcas naturales que puedan ser localizadas con precisión.

Por ello, en la última década se ha desarrollado una familia de métodos alternativos que tratan de crear un mapa del entorno usando la información de las imágenes en su conjunto, sin tomar en consideración patrones locales ni marcas características estructurales. Estos son los llamados métodos basados en apariencia.

Las imágenes son memorizadas sin ninguna extracción de características previa, y el reconocimiento se consigue mediante la comparación de imágenes, bien directamente o bien previa transformación matemática, pero siempre basado en su apariencia global. De este modo, se logra un método útil en escenas complejas en el mundo real, donde es difícil crear

modelos adecuados para reconocimiento.

Actualmente en robótica, las técnicas de aprendizaje y reconocimiento basadas en apariencia global están siendo tenidas muy en cuenta en la autolocalización aplicada a la navegación robótica, utilizando únicamente cámaras como sensores. Como se ha comentado, los métodos basados en apariencia global de la escena tienen una ventaja importante, que es la simplicidad de los algoritmos, ya que no se realiza la búsqueda de correspondencias ni características locales, como se hace en otros métodos. A su vez, la construcción de mapas del entorno basados en apariencia es mucho menos compleja que la reconstrucción a partir de la posición de marcas en el espacio 3D.

Sin embargo, puede surgir el inconveniente de almacenamiento de los datos, ya que se trabaja con toda la información proporcionada por el sensor de visión, de manera que se tenga que manejar gran cantidad de datos al estudiar las imágenes al completo.

De este modo, el estudio de los descriptores basados en apariencia global se centra en el tipo, forma y cantidad de información almacenada a partir de las imágenes. También debe considerarse cómo se lleva a cabo la comparación entre la vista actual y las almacenadas para estimar la pose del robot con precisión, pero dentro de unos límites de tiempo y requerimientos de memoria.

En el Capítulo 4 se presenta un estudio detallado de los descriptores basados en apariencia utilizados en este trabajo.

2.4 Navegación y Creación de Mapas

La navegación robótica puede ser descrita como el proceso de determinar una ruta segura entre dos puntos por la que el robot pueda navegar a un destino. De nuevo, son numerosos los sensores disponibles para esta tarea, así como las posibles soluciones. Dentro de las tareas de navegación, es posible considerar tres problemas fundamentales:

- ¿Dónde estoy?
- ¿Cómo es mi entorno?
- ¿Qué trayectoria debo seguir para llegar a un punto?

Estas preguntas dan lugar a tres áreas distintas de investigación en navegación, que pueden ser tratadas por separado, o de forma conjunta. Estas áreas son:

- **Localización**: La localización es el proceso mediante el cual el robot determina su pose en el entorno, y se lleva a cabo mediante la comparación de la información actual

proporcionada por los sensores del robot con información recogida previamente. Esta información puede estar ordenada en forma de mapa, aunque también podemos considerar la localización utilizando únicamente la información de la pose inmediatamente anterior, sin necesidad de mapa.

- ***Mapping***: El robot puede construir un mapa que represente el área de navegación. El proceso de *mapping* hace referencia precisamente a la interpretación de la información recibida por los sensores para describir su entorno. Esta tarea puede ser realizada por uno [201] o varios robots [162].

- ***Path Planning***: El *Path Planning* o estimación de la trayectoria determina el camino que debe seguir el robot para alcanzar una determinada posición desde su pose actual. La trayectoria calculada puede estar condicionada por distintos criterios de optimización diseñados por el usuario en función de la tarea a realizar o las condiciones del entorno.

Como se ha comentado anteriormente, estos conceptos pueden ser tratados de forma conjunta, ya que están intrínsecamente relacionados dentro de la navegación. En tareas reales de navegación, es muy complicado entender la planificación de trayectorias sin conocer la posición actual el robot, es decir, sin llevar a cabo un proceso de localización. De igual forma, la creación de un mapa suele requerir de la correcta estimación de la pose del robot dentro del mismo mapa para ser capaz de integrar la nueva información recibida por los sensores del robot en el mapa. La Figura 2.7 muestra un esquema con las relaciones entre las distintas tareas.

Se conoce como Localización Activa a la unión de la localización y la planificación de la trayectoria. En [83], sus autores describen un algoritmo de localización activa para un robot equipado con un sensor visual catadióptrico y un sensor de infrarrojos. El sensor de infrarrojos, además de ser utilizado para la localización, se utiliza para evitar obstáculos en la trayectoria. Los experimentos se llevan a cabo en entornos estructurados. La estimación de la ruta se hace buscando el camino más corto en un grafo cuyos nodos están formados por las imágenes omnidireccionales.

Por otro lado, la combinación de *Path-Planning* más mapping se denomina exploración. La exploración trata de optimizar la ruta del robot o conjunto de robos de tal forma que se realice una navegación eficiente, tratando de recoger la máxima información del entorno con el menor coste posible de tiempo y recursos. En [93], Juliá et al. presentan una extensa comparación de distintas técnicas de exploración y creación de mapas.

Pero sin duda alguna, la tarea conjunta de localización y creación de mapas es la que ha tenido un mayor desarrollo durante las últimas décadas. Este proceso es más conocido por

Figura 2.7: Esquema de tareas de navegación y sus relaciones.

sus siglas en inglés, SLAM (*Simultaneous Localization and Mapping*), pudiendo aparecer también bajo las siglas CMBL (*ConcurrentMap Building and Localization*).

Esta tarea representa un problema complejo, pues el error en la localización deriva en una construcción errónea del mapa, y viceversa.

Cuando la información del entorno que utiliza el robot es proporcionada por un sistema visual, podremos hablar de SLAM visual. Normalmente, el mapa se representa mediante puntos característicos (o *landmarks*) estimados en espacio tridimensional. Esas *landmarks* formarán el mapa, y serán utilizadas para estimar la pose del vehículo cuando vuelvan a ser identificadas por el robot.

Debido a que las medidas de los sensores tienen una incertidumbre asociada, al igual que la estimación del desplazamiento del robot entre dos poses distintas, las soluciones de SLAM suelen introducir filtros probabilísticos que permiten tener en cuenta dicha incertidumbre en el sistema. Por ejemplo, [71] emplea un filtro de partículas Rao-Blackwellized. Otra característica de este trabajo es el empleo de un grupo de robots en la tarea de SLAM de forma cooperativa.

Uno de los filtros más empleados es el EKF (o *Extended Kalman Filter*). En [192] se compara este filtro con el método de optimización SGD (*Stochastic Gradient Descent*) en

tareas de SLAM visual. Específicamente, la información visual empleada por el robot son imágenes omnidireccionales capturadas por un sistema catadióptrico, sobre las que se extraen puntos característicos mediante SURF.

En la gran mayoría de los trabajos incluidos como referencia, se emplea la información proporcionada por la odometría interna para obtener una primera estimación del movimiento del robot. Cuando el movimiento de la cámara entre dos escenas se calcula usando la información contenida por las imágenes, es posible hablar de odometría visual. En [166, 190] se incluyen dos ejemplos diferentes de odometría visual utilizando información omnidireccional.

Según el tipo de mapa que utiliza el robot, es posible encontrar:

- **Mapas métricos**: Como su nombre indica, el propósito de este tipo de mapas es reconstruir un modelo de la distribución espacial de los elementos del entorno con precisión métrica. Esto puede ser llevado a cabo, por ejemplo, a través de la extracción de *landmarks* del entorno [176], o creando rejillas de ocupación del área de navegación [94].

- **Mapas Topológicos**: Los mapas topológicos utilizan grafos para representar el entorno que le rodea. En esos grafos, los nodos corresponden a características distintivas o zonas diferenciadas del mapa, mientras que los ejes representan relaciones entre los distintos nodos. Los nodos pueden llevar asociadas acciones de control.

  La característica más destacable de este tipo de mapas es que no hay distancias absolutas o sistemas de coordenadas que midan el espacio. Suelen ser mapas más compactos y simples que los métricos, lo que se traduce en un menor coste computacional. En [196], se presenta un trabajo de navegación topológica usando imágenes omnidireccionales. En una primera fase, se crea el mapa, cuyos nodos son imágenes de lugares característicos, y los ejes están formados por imágenes consecutivas entre los nodos.

- **Mapas híbridos**: En la actualidad, están apareciendo muchos trabajos que tratan de aprovechar las ventajas de ambos tipos de mapas con la creación de mapas híbridos.

  Si consideramos la navegación del robot en entornos grandes, los requerimientos computacionales necesarios para construir un mapa métrico pueden ser excesivamente altos para aplicaciones en tiempo real.

  Normalmente, los mapas híbridos utilizan la métrica para construir submapas de distintas zonas separadas, mientras que las relaciones topológicas sirven para relacionar las zonas incluidas en el mapa y llevar a cabo tareas como, por ejemplo, los cierres de

bucle o las uniones entre zonas alejadas en las que la precisión métrica no sea requerida. En [58, 103, 132] se incluyen distintas aproximaciones a la construcción de mapas híbridos.

Por último, cabe decir que también puede llevarse a cabo tareas de navegación sin mapa, como los basados en flujo óptico para detectar el desplazamiento de la cámara [32]. Esta técnica todavía presenta distintas áreas de mejora. Por ejemplo, en las fronteras de movimiento, el error de estimación del flujo es muy alto. Además, no existen indicadores para medir de forma fiable el error en la estimación [123].

CAPÍTULO

3

Visión Omnidireccional en Navegación Robótica

La visión en aplicaciones de navegación robóticas ha tenido un largo recorrido de desarrollo. Ya en el año 1979 aparece en [72] un trabajo en el que se presenta un robot modular destinado a navegación equipado con un sensor visual además de láser como sensores del sistema.

Un año más tarde, Moravec presenta en [134] el estudio desarrollado sobre una plataforma robótica construida por la Universidad de Standford durante los años 60, equipada con una cámara que permite la adquisición de imágenes con una resolución de 256×256 pixeles, y una profundidad de color de 32 escalas de gris.

Durante este tiempo, el avance en los sistemas de adquisición de la información visual y la mejora de los equipos de procesamiento de imágenes han permitido el avance del uso de la visión artificial en aplicaciones de navegación robótica. De este modo, se han podido desarrollar algoritmos más avanzados de extracción y detección de características aplicado a imágenes. En [48] y [23] están recogidos dos estudios detallados sobre el desarrollo de la navegación visual en los últimos 30 años.

Además de la mejora en el procesamiento de información, los sistemas visuales también han evolucionado en su capacidad de captar la información de su entorno, con una mayor velocidad de captura y una resolución muy superior a la de las primeras cámaras. Sin embargo, los sistemas de visión artificial no sólo han mejorado las características técnicas de las cámaras.

El mejor ejemplo de esa evolución son los sistemas que permiten la captura de un gran campo de visión alrededor del sensor. Muestra de ello son las cámaras esféricas, como la *Ladybug* [1], compuesta por varios sensores visuales distribuidos alrededor de la cámara, que logran recoger hasta el 90 % del campo visual a su alrededor. Otro ejemplo son los sistemas catadióptricos [158], basados en la utilización de espejos que recogen la escena a su alrededor. Estos espejos pueden tener distintas geometrías, como esféricos, hiperbólicos, parabólico, elípticos o cónicos, y la cámara adquiere la información del entorno a través del reflejo en los espejos. También pueden considerarse los sistemas que utilizan lentes con un amplio ángulo de visión, como las lentes de ojo de pez.

En este capítulo se incluye una descripción detallada de los sistema de adquisición de imágenes utilizados en este trabajo, y las distintas transformaciones que se pueden llevar a cabo sobre las imágenes omnidireccionales. En concreto, se han empleado sistemas de visión catadióptricos, y una cámara con lente de ojo de pez.

Gracias a la calibración de los sistemas de visión, es posible conocer la dirección de los rayos en el mundo real que llegan al sensor. De esta forma, podemos simular la proyección de la información visual en diferentes planos que aportan distintas vistas de la escena. Cada proyección presenta unas características propias que pueden ser de utilidad en tareas de navegación, tal y como se estudiará en capítulos posteriores de esta tesis.

Además, se describen las distintas plataformas utilizadas en la adquisición de las bases de imágenes empleadas en la parte experimental de la tesis para validación y comparación del comportamiento de los algoritmos propuestos.

3.1 Sistemas de Visión Catadióptricos

Los sistemas de visión catadióptricos son aquellos que usan superficies reflectantes para mejorar el campo de visión de la cámara. El sensor visual recoge la información a través de dicha superficie. La forma, posición y orientación del espejo determinará la proyección del mundo real en la cámara.

Existen multitud de sensores catadióptricos. El primero del que tenemos constancia es el desarrollado por Rees, y publicado en 1970 [161]. En [21, 74, 105, 202] es posible encontrar ejemplos de distintos sistemas catadióptricos.

La principal ventaja que presentan estos sistemas frente al resto es el aumento del campo visual alrededor del eje generatriz del espejo. De hecho, la mayoría de ellos permiten una visión omnidireccional, con un ángulo de visión de 360^o medido sobre un plano perpendicular al eje del espejo. Respecto al ángulo lateral de visión proporcionado por el espejo, depende directamente de su geometría y de la configuración del sistema catadióptrico.

En la literatura, es posible encontrar sistemas de visión omnidireccionales que utilizan espejos esféricos [142], cónicos [200], parabólicos [138], elípticos [109] o hiperbólicos [30]. En [205] se presenta un trabajo sobre la posibilidad de crear un sistema catadióptrico compuesto por dos espejos, centrando el estudio en la modificación de la geometría de dichos espejos y en el análisis de los resultados obtenidos.

Por otro lado, Baker y Nayar muestran en [17] una comparación sobre el uso de distintas geométricas de espejos en sistemas catadióptricos.

Tal y como aparece en [17], no se puede afirmar que una geometría sea mejor al resto. Cada forma presenta unas propiedades de reflexión características que pueden ser aprovechables. Sin embargo, si se desea extraer proyecciones perspectivas de la imagen, los espejos parabólicos e hiperbólicos son los más recomendables.

Otra propiedad deseable es que los sistemas de visión tengan un único centro de proyección, es decir, que sean cámaras centrales. En dichos sistemas, todos los rayos que llegan a la superficie del espejo tienen un único punto de intersección común, que es el centro óptico. Esta propiedad aparece en [17] y [74] bajo el nombre de *single effective viewpoint*.

Según estos trabajos, únicamente hay dos formas de conseguir un sistema de visión catadióptrico con esta propiedad: mediante un conjunto formado por un espejo hiperbólico con una cámara con lente de proyección perspectiva (lente convencional, o modelo *pin-hole*), o bien usando un sistema compuesto por un espejo parabólico y una lente de proyección ortográfica (aquella que hace converger los rayos paralelos al foco de la cámara). En la Figura 3.1 se representa un ejemplo de ambas configuraciones.

En el caso del sistema que utiliza el espejo hiperbólico, el centro óptico de la cámara tiene que coincidir con el foco de la hipérbola que se encuentra fuera del espejo. De esta

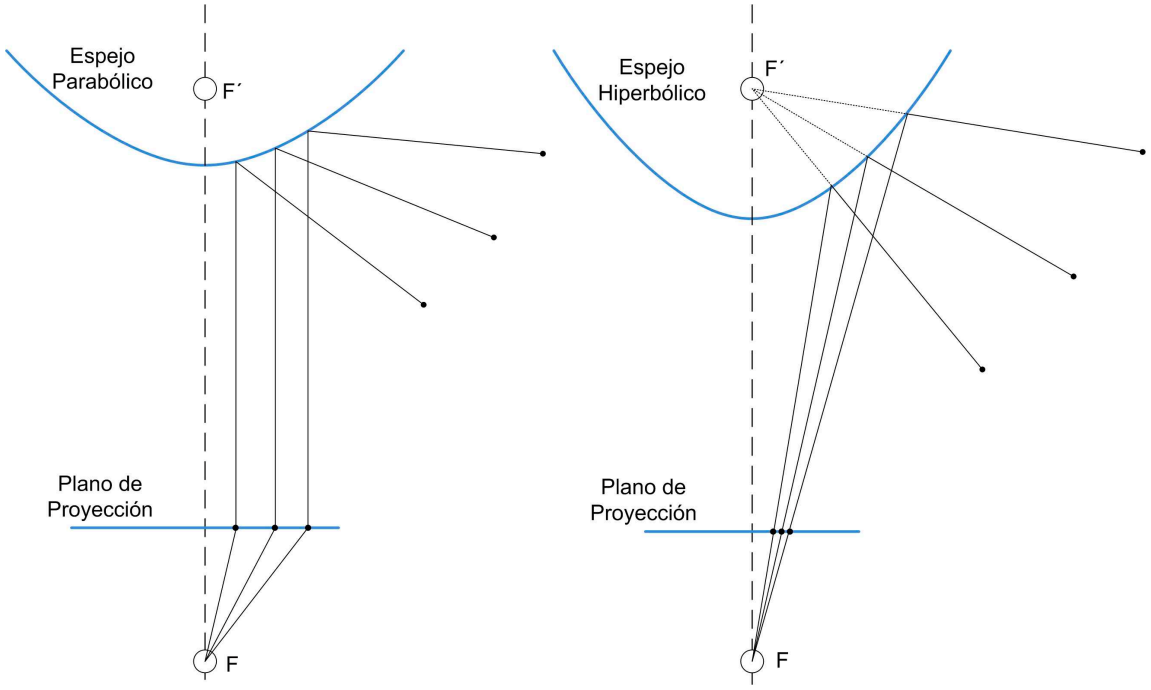


Figura 3.1: Representación de sistemas cadadióptricos centrales. Sistemas formados por (a) espejo hiperbólico con lente perspectiva, y (b) espejo parabólico con lente ortográfica.

forma, los puntos intersectan en un único punto. La imagen recogida equivaldría a colocar el sensor óptico en el foco del espejo.

Por otro lado, las cámaras no centrales son aquellas en los que los rayos que inciden en el espejo presentan un foco distinto dependiendo del ángulo de incidencia vertical. En [142], Ohte et al. describen este problema sobre un espejo esférico. Cuando un objeto está situado lo suficientemente lejos, el error provocado por la existencia de distintos puntos focales es prácticamente nulo. Sin embargo, cuando el objeto está situado cerca del sistema de visión, aparecen problemas de paralaje o errores en los ángulos de proyección.

Gracias a la calibración del sistema omnidireccional, es posible conocer la dirección en el mundo real del rayo correspondiente a cada uno de los píxeles recogidos en el plano de proyección con respecto al foco del espejo.

Consideramos u y v las coordenadas del punto **p** en la imagen recogida en el sensor, cuyo origen de coordenadas corresponde con el punto que corta la línea $\overline{FF'}$ y el plano de proyección. El punto **p** es la proyección de **P**, un punto situado en el mundo real.

Mediante la calibración del sistema catadióptrico, es posible conocer la dirección del rayo que une el foco del espejo con el punto **P**, indicada mediante el vector $\vec{P}$. Definiremos $\vec{P}$ con tres componentes respecto al sistema de coordenadas del mundo real, situado en el

centro óptico del espejo.

En la Figura 3.2 es posible ver la proyección del punto **P** en el plano de proyección de la cámara **p**, y el vector de dirección respecto al centro óptico del espejo $\vec{P}$.

Siendo UV el sistema de coordenadas del plano imagen, y el sistema XYZ el sistema correspondiente al mundo real, centrado en el foco del espejo, consideramos el sistema de coordenadas UV alineado con las direcciones XY. Las coordenadas x e y del vector del vector son proporcionales a las coordenadas u y v del punto **p** respectivamente:

$$\begin{bmatrix} x \\ y \end{bmatrix} = \alpha \cdot \begin{bmatrix} u \\ v \end{bmatrix} \tag{3.1}$$

Por lo tanto, el vector $\vec{P}$ se puede definir como:

$$\vec{P} = \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} \alpha \cdot u \\ \alpha \cdot v \\ f(u,v) \end{bmatrix} \tag{3.2}$$

Del vector $\vec{P}$ únicamente nos interesa su dirección, por lo que es posible incluir el parámetro α dentro de la función $f(u,v)$, quedando la Ecuación anterior como:

$$\vec{P} = \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} u \\ v \\ f(u,v) \end{bmatrix} \tag{3.3}$$

Debido a la geometría simétrica del espejo, se puede deducir que la altura a la que se encuentra el punto **P** es proporcional a la distancia de la proyección respecto al centro de coordenadas de la imagen omnidireccional (ρ). Esto se traduce en:

$$\vec{P} = \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} u \\ v \\ f(\rho) \end{bmatrix}, \quad \text{con } \rho = \sqrt{u^2 + v^2}. \tag{3.4}$$

La función f depende de la geometría del espejo. Se implementa mediante un polinomio sin orden predefinido:

$$f(\rho) = a_0 + a_1 \cdot \rho + a_2 \cdot \rho^2 + a_3 \cdot \rho^3 + \ldots + a_n \cdot \rho^n. \tag{3.5}$$

La calibración de la cámara nos proporciona los parámetros del polinomio de la ecuación $f(\rho)$.

En la Figura 3.4 se incluyen dos ejemplos de sistemas de visión catadióptricos e imágenes capturadas en un mismo entorno con ambos sistemas. Los dos espejos son hiperbólicos. Comparando ambas imágenes, podemos observar la diferencia de distribución de los elementos como consecuencia de la distinta geometría del espejo.

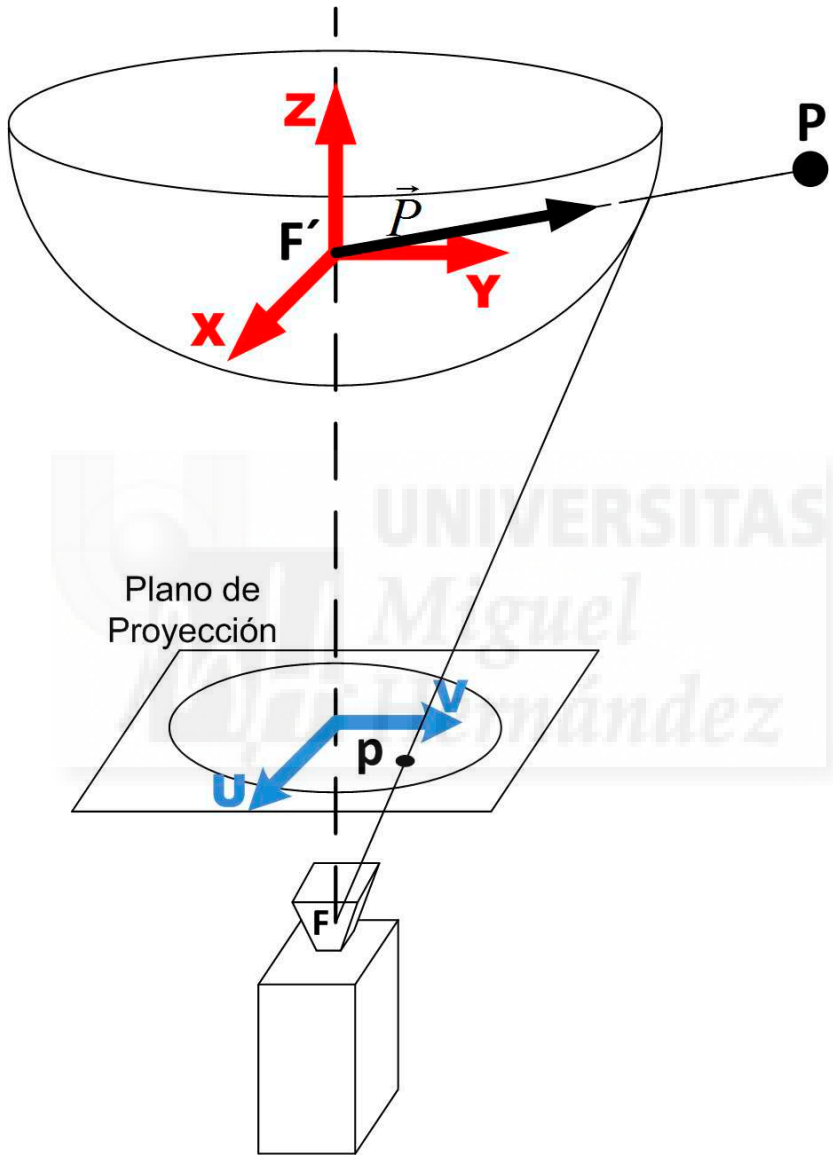
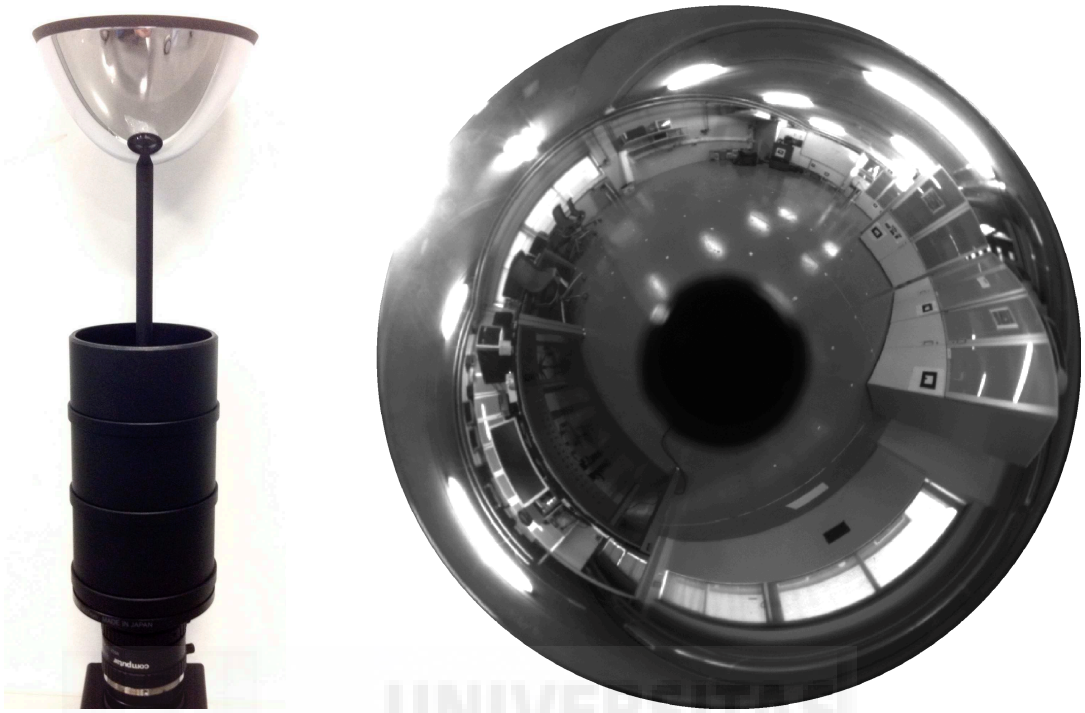


Figura 3.2: Modelo de proyección de un punto **P** del mundo real en el plano de proyección del sistema catadióptrico (**p**).



(a)

(b)

Figura 3.3: Ejemplo de dos imágenes capturadas en un mismo entorno por sistemas catadióptricos distintos.

3.1.1 Sistema Catadióptrico usado en este trabajo

En este trabajo se ha empleado un sistema de visión catadióptrico para adquirir bases de imágenes omnidireccionales en distintos entornos.

Los equipos estudiados se corresponden con los mostrados en la Figura 3.4. Ambos se componen de un espejo hiperbólico y una cámara CCD color.

Se han usado dos sensores de adquisición de imágenes distintos. Ambos pertenecen a la compañía *The Imaging Source*, y son cámaras de tipo CCD color provistas de conexión Firewire. Los modelos exactos son el DFK-21BF04 [183] y el DFK-41BF02 [184]. En la Figura 3.4 aparece una imagen del modelo DFK-21BF04, aunque cabe destacar que el modelo DFK-41BF04 presenta el mismo aspecto exterior.

Figura 3.4: Cámara CCD DFK-21BF04

En la Tabla 3.1 podemos ver una comparación de las principales características de ambos modelos. Como se puede apreciar, la mayor diferencia entre ambos sensores es la resolución de la imagen. Por ello, durante este trabajo nos referiremos al modelo DFK-21BF04 como cámara de baja resolución, y al DFK-41BF02 como de alta resolución.

Con respecto a los espejos, se trata del modelo Wide70 de *Eizoh* [52] y del modelo Super-Wide view Large de *Accowle* [3]. En la Tabla 3.2 se recogen las características de ambos espejos hiperbólicos.

El espejo Eizoh Wide70 viene con unos adaptadores que permiten modificar la distancia del espejo con respecto al de la cámara. Tal y como se indica en la tabla, la distancia óptima para conseguir los mejores resultados en cuanto a cantidad de información recogida por el sistema omnidireccional es de 160 mm.

Por otro lado, el espejo Accowle SuperWide Large tiene una distancia fija entre el espejo y la cámara, y está protegido con un cristal a su alrededor de posibles impactos, que es el que sujeta el espejo a la estructura del sistema catadióptrico.

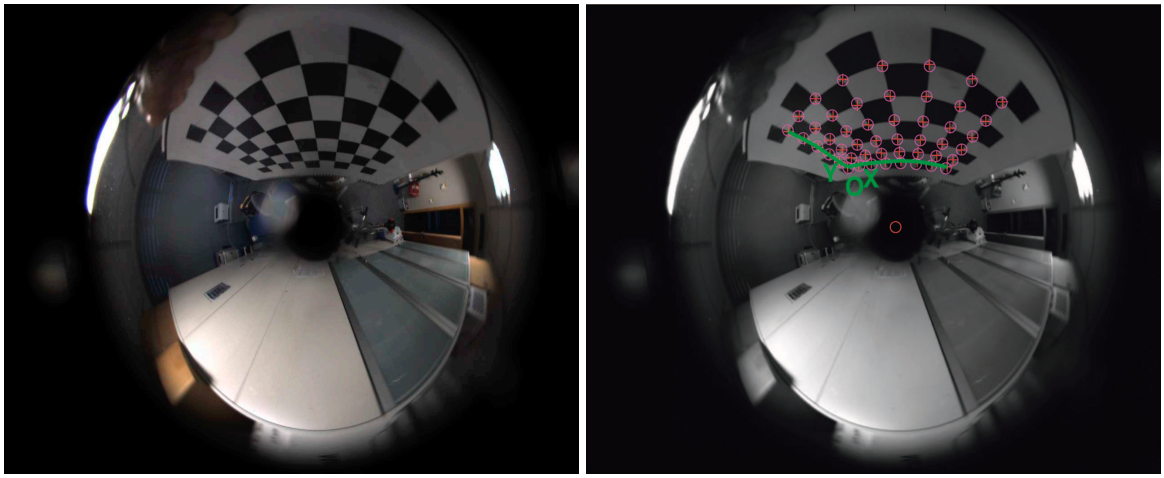
Como se ha comentado anteriormente, para conseguir la ecuación que modela la dirección de los rayos proyectados en la imagen omnidireccional, es necesario calibrar la cámara. Este proceso se ha realizado mediante la *toolbox* de Matlab *OCamCalib* [165].

Modelo	**DFK-21BF04**	**DFK-41BF02**
Fabricante	The Imaging Source	The Imaging Source
Resolución de Imagen	640x480 pixel	1280x960 pixel
Conexión	FireWire	FireWire
Formatos de Video	UYVY / BY8	UYVY / BY8
Frecuencia de Imágenes	de 3.75 a 60 fps	de 3.75 a 60 fps
Intensidad	0.10 lx	0.15 lx
Sensor	CCD Sony ICX098BQ	CCD Sony ICX205AK
Formato CCD	1/4”	1/2”
Tamaño de píxel	H: 5.6μm, V: 5.6μm	H: 4.65μm, V: 4.65μm
Montaje de lente	C/CS	C/CS
Tensión de alimentación	8 a 30 VDC	8 a 30 VDC
Corriente Eléctrica	aprox. 200 mA 12 VDC	aprox. 200 mA 12 VDC

Tabla 3.1: Especificaciones de los sensores CCD usados en este trabajo.

Fabricante/Modelo	**Eizoh Wide70**	**Accowle SuperWide L**
Ángulo de Visión Superior	60^o sobre el horizonte	55^o sobre el horizonte
Ángulo de Visión Inferior	60^o bajo el horizonte	65 ^{o} bajo el horizonte
Ángulo Lateral de Visión	360^o	360^o
Geometría	Hiperbólico	Hiperbólico
Diámetro Máximo Espejo	70 mm	74 mm
Altura Espejo	35 mm	-
Diámetro de Anclaje	75 mm	-
Distancia óptima centro espejo y centro proyección cámara	160 mm	Fija
Peso Conjunto	170g aprox.	-

Tabla 3.2: Especificaciones técnicas de los espejos Eizoh Wide 70 y Accowle SuperWide Large.



(a) (b)

Figura 3.5: Imagen omnidireccional con patrón usado para la calibración del conjunto catadióptrico. (a) Imagen original y (b) imagen con esquinas del patrón marcadas por el algoritmo de calibración.

Esta librería dispone de su propio patrón de calibración, aunque cabe destacar que es posible usar otro patrón distinto. En la Figura 3.5 se muestra una imagen usada durante el proceso de calibración, en su apariencia original y una vez se han encontrado las esquinas del patrón.

Tras llevar a cabo la calibración, obtenemos los coeficientes del polinomio de la función de proyección $f(\rho)$ definida en la Ecuación 3.5.

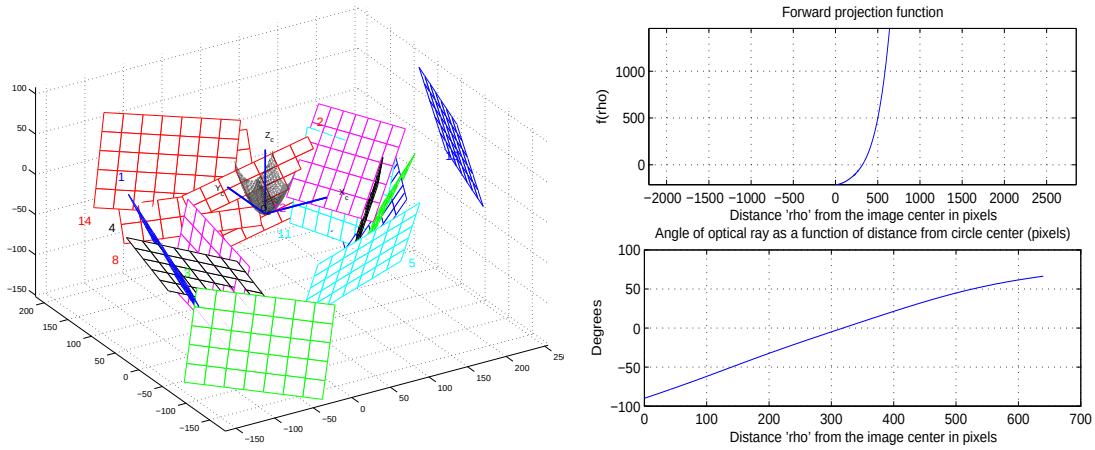
También es posible obtener una representación gráfica de la posición de los patrones utilizados en las imágenes de calibración respecto el sistema de referencia del espejo, además de gráficas de la función $f(\rho)$, y del el ángulo del rayo óptico con respecto a ρ, es decir, con respecto a la distancia de cada pixel al centro de la imagen omnidireccional. Las Figuras 3.6 y 3.7 recogen las gráficas obtenidas en la calibración de los dos sistemas catadióptricos disponibles.

En el caso concreto del conjunto catadiótrico formado por el espejo Eizoh Wide70 y la cámara de alta resolución, el error mínimo de calibración se ha obtenido para un polinomio de grado 4, con coeficientes $a_0 = -212{,}5180$, $a_1 = 0$, $a_2 = 3{,}200 \cdot 10^{-3}$, $a_3 = -8{,}1262 \cdot 10^{-6}$ y $a_4 = 1{,}4931 \cdot 10^{-8}$. Por lo tanto, el polinomio queda como:

$$f(\rho) = -212{,}5180 + 3{,}200 \cdot 10^{-3} \cdot \rho - 8{,}1262 \cdot 10^{-6} \cdot \rho^3 + 1{,}4931 \cdot 10^{-8} \cdot \rho^4 \qquad (3.6)$$

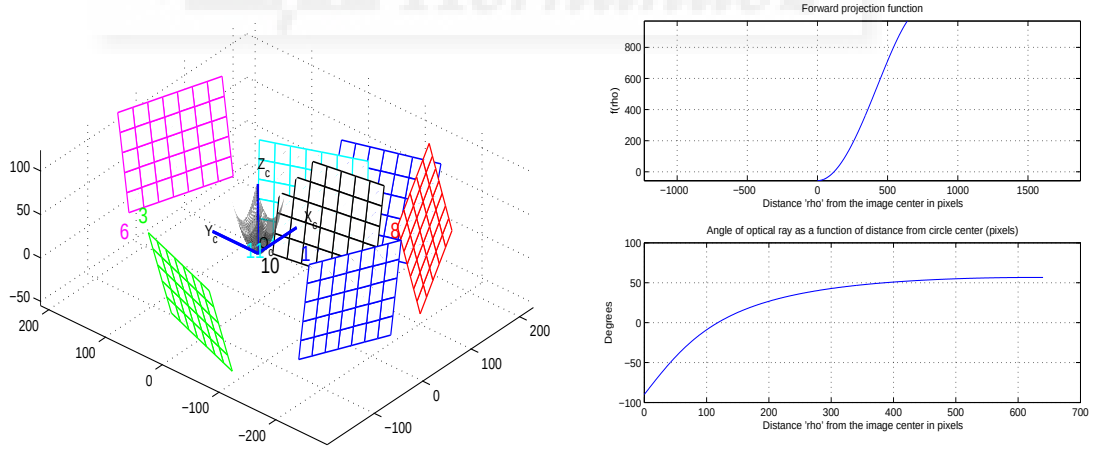
El error de calibración es de 0.671372 pixel.

Por su lado, el conjunto catadióptrico formado por el espejo Accowle SuperWide y la cámara DFK-41BF02 presenta un error muy similar usando polinomios de grado 3 o 4. Por



(a) (b)

Figura 3.6: Calibración del conjunto catadióptrico formado por cámara DFK-41BF02 y espejo Eizo Wide70. (a) Estimación de la posición de los patrones de calibración de las distintas imágenes respecto al sistema de referencia del espejo, y (b) representación de la función de proyección y del ángulo del rayo óptico con respecto a ρ.



(a) (b)

Figura 3.7: Calibración del conjunto catadióptrico formado por cámara DFK-41BF02 y el espejo Accowle SuperWide Large. (a) Estimación de la posición de los patrones de calibración de las distintas imágenes respecto al sistema de referencia del espejo, y (b) representación de la función de proyección y del ángulo del rayo óptico con respecto a ρ.

ser una ecuación más sencilla, elegimos un polinomio para la función de calibración de grado 3, cuyos coeficientes son $a_0 = -57{,}5064$, $a_1 = 0$, $a_2 = 4{,}387 \cdot 10^{-3}$ y $a_3 = -2{,}1416 \cdot 10^{-6}$.

La función de calibración queda como:

$$f(\rho) = -57{,}5064 + 4{,}387 \cdot 10^{-3} \cdot \rho a_3 = -2{,}1416 \cdot 10^{-6} \cdot \rho^3. \tag{3.7}$$

siendo el error de calibración igual a 0.4670 pixel.

Como se puede apreciar en las gráficas obtenidas en la calibración (Figuras 3.6(b) y 3.7(b)), la distribución del ángulo óptico con respecto a ρ es distinto para cada sistema catadióptrico.

El espejo Eizoh Wide 70 (Figura 3.6(b)) presenta una relación lineal del ángulo del rayo óptico recogido por el espejo y la distancia de cada pixel con respecto al centro de la imagen omnidireccional.

Por otro lado, el modelo Accowle SuperWide (Figura 3.7(b)) presenta una distribución no lineal. Los rayos con ángulo de incidencia negativos, es decir, los provenientes de puntos situados por debajo del foco del espejo en el mundo real, se recogen en los píxeles de la imagen omnidireccional situados dentro de un radio cercano al centro (aproximadamente en un 20% del radio exterior de la imagen omnidireccional). Por contra, la mayor parte de la imagen corresponde únicamente a los rayos con ángulo de incidencia más elevados, otorgando una mayor importancia a dicha información.

El Accowle SuperWide resulta interesante en aplicaciones en las cuales sea necesario centrarse en la parte más alta de la escena (a partir de 25^o sobre el eje Z del foco del espejo). Además, la oclusión central es menor, ya que la la sujeción del espejo a la estructura no se realiza a través de un adaptador central, como en el caso del Eizoh Wide 70.

Sin embargo, en este trabajo se usará el sistema catadióptrico que utiliza el espejo Eizoh Wide 70, pues preferimos una distribución lineal de los rayos ópticos con respecto a la distancia de proyección en la imagen omnidireccional. Esta característica nos será de utilidad al obtener la proyección panorámica de la información visual. Aunque la oclusión central es mayor, la distribución angular de los rayos ópticos permite que el ángulo inferior proyectado en el espejo sea muy similar al del Accowle SuperWide Large.

La posición del sistema catadióptrico será siempre la misma, con el eje del espejo perpendicular al plano del suelo. Cuando el plano del suelo no presente inclinación, el eje de la cámara será además paralelo al eje Z del mundo real.

3.2 Representación de la Información Omnidireccional

Como hemos visto, el conjunto catadióptrico es capaz de capturar una escena con un campo de visión de 360º alrededor del eje Z, con un ángulo lateral efectivo de 120º en el caso de nuestro sistema.

La imagen obtenida en el sensor visual corresponde al reflejo de los rayos en el espejo. Dicha imagen está representada en coordenadas polares, cuyo origen de coordenadas se corresponde con la proyección del foco del espejo en el sensor visual. En la Figura 3.5(b) se muestra el centro de la imagen marcada con un círculo rojo. Dicho centro no tiene por qué coincidir exactamente con los píxeles centrales de la imagen omnidireccional.

Debido a la geometría simétrica del espejo, el ángulo del pixel en la imagen omnidireccional se corresponde con la dirección exacta de reflexión del rayo en el espejo en el plano XY. Por otro lado, la distancia de cada pixel al centro de la imagen es función del ángulo lateral de entrada (medido en el plano YZ). Este ángulo depende de la geometría del espejo, y se modela mediante la función $f(\rho)$ (Ecuación 3.5), que es característico de cada sistema catadióptrico. La distribución vertical de la información visual en la imagen omnidireccional estará condicionada por dicha función.

En la Figura 3.8 se muestra un ejemplo de imagen omnidireccional capturada en un entorno de exterior.

La escena omnidireccional puede utilizarse directamente para obtener información útil en tareas de navegación robóticas. Como ejemplo, Scaramuzza et al. presentan en [167] un descriptor basado en la extracción de líneas radiales de la imagen omnidireccional para caracterización de escenas omnidireccionales capturadas en entornos reales. Nótese que las líneas radiales de la imagen omnidireccional corresponden con las líneas verticales de la escena en el mundo real, mientras que las coordenadas angulares (que forman líneas concéntricas al origen de coordenadas de la imagen) se corresponden con líneas horizontales en la escena real.

Además, a partir de la imagen omnidireccional, es posible obtener distintas representaciones de la escena mediante la proyección de la información visual en distintas planos y geometrías. En concreto, en este trabajo vamos a ver la proyección de la escena en la esfera unitaria, la formación de la imagen panorámica, la obtención de imágenes perspectivas, y la imagen ortográfica.

En los siguientes puntos se desarrolla cada una de las representaciones de forma conceptual. En [74], se presenta un trabajo que trata las distintas proyecciones desde un punto de vista más matemático.

En cada apartado se incluye un ejemplo de cada proyección a partir de la imagen omnidireccional mostrada en la Figura 3.8.

Figura 3.8: Imagen Omnidireccional

3.2.1 Imagen Esférica

Una imagen omnidireccional puede ser proyectada sobre una esfera unitaria cuyo centro se sitúe en el foco del espejo del sistema catadióptrico. Cada pixel de la esfera unitaria adquiere el valor del rayo del mundo real que tiene la misma dirección con respecto al foco de la hipérbola.

En la Figura 3.9 se representa el esquema de proyección de los rayos recogidos en el mundo real por el espejo hiperbólico sobre la esfera.

La dirección del rayo en el plano XY es la misma que la correspondiente a su proyección en la imagen omnidireccional en el plano UV. La tercera componente del vector en el espacio se obtiene introduciendo la norma del vector correspondiente a cada pixel en el plano imagen $\rho = \sqrt{u^2 + v^2}$ como argumento de la función de calibración, como se puede ver en la Ecuación 3.4. Obtenidas las tres componentes, se puede normalizar el vector para que su módulo sea unitario.

Nuestro sistema catadióptrico no recoge el campo visual correspondiente a los ángulos verticales superiores a $+60^o$ ni inferiores a -60^o respecto al centro de proyección del espejo. Por ello, al realizar la proyección sobre la esfera unitaria quedarán regiones sin información asociada. Esto se puede ver en la imagen de la Figura 3.10.

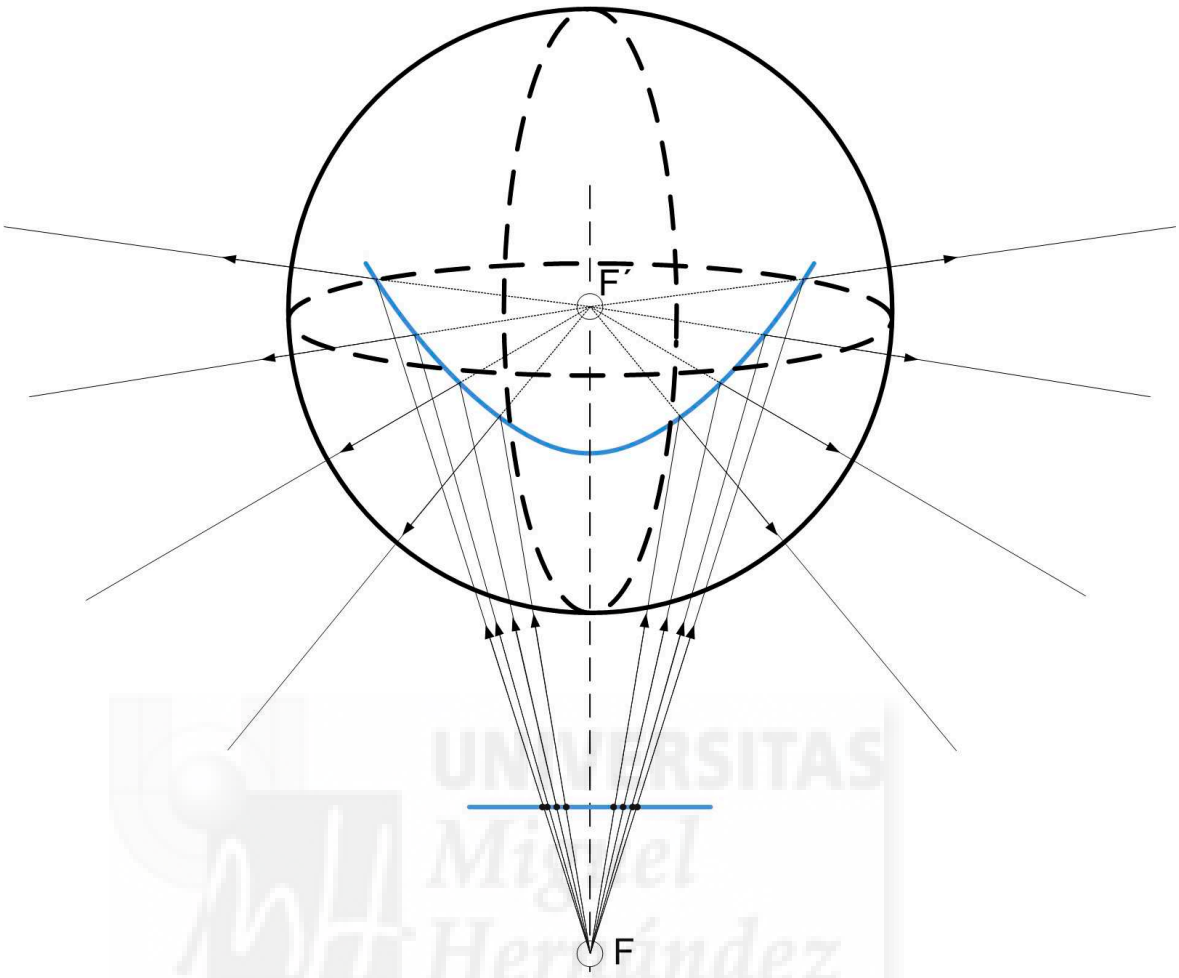


Figura 3.9: Modelo de proyección de la imagen esférica.

3.2.2 Imagen Panorámica

La representación panorámica de la información omnidireccional es una de las más utilizadas en trabajos de navegación robótica, como [39, 66, 114, 131, 153, 199], ya que es una representación más comprensible que la omnidireccional, y además, permite aplicar la mayoría de algoritmos de procesamiento de imágenes pensados para imágenes con proyección perspectiva.

Tal y como se puede ver en el esquema de la Figura 3.11, se trata de la proyección de la información omnidireccional sobre una superficie cilíndrica.

La imagen omnidireccional se forma dependiendo de la geometría del espejo en lo que puede ser definido como un mapeado polar no lineal, ya que las coordenadas radiales dependen de la función de proyección $f(\rho)$.

La manera más sencilla de obtener una imagen panorámica es cambiar el sistema de coordenadas de la imagen omnidireccional (mapeada en coordenadas polares) a un sistema

Figura 3.10: Imagen Esférica

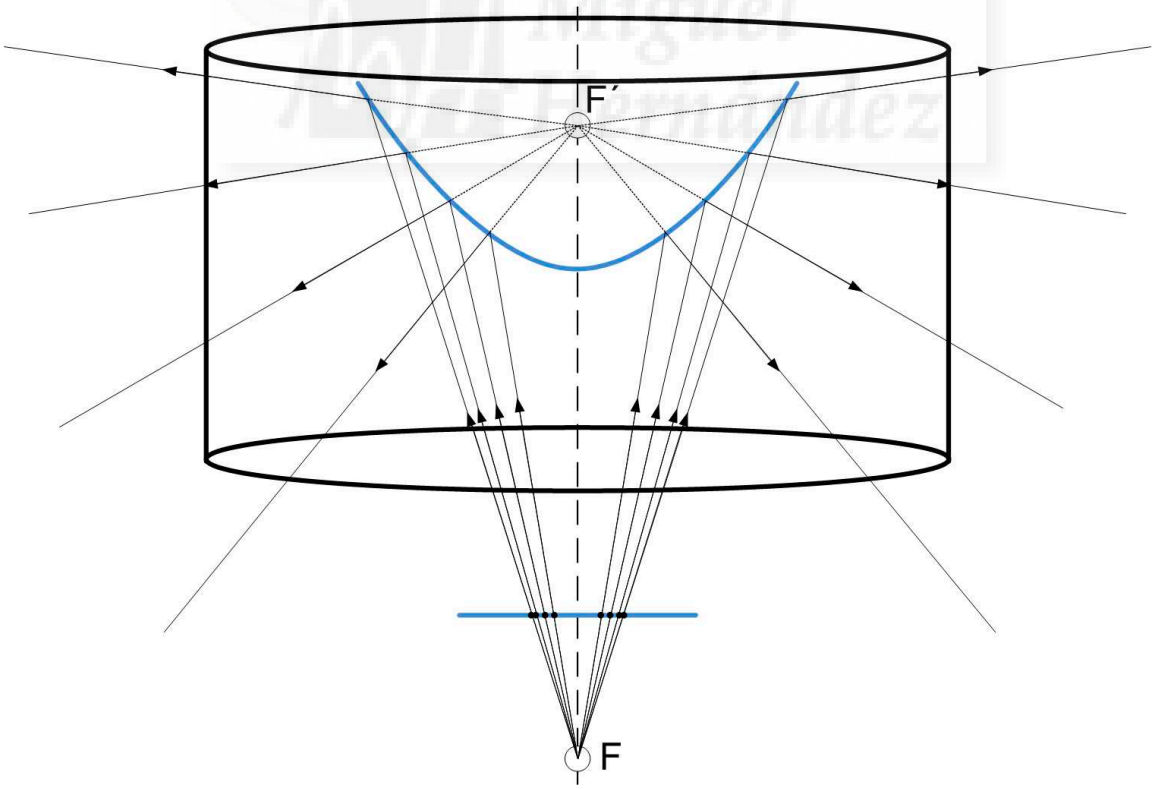


Figura 3.11: Modelo de proyección de la imagen panorámica.

(a) (b)

Figura 3.12: Imágenes panorámicas obtenidas mediante cambio de sistema de coordenadas a partir del sistema catadióptrico que utiliza el espejo (a) Eizoh Wide70 y el (b) Accowle SuperWide.

de coordenadas rectangulares. Esta transformación hace corresponder cada circunferencia de la imagen a una linea horizontal de la imagen panorámica, y las coordenadas radiales de la imagen omnidireccional pasan a ser líneas verticales de la imagen panorámica.

Sin embargo, al obtener la imagen panorámica de esta forma, se hace una correspondencia lineal entre el radio de la imagen omnidireccional y la altura en la imagen panorámica.

Retomando el ejemplo de los dos sistemas catadióptricos que disponemos en nuestro trabajo, en la Figura 3.12 podemos ver la vista panorámica obtenida a partir de las imágenes de los distintos sensores mostradas en la Figura 3.4.

Si comparamos ambas imágenes, podemos percatarnos que la Figura 3.12(b) introduce una distorsión vertical en la escena, pues amplía la zona correspondiente con los ángulos más elevados, y concentra la información de la zona más baja. La razón se encuentra en la función de proyección del espejo Accowle SuperWide (Figura 3.7(b)), pues no es lineal. Sin embargo, la imagen panorámica correspondiente con el Eizoh Wide70 (Figura 3.12(a)) no presenta distorsión apreciable, ya que su función de proyección sigue una distribución casi lineal (Figura 3.6(b)).

Este método no requiere la calibración del sistema catadióptrico para obtener la proyección panorámica de la imagen.

Para eliminar la distorsión producida por espejos con ecuación no lineal, es necesario usar el modelo de proyección al obtener la vista panorámica. La imagen se formará buscando la correspondencia entre la proyección de los rayos en el plano imagen y su intersección en el plano cilíndrico, donde se obtiene la proyección panorámica.

Este segundo método sí requiere la calibración de la cámara, y es computacionalmente más costoso que el primero.

Como se ha comentado al final de la Sección 3.1.1, en este trabajo hemos elegido el sistema catadióptrico que utiliza el espejo Eizoh Wide70 para capturar las bases de imágenes utilizadas en la parte experimental, y la transformación de las imágenes omnidireccionales a panorámicas se realiza con el cambio de coordenadas polares a rectangulares.

Figura 3.13: Imagen Panorámica.

3.2.3 **Imagen Perspectiva**

A partir de la información recogida en la imagen omnidireccional y de la calibración del sistema de visión, pueden obtenerse vistas proyectivas de la escena. El esquema de proyección se corresponde con el mostrado en la Figura 3.14.

La imagen obtenida se aproxima a la proporcionada por una cámara convencional situada en el foco del espejo.

La posición y orientación del plano donde se proyectan los rayos puede ser modificado, cambiando con ello la apariencia de la escena proyectada en él.

En nuestros experimentos, los planos de proyección utilizados para obtener imágenes perspectivas van a ser siempre perpendiculares al plano XY a no ser que se indique lo contrario.

Será posible modificar el tamaño del plano de proyección, la distancia respecto al eje F-F'del sistema catadióptrico, y el ángulo que forma el vector director del plano de proyección con U medido en el plano UV. Este último ángulo es el mismo que el medido sobre el eje X en el plano XY de mundo real, ya que recordemos que el sistema de referencia de la cámara UV es paralelo a XY.

Siguiendo el esquema mostrado en la Figura 3.14, el plano de proyección quedará definido mediante la distancia f_p, que representa la distancia en píxeles desde el punto focal del espejo hiperbólico al plano, y el ángulo θ.

Cabe destacar que f_p actuará como un parámetro de zoom de la imagen perspectiva. Cuanto menor sea la distancia entre el foco del espejo y el plano imagen, mayor será la ampliación. Si utilizamos una distancia demasiado pequeña, pueden aparecer en la imagen problemas de pixelación por falta de resolución, sobre todo en los ángulos inferiores de la proyección, que se corresponden con las circunferencias de menor radio de la imagen omnidireccional. Por otro lado, el ángulo θ determinará la parte de la escena omnidireccional que va a proyectarse.

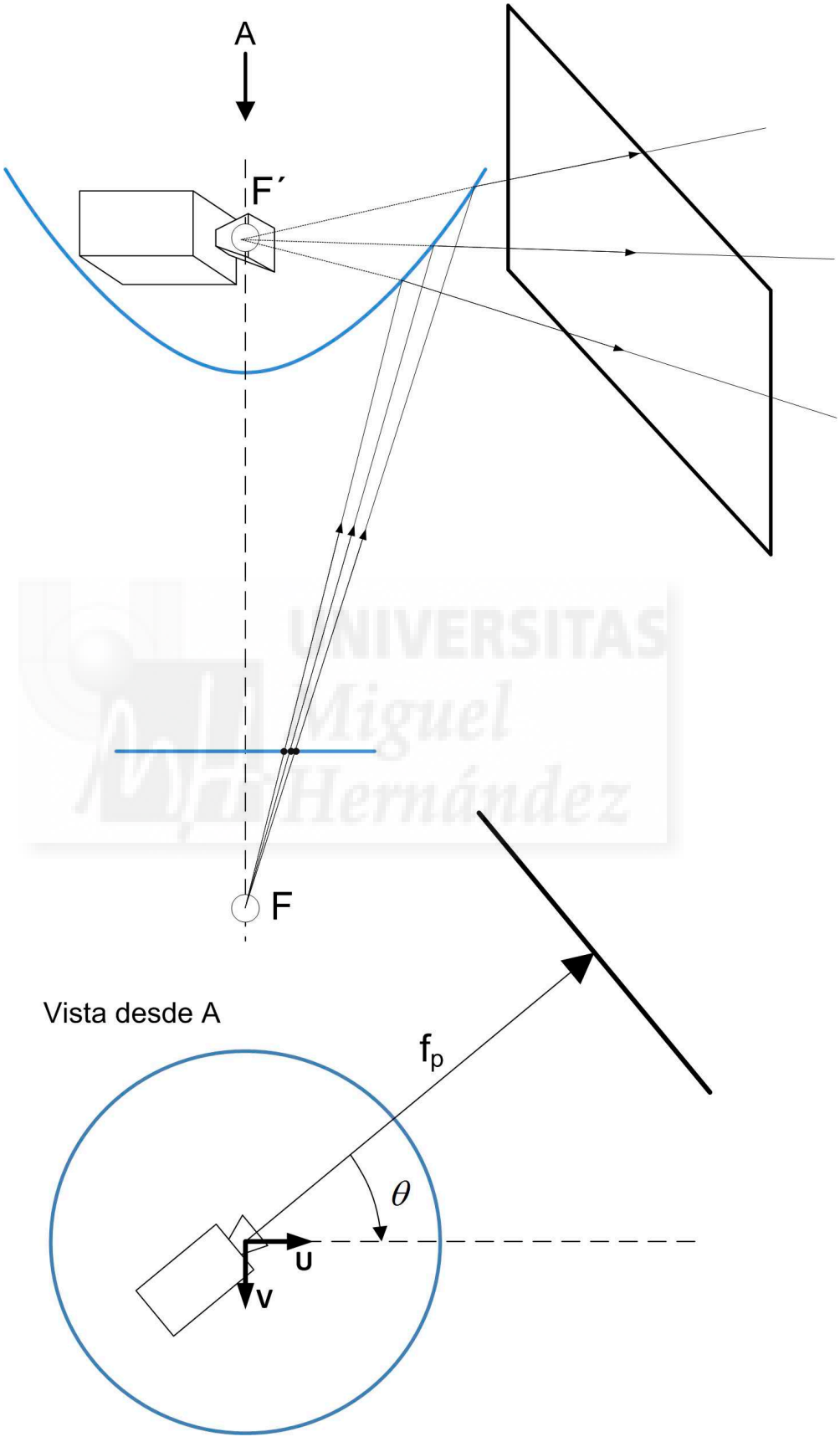
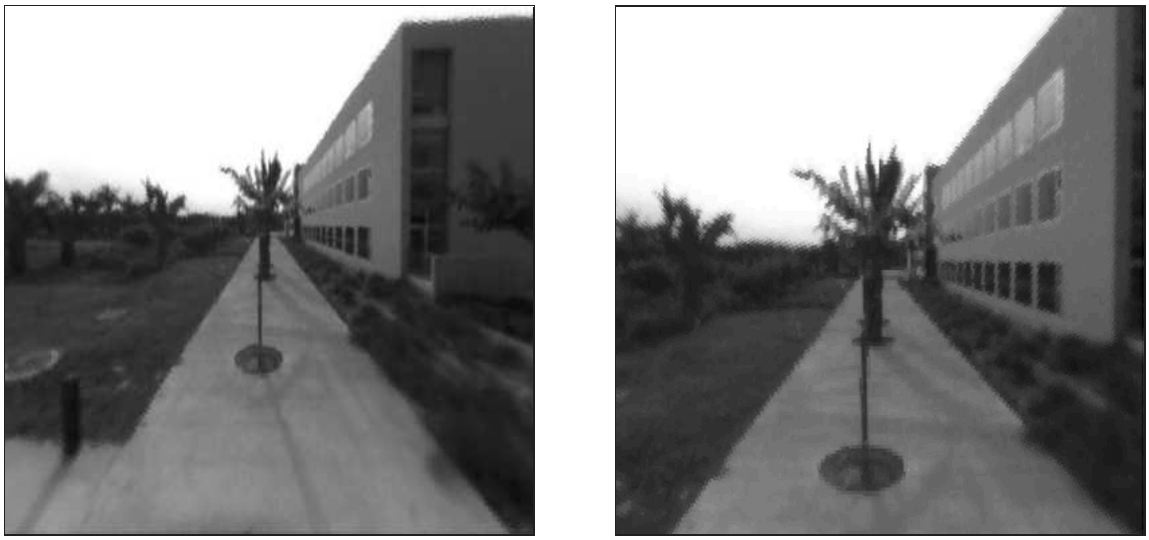


Figura 3.14: Modelo de proyección de una imagen perspectiva.



(a) (b)

Figura 3.15: Imágenes Perspectivas usando distintas distancias entre el foco del espejo y el plano de proyección.

En la Figura 3.15 se recogen dos imágenes proyectivas con mismo ángulo θ pero distinta distancia en el plano de proyección. En la Figura 3.15(b) puede apreciarse un efecto de pixelado en la imagen al usar un plano con una distancia f_p demasiado pequeña.

3.2.4 **Imagen Ortográfica**

La imagen ortográfica, también conocida como vista de pájaro, puede considerarse un caso particular de proyección perspectiva.

En la literatura, es posible encontrar distintos trabajos que usan la proyección ortográfica de la imagen omnidireccional. En [65] se usa la vista de pájaro para extraer las líneas paralelas de un pasillo en tareas de navegación robóticas en entorno de interior. En [22, 39, 163] se encuentran otros trabajos donde se aprovechan las propiedades de la proyección ortográfica en aplicaciones de localización, navegación y construcción de mapas.

Para obtener la imagen ortográfica, el plano de proyección de la imagen está situado perpendicularmente al eje de la cámara. Por la configuración de nuestro sistema catadióptrico, este plano también será paralelo al plano XY. Por lo tanto, obtenemos una imagen que equivale a la capturada por una cámara convencional situada en el foco del espejo con dirección hacia el plano del suelo. En la Figura 3.16 aparece el modelo de proyección de la vista ortográfica.

Debido al eje que sujeta el espejo y al propio sensor CCD, se produce una oclusión en la parte central de la imagen, como puede verse en el ejemplo recogido en la Figura 3.17.

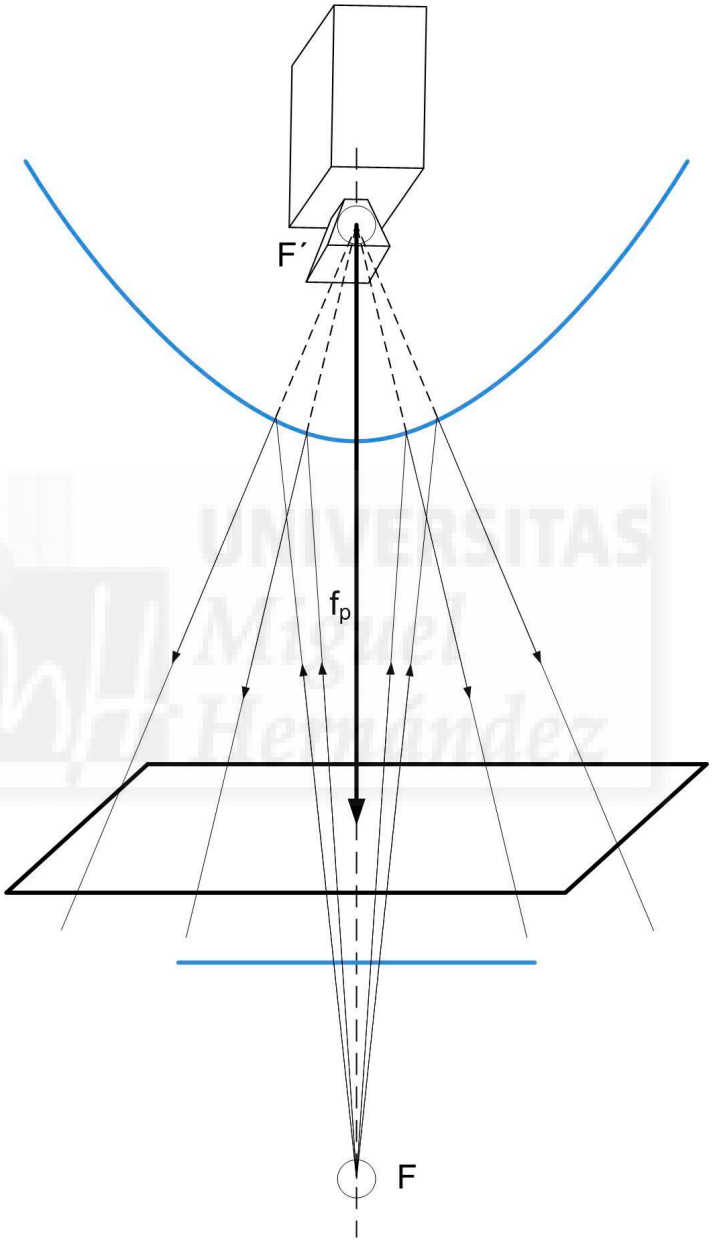


Figura 3.16: Modelo de proyección de la vista ortográfica.

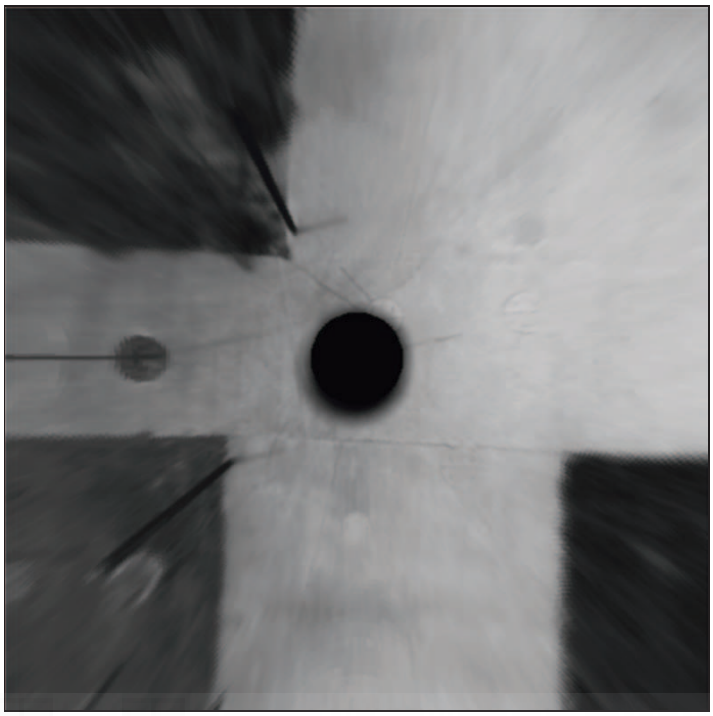


Figura 3.17: Imagen Ortográfica

Al igual que ocurre con la proyección perspectiva, la variación de la distancia del plano de proyección f_p producirá un efecto de zoom en la imagen ortográfica resultante.

Figura 3.18: Imagen de la cámara GoPro Hero.

3.3 Sistemas de Visión con Lente Ojo de Pez.

Una alternativa para conseguir imágenes con un ángulo de visión amplio son las cámaras que utilizan lente de ojo de pez. El primer artículo donde se tiene en cuenta este tipo de formación de imágenes se recoge en [197]. En [75] podemos ver un ejemplo más reciente de utilización de una lente de ojo de pez en un sistema de navegación para un vehículo aéreo no tripulado. En [198] se presenta un algoritmo de seguimiento para un robot equipado con un sistema visual con lente de ojo de pez. En el trabajo incluido en [171], se obtienen vistas panorámicas desde imágenes obtenidas con una lente ojo de pez que cubre los 360º.

Una de las principales características de las imágenes capturadas con sistemas de visión con lente de ojo de pez es que presentan distorsión radial. En la Figura 3.20(a) podemos ver una imagen capturada con uno de estos sistemas.

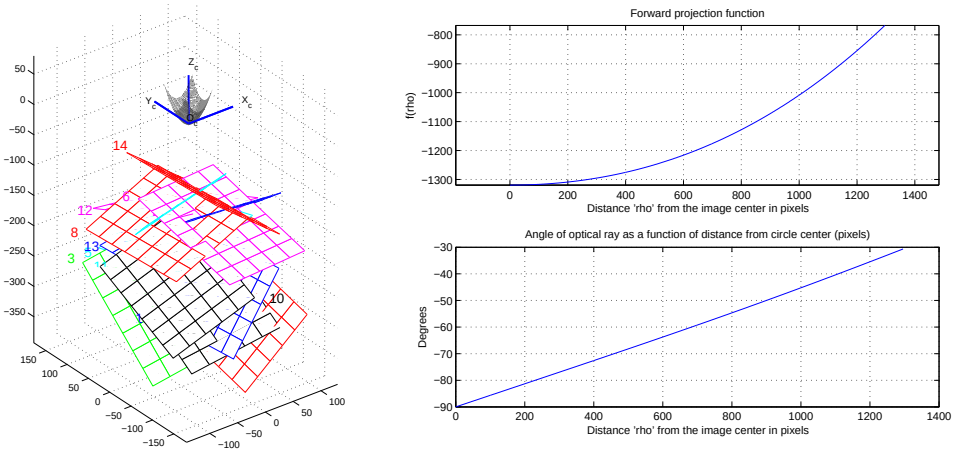
Para corregir esta distorsión, necesitamos conocer la función de proyección de la cámara.

Tal y como se describe en [41], es posible utilizar el modelo de calibración de las cámaras catadióptricas para la calibración de una cámara con lente de ojo de pez. Aunque en general no son sistemas de visión centrales, se aproximan bastante a la propiedad de único centro óptico. La librería *OCamCalib* empleada para la calibración del sistema catadióptrico puede ser empleada para la calibración de lentes de ojo con un campo de visión de hasta 195º.

3.3.1 Sistema con Lente Ojo de Pez usado en este trabajo

La cámara utilizada es el modelo Hero 960 de *GoPro*[2]. En la Figura 3.18 se puede ver una imagen del dispositivo. Se trata de una cámara compacta que permite una gran resolución y un amplio campo de visión (de hasta 170º) gracias a su lente. En la Tabla 3.3 se incluyen las principales características de la cámara utilizada.

Como se ha comentado anteriormente, el tipo de lente utilizada (de distancia focal corta), introduce distorsión radial (o distorsión de barril). Para corregir dicha distorsión, necesitamos



(a) (b)

Figura 3.19: Calibración del sistema dióptrico GoPro Hero 960. (a) Estimación de la posición de los patrones de calibración de las distintas imágenes respecto al sistema de referencia del espejo, y (b) representación de la función de proyección y del ángulo del rayo óptico con respecto a ρ.

conocer la función de calibración de la cámara.

El error mínimo de calibración se obtiene con un polinomio de función de grado 3, con coeficientes $a_0 = -13194,8877$, $a_1 = 0$, $a_2 = 2,2520 \cdot 10^{-4}$, y $a_3 = 5,8996 \cdot 10^{-8}$. La función queda como:

$$f(\rho) = -13194,8877 + 2,2520 \cdot 10^{-4} \cdot \rho + a_3 = 5,8996 \cdot 10^{-8} \cdot \rho^3 \tag{3.8}$$

El error de calibración es de 1.132 pixeles.

La corrección de la distorsión de este sistema dióptrico se lleva a cabo usando la teoría del centro óptico para conocer la dirección de los rayos proyectados en la imagen, de la mismos forma que se ha descrito en la Sección 3.2 con los sistemas catadióptricos.

La imagen con la distorsión corregida se corresponde a la proyección de los rayos en un plano paralelo al del sensor de la imagen. El modelo se aproxima a la proyección ortográfica de la imagen omnidireccional por la posición relativa del plano del sensor CCD y el plano de proyección.

En la Figura 3.20 se muestra el ejemplo de una imagen capturada con esta cámara, y la misma imagen con la distorsión corregida.

(a) (b)

Figura 3.20: (a) Imagen capturada con cámara GoPro Hero y (b) imagen con distorsión corregida.

Óptica	Tipo de lente	Foco fijo
	Apertura	f/2,8
	Ángulo de visión	170º gran angular en 720p/960p
Foto	Resolución	5 Megapixeles
	Tamaño original	2592 x 1944 pixeles
	Modos de captura	una foto, foto cada 2,5,10,30 o 60 seg., ráfaga de 3 fotos, temporizador
Almacenamiento	Memoria	Tarjeta de memoria SD, de hasta 32 GB
Conectores	Conexión a PC	USB 2.0 (conexión de datos y de carga de batería)
	Salida de TV	HD NTCS y PAL
	Salida Audio	Jack 2.5mm
Alimentación	Batería	recargable de litio 1100 mAh
	Duración	Aprox. 2,25 horas
	Carga	a través de USB de ordenador o adaptador de corriente
	Tiempo de Carga	80 % de su capacidad después de 1 hora
Información Adicional	Dimensiones	Largo x Ancho x Largo 42mm x 60mm x 30mm
	Peso	150 g. con batería

Tabla 3.3: Especificaciones Técnicas de la cámara GoPro Hero 960.

3.4 Plataformas de Adquisición de Imágenes

Para la adquisición de las bases de imágenes, se han empleado dos plataformas distintas. Por un lado, se ha utilizado un robot modelo Pioneer P3-AT para la captura de rutas y mapas densos. Por otro lado, cuando las bases de imágenes han requerido la variación de la altura del sistema de visión con respecto al suelo, se ha empleado un trípode que permite un rango de altura amplio.

A continuación se detallan las características de cada una de las plataformas.

3.4.1 Robot Pioneer P3-AT

El robot Pioneer pertenece a una familia de robots de la compañía americana *Mobile Robots*. El modelo específico utilizado en este trabajo es el P3-AT. Dispone de 4 ruedas motrices, siendo un robot todo-terreno que cubre un amplio rango de aplicaciones robóticas tanto de interior como de exterior. En la Tabla 3.4 se incluyen las características técnicas del robot.

La base del robot incluye un cuerpo correspondiente por 3 baterías, 4 ruedas, 4 motores con encóders, alimentador de los motores a bordo, microcontroladores, bus I/O integrado en el hardware y software ARIA. Los motores contienen encóders con una resolución de 100 ticks por vuelta. La plataforma puede rotar sobre sí misma haciendo uso de las cuatro ruedas en el movimiento de cambio de dirección. En [5] se puede ampliar información sobre el robot y las diferentes posibilidades de configuración.

El Pioneer P3-AT admite una serie de accesorios opcionales. De los complementos opcionales disponibles para el robot, nuestro modelo va equipado con el telémetro láser, los sónares delanteros, un par estéreo, y un sistema de visión omnidireccional. Es posible ver una imagen del robot utilizado en la Figura 3.21.

Las ruedas con las que va equipado son las de interior, aunque dispone de un juego de ruedas para exteriores que evita el deslizamiento de los neumáticos sobre el terreno de navegación. Sin embargo, cabe destacar que es común que se produzcan pequeños deslizamientos que introducen error en las mediciones odométricas del robot, especialmente cuando se producen cambios de dirección durante la navegación.

La comunicación con el robot se hace a través de un PC a bordo del robot. Se hace uso de ROSARIA, una interfaz entre ROS (siglas en inglés de *Robot Operating System*) y la librería ARIA (*Advanced Robot Interface for Applications*) nativa del robot. Dicha interfaz permite el control y la estimación de la posición del robot. La comunicación con el robot se hace a través de instrucciones ARIA, mientras que usa nodos específicos de ROS para los sensores y otros dispositivos que están en el robot.

Físico	Largo	501 mm
	Ancho	493 mm
	Alto	277 mm
	Peso (sin Batería)	14 kg
	Peso (con Batería)	14 kg
	Carga	Hormigón: 20kg, Asfalto: 7kg, Hierba: 13.5kg
Construcción	Cuerpo	Aluminio 1.6 mm
	Acceso Batería	Puerta con bisagra y pestillo
	Montaje	Tornillos Allen
Alimentación	Batería	12V
	Número Baterías	3
	Capacidad	7.2 *Ah* (cada una)
	Composición	Plomo y ácido
	Autonomía	2-4 horas
	Tiempo Carga	6 hrs/batería (cargador estándar) 2.4 hrs/bat (cargador alta capacidad)
Movilidad	Conducción	4 ruedas
	Radio de Giro	0 cm
	Velocidad Máx. Avance	0.7 m/s
	Velocidad Max. Rotación	140 $^{o}/s$
	Escalón máximo	10 cm
	Hueco máximo	15 cm
	Pendiente máxima	35 %
Panel Control	Indicador principal de alimentación	
	Indicador de carga de batería	
	Botón de reseteo	
Accesorios	Sónar delantero y trasero	
	Sistemas de visión Mono y Estéreo	
	Telémetro Laser	
	PC a bordo	
	...	

Tabla 3.4: Especificaciones Técnicas del Robot Pioneer P3-AT

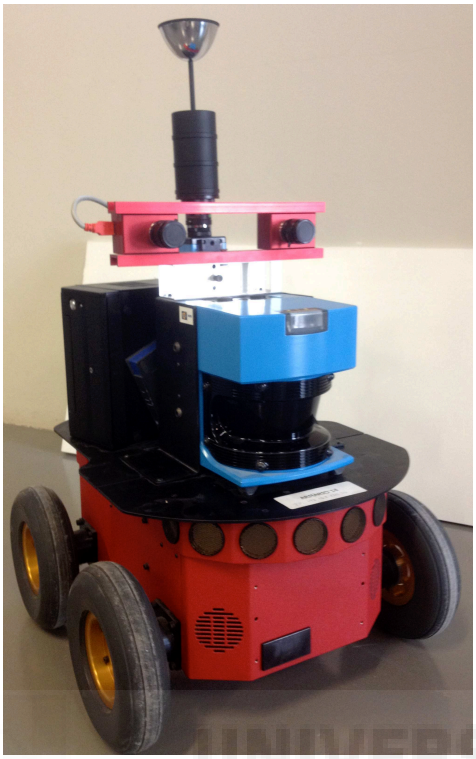


Figura 3.21: Imagen del robot P3-AT utilizado para la captura de imágenes.

3.4.2 **Trípode**

En este trabajo se incluye un estudio comparativo de algoritmos orientaos a la estimación topológica de variación de la altura del sensor visual. Por ello, se ha requerido la adquisición de una base de imágenes que incluyen la variación de la posición de la cámara con respecto al eje Z.

Con el objetivo de obtener un rango de altura suficientemente amplio, se hace uso de un trípode. Las características de dicho trípode vienen recogidas en la Tabla 3.5. El rango total de altura es de 1.720 mm. En [102] es posible ampliar la información sobre el trípode utilizado.

Además, se ha añadido un adaptador que permite la rotación de la cámara alrededor del eje de simetría del espejo, lo que permite variar la orientación de la cámara en el plano XY. También se ha preparado el trípode para poder medir los desplazamientos verticales.

En la Figura 3.22 se presenta una imagen del trípode utilizado y del adaptador del sistema catadióptrico.

Fabricante	König & Meyer
Modelo	20811-409-55
Referencia	M20811B
Diámetro de la base	1.485 mm
Altura Mínima	1.290 mm
Altura Máxima	3.010 mm
Rosca Adaptador	Withworth 3/8”
Material	Acero
Estructura	2 tubos extensibles
Fijación Altura	Tornillo
Peso	9,5 kg

Tabla 3.5: Especificaciones del Trípode K&M 20811.

Figura 3.22: Imagen del trípode K&M 20811 y detalle de la adaptación de sistema catadióptrico.

UNIVERSITAS
Miguel
Hernández

CAPÍTULO

4

Apariencia Global de Información Visual: Descriptores

La riqueza de la información visual hace necesario buscar formas alternativas de describir las escenas, pues los requerimientos de tiempo y memoria hace inviable su uso en la mayoría de aplicaciones reales de forma directa.

En este capítulo se presenta un recopilatorio de técnicas basadas en la apariencia global de las imágenes para obtener descriptores útiles para ser usados en tareas de navegación robótica.

La reducción de la vasta cantidad de información recogida por las imágenes es una de las barreras más importantes del uso de sistemas basados en visión en aplicaciones en tiempo real. El funcionamiento de los sistemas de navegación depende de una representación eficiente de la información proporcionada por el sensor para obtener una asociación entre imágenes que combine rapidez y precisión.

Los trabajos clásicos en robótica móvil usando visión han estado centrados principalmente en la extracción de marcas o puntos característicos para crear un descriptor de la imagen. En este sentido, podemos encontrar trabajos basados en la extracción de características SIFT (Scale-Invariant Feature Transform) como [68, 113, 115, 116].

En [20], Bay et al. presentan SURF (Speed Up Robust Features), junto con una comparación experimental con otros descriptores (como SIFT). Murillo et al. presentan en [136] un sistema piramidal de búsqueda de correspondencias sobre imágenes omnidireccionales para llevar a cabo la localización de un robot móvil de forma topológica y métrica. En [118]

se describe un sistema que combina la información proporcionada por una cámara óptica y otra térmica para extraer características SIFT de ambas imágenes. También se pueden utilizar marcas artificiales colocadas a lo largo del entorno de navegación. En [28] se detalla un sistema de navegación que utiliza códigos identificables por el algoritmo para la localización del robot.

Otro ejemplo de utilización de SURF como descriptor es FAB-MAP [42, 43]. FAB-MAP es un sistema de SLAM probado en rutas de exterior de gran escala, que combina la información de una cámara esférica un par estéreo y un GPS. Scaramuzza et al. proponen en [167] la extracción de las líneas verticales directamente sobre las imágenes omnidireccionales para generar un descriptor de la imagen a partir de dicha información. En [70], Gil et al. presentan un extenso estudio comparativo de distintos descriptores basados en la extracción de puntos característicos y descriptores globales en tareas de SLAM. Sin embargo, este tipo de descriptores suelen tener asociado un alto coste computacional, y ser muy sensibles a oclusiones en el entorno.

Todos las técnicas enunciadas hasta este punto permiten caracterizar la imagen a partir de puntos o regiones significativas de la escena de la escena, añadiendo un descriptor del entornos de vecindad de estas regiones. Frente a estos descriptores, podemos encontrar técnicas que evitan extraer información local de la escena para centrarse en una descripción global de la escena.

Este trabajo se va a centrar en descriptores basados en la apariencia global de las imágenes. Dichas técnicas obtienen información de la imagen en su conjunto, sin centrarse en ninguna región específica ni punto característico, formando vectores que recogen propiedades como la distribución frecuencial de la escena, la dirección de los elementos representados, o información sobre el color de la imagen.

Además, como contribución de esta tesis, en la Sección 4.1.1 se presenta un descriptor basado en la apariencia global, bajo el nombre de Fourier 1D.

A continuación, se van a detallar cada una de las técnicas utilizadas.

4.1 Técnicas basadas en la Transformada de Fourier

El análisis de Fourier es una herramienta matemática utilizada para analizar funciones periódicas a través de su descomposición en una suma infinita de funciones senoidales mucho más simples. Las series de Fourier poseen la forma:

$$y(x) = \frac{a_0}{2} + \sum_{n=1}^{\infty} \left[a_n \cos(nx) + b_n \sin(nx) \right], \tag{4.1}$$

donde a_n y b_n son los denominados coeficientes de Fourier de la transformada de la función $y(x)$.

Por lo tanto, la transformada de Fourier nos permite representar una función cualquiera en el dominio de la frecuencia. La transformada de Fourier de una función contínua se define como:

$$\mathcal{F}\{f(t)\} = F(\omega) = \int_{-\infty}^{+\infty} f(t) e^{-j\omega t} dt, \tag{4.2}$$

siendo e la base de los logaritmos naturales, y j la unidad imaginaria.

Sin embargo, como los datos de partida de este trabajo son de naturaleza discreta, utilizaremos la Transformada de Fourier Discreta, o DFT. Debe tenerse en cuenta que cada vez que se haga mención a la transformada de Fourier, se hará referencia la transformada discreta de Fourier si no se especifica lo contrario.

La transformada de Fourier discreta nos permite expandir la secuencia de números $\{a_n\} = \{a_0, a_1, \ldots, a_{N-1}\}$ en la secuencia compleja $\{A_n\} = \{A_0, A_1, \ldots, A_{N-1}\}$ según la siguiente ecuación:

$$\{A_n\} = \mathcal{F}[\{a_n\}] = \sum_{n=0}^{N-1} a_n e^{-j\frac{2\pi}{N}kn}; k = 0, \ldots, N-1, \tag{4.3}$$

con N igual al número de elementos de la secuencia.

Así pues, toda señal puede representarse mediante su espectro en el dominio de las frecuencias. En el caso de una imagen, existe un paralelismo absoluto entre la misma y su espectro de frecuencias espaciales. La frecuencia espacial, también llamada red sinusoidal, se define como la distribución espacial de iluminaciones, que sigue una ley sinusoidal, caracterizada por su amplitud, fase y frecuencia.

La información obtenida con la transformada de Fourier puede dividirse en dos partes: por un lado, los módulos de los números complejos (que representan su espectro de potencia), y por otro su fase. Además, presenta ciertas propiedades que son muy interesantes. Por ejemplo, la información más relevante se concentra en las bajas frecuencias de la serie, que se corresponden con los primeros términos de la transformada. Por otro lado, las altas

frecuencias suelen estar más afectadas por ruido. Así pues, seleccionando únicamente los primeros elementos de la transformada, conservamos la mayor parte de la información de la serie original, eliminamos frecuencias con mayor probabilidad de estar afectada por ruido, y reducimos la memoria necesaria para almacenar la información.

Otra propiedad muy interesante es la invariancia rotacional. Gracias a esta propiedad, la transformada de Fourier de una serie numérica, y la transformada de esa misma serie rotada un cierto número de posiciones tienen el mismo módulo. Esta propiedad puede demostrarse a través del Teorema del Desplazamiento, que se recoge en la siguiente ecuación:

$$\mathcal{F}[\{a_{n-q}\}] = A_k e^{-j\frac{2\pi qk}{N}}; \quad k = 0, \ldots N-1 \tag{4.4}$$

$\mathcal{F}[\{a_{n-q}\}]$ es la Transformada de Fourier de la secuencia desplazada, y A_k son los componentes de la transformada de la serie no desplazada.

En la Ecuación 4.4 se puede apreciar que la transformada de Fourier de una secuencia desplazada es igual a la transformada de la secuencia original, multiplicada por un número complejo cuya magnitud es la unidad. Por tanto, de acuerdo con esta expresión, la amplitud de los coeficientes de la secuencia rotada es la misma que los de la original, cambiando únicamente su fase. En la figura 4.1(b) se representa el módulo de la transformada de una secuencia de valores (A) y la misma secuencia rotada un elemento (A'). En las gráficas se aprecia que el espectro de potencia es invariante a una rotación circular de la secuencia.

Por otro lado, la fase de los coeficientes nos proporciona información suficiente para estimar la rotación relativa entre las series.

- Fase de la secuencia original:

$$\theta = e^{-j\frac{2\pi k}{N}} \tag{4.5}$$

- Fase de la secuencia desplazada q elementos:

$$\theta' = e^{-j\frac{2\pi qk}{N}} \tag{4.6}$$

Tal y como se puede ver en las ecuaciones (4.5) y (4.6), la diferencia entre fases de cada elemento de la serie de Fourier depende del desplazamiento entre series (q), del números de elementos de la serie N y de la posición del término de la serie en el que se estudia el desfase k. Los términos q y N son fijos para todos los elementos de las transformadas, pero no así el término k, que depende directamente de la posición del elemento. Así pues, la diferencia de fase entre los términos de dos series rotadas será multiplicada por la posición del elemento calculado.

En el ejemplo de la Figura 4.1, el desfase entre las dos series es de un término. Como $N = 10$, este desfase corresponde a $360^o/N = 36^o$. En la Figura 4.1(f) es posible apreciar

cómo este desfase se va multiplicando por la posición del coeficiente de la transformada en el que se calcula la diferencia (k).

La representación en el espacio de la frecuencia de la información visual presenta numerosas posibilidades dependiendo de la información de partida y su tratamiento. Los siguientes apartados recogen cuatro descriptores distintos que concentran la información visual mediante su representación en el dominio de la frecuencia.

4.1.1 Transformada 1D de la Imagen

En [25, 26, 27], Briggs et al. proponen un descriptor que reduce una imagen panorámica en un vector unidimensional para aplicaciones de localización y navegación robótica. El descriptor parte del filtrado de la imagen original con máscaras Gaussianas a distintas escalas. Tras ello, buscando el máximo y el mínimo en los niveles de intensidad de las escenas filtradas a distintas escalas, se extraen puntos característicos que son localmente invariantes ante cambios en la escala de filtrado. Por tanto, no es un descriptor basado en la apariencia global.

Sin embargo, la idea de crear un descriptor unidimensional de una imagen panorámica parte de dichos trabajos. En este punto se propone crear imágenes 1D calculando el valor medio de las columnas de la imagen panorámica. Tras ello, calculamos la transformada de Fourier de la imagen 1D. Como una rotación del móvil en el plano del suelo se traduce en una rotación en el orden de las columnas de la imagen panorámica, tal y como se aprecia en la Figura 4.2 la transformada de Fourier aporta invariancia rotacional al descriptor.

Las ventajas de este método es el bajo coste computacional que tiene asociado. Además, un cambio en la altura de la cámara debido a posibles vibraciones no introduce una variación notable en el nivel medio de intensidad de los píxeles por columna, por lo que el descriptor no se verá afectado de forma importante. Como contrapartida, la reducción de la información visual a un solo vector puede dificultar la obtención de un descriptor fiable y único para cada una de las imágenes dentro de un entorno de navegación.

La Figura 4.3 recoge gráficamente la construcción del descriptor. Como se puede apreciar, el proceso de creación del descriptor incluye el cálculo del valor medio de los píxeles de la imagen panorámica, para luego calcular la transformada de Fourier del vector obtenido.

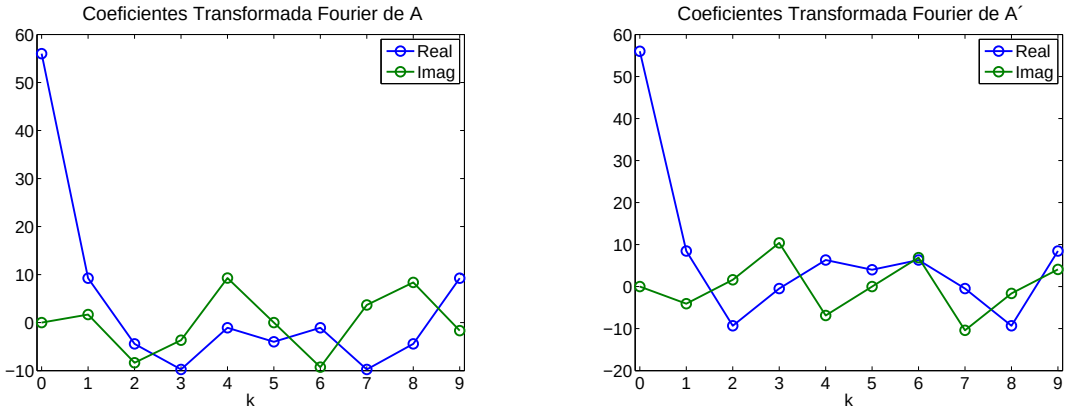
Aplicado a tareas de navegación, la localización se valdrá del módulo de la transformada de Fourier de la serie, ya que es invariante a rotación. En cuanto a la orientación, se estimará usando la fase de los coeficientes y el teorema del desplazamiento (Ecuación 4.4).

Secuencia A

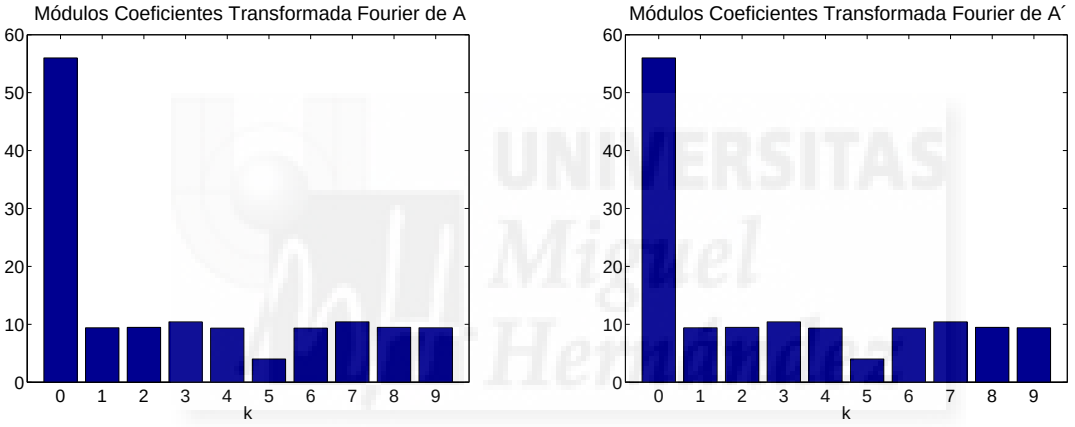
[4 9 10 1 3 5 3 8 6 7]

Secuencia A'

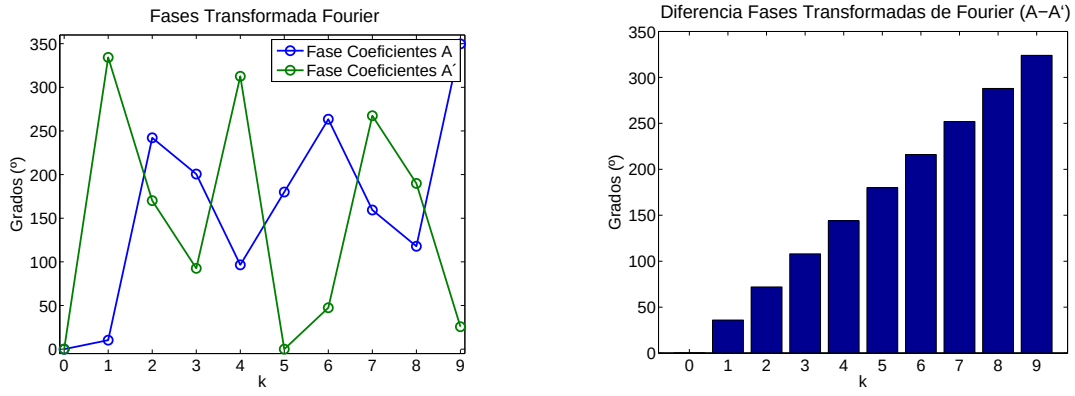
[7 4 9 10 1 3 5 3 8 6]



(a)



(b)



(c) (d)

Figura 4.1: (a) Transformada de Fourier de una serie numérica de 10 elementos (Secuencia A) y la de la misma serie rotando un elemento (Secuencia A'); (b) Módulo de los coeficientes de Fourier; (c) Fase de los coeficientes de ambas series, y (d) Diferencia de fases calculadas entre 0^o y 360^o.

Figura 4.2: Imágenes panorámicas rotadas 68º entre sí.

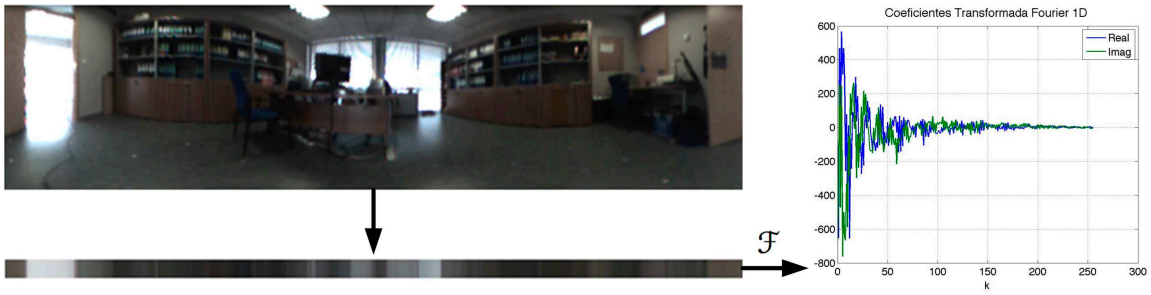


Figura 4.3: Construcción de la Transformada de Fourier 1D de una imagen panorámica.

4.1.2 Firma de Fourier

En [87], Ishiguro y Tsuji proponen la creación de mapas visuales aplicando la Transformada de Fourier sobre imágenes panorámicas. En concreto, sus autores pasan cada una de las filas de la imagen al dominio de la frecuencia. En [127] se desarrolla esta misma idea bajo el nombre de Firma de Fourier (*FourierSignature*).

De esta forma, el descriptor aprovecha las propiedades de la transformada de Fourier de una secuencia de datos. Por un lado, como la información más relevante se concentra en las frecuencias más bajas, se seleccionan únicamente los primeros términos de la secuencia transformada para representar cada fila. En la Figura 4.4 se puede ver gráficamente la distribución de la densidad espectral de potencia de los primeros términos de la Firma de Fourier.

Además, si se trabaja con imágenes panorámicas, se obtiene invariancia rotacional. Un giro en el plano del suelo se traduce en un desplazamiento de sus columnas, es decir, un desplazamiento de los píxeles de cada fila de la imagen a lo largo del eje horizontal. Esta rotación circular del orden de los píxeles no afecta al valor del módulo de la Transformada de Fourier, ya que únicamente varía la fase de los coeficientes, tal y como se ha visto en la ecuación (4.4).

Por tanto, el módulo de la Firma de Fourier nos permitirá buscar la correspondencia entre imágenes indistintamente de su orientación, mientras que la fase nos aporta información sobre el desfase relativo entre escenas.

4.1.3 Transformada 2D de Fourier

La forma transformada de Fourier bidimensional puede ser aplicada directamente sobre la matriz que recoge una imagen digital para pasar la información visual al dominio de la frecuencia.

Si representamos la imagen con la función de elementos discretos $f(x,y)$ con N_x columnas y N_y filas, la transformada de Fourier 2D se define como:

$$\begin{aligned} &\mathcal{F}[\{f(x,y)\}] = F(u,v) = \frac{1}{N_y N_x} \sum_{x=0}^{N_x-1} \sum_{y=0}^{N_y-1} f(x,y) \cdot e^{-j2\pi\left(\frac{ux}{N_x}+\frac{vy}{N_y}\right)} \\ &u = 0,\ldots N_x-1, \quad v = 0,\ldots N_y-1. \end{aligned} \tag{4.7}$$

La transformada de Fourier 2D puede interpretarse como una doble transformada de Fourier de la matriz $f(x,y)$, es decir, una primera transformada de la imagen por filas, seguida de una transformada por columnas de los coeficientes de la primera transformada, o viceversa. Matemáticamente, este concepto viene recogido por la propiedad de separabilidad. En la Ecuación 4.8 se recoge la propiedad de separabilidad aplicada a la Ecuación 4.7.

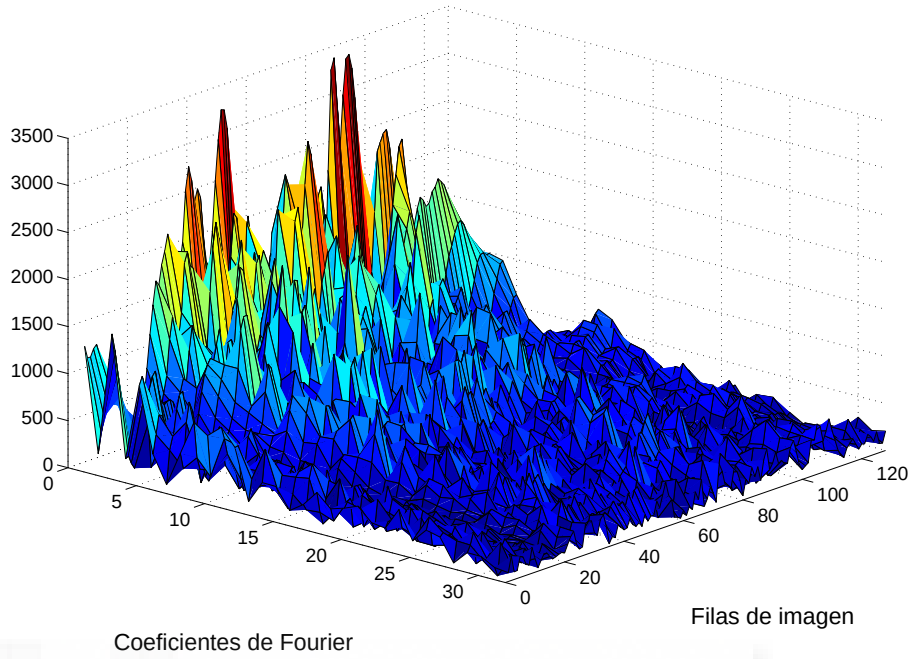


Figura 4.4: Módulos de la Firma de Fourier.

$$\begin{aligned} F(u,v) &= \frac{1}{N_x}\sum_{x=0}^{N_x-1} e^{-j2\pi\left(\frac{ux}{N_x}\right)} \frac{1}{N_y}\sum_{y=0}^{N_y-1} f(x,y)\cdot e^{-j2\pi\left(\frac{vy}{N_y}\right)} \\ &= \frac{1}{N_x}\sum_{x=0}^{N_x-1} F(x,v) e^{-j2\pi\left(\frac{vy}{N_y}\right)} \end{aligned} \tag{4.8}$$

Al igual que en los otros descriptores basados en el dominio de la frecuencia, los componentes de la transformada son números complejos que pueden ser divididos en dos matrices, una con los módulos, y otra con las fases.

Cuando trabajamos con imágenes panorámicas, la transformada de Fourier bidimensional también presenta invariancia ante rotaciones alrededor del eje perpendicular al sistema catadióptrico de adquisición de imágenes. Para Fourier 2D, el teorema del desplazamiento se expresa como:

$$\begin{aligned} &\mathcal{F}[\{f(x-x_0, y-y_0)\}] = F(u,v)\cdot e^{-j2\pi\left(\frac{ux}{N_x}+\frac{vy}{N_y}\right)} \\ &u = 0,\ldots N_x-1, \quad v = 0,\ldots N_y-1. \end{aligned} \tag{4.9}$$

De acuerdo con la Ecuación 4.9, el espectro de potencia de la imagen rotada permanece igual al de la imagen original, cambiando únicamente la fase de los coeficientes de Fourier. El cambio en la fase dependerá de los desplazamientos tanto en el eje x (x_0), como en el eje y(y_0).

Si se mantiene la altura de captura de las escenas, la variación en la fase entre los coeficientes de la transformada de dos imágenes adquiridas en un mismo punto en el plano del suelo será únicamente debida a la rotación del robot alrededor del eje focal de la cámara.

Sin embargo, si la altura del robot no es constante, el cambio de fase por variación de las filas de la imagen y el debido a la variación del orden de las columnas, se sumarán y afectarán conjuntamente a los coeficientes de la transformada.

Cabe destacar que la fase de los coeficientes de la primera fila, que tiene asociada $v = 0$, recoge únicamente la variación en el eje x, mientras que la fase de los coeficientes de la primera columna ($u = 0$), estarán asociados con una variación de la información visual en el eje y.

Los componentes de la transformada de Fourier 2D están ordenados según la frecuencia que representa. En la Figura 4.5(a), podemos ver gráficamente los módulos de la transformada de la imagen que se encuentra a su derecha. Las componentes centrales recogen las bajas frecuencias, mientras que las más alejadas del centro corresponden a las altas frecuencias. En la Figura 4.5(b) hemos filtrado las bajas frecuencias. La imagen resultante en el dominio espacial recoge únicamente las zonas en las que se produce un cambio brusco de intensidad en la escena original. Por contra, si nos quedamos con las bajas frecuencias (Figura 4.5(c)), el resultado en el dominio espacial es equivalente al de un filtro de suavizado, recogiendo la mayor parte de la información de partida. Además, consigue reducir el ruido de la imagen, que suele localizarse en las frecuencias más altas del espectro.

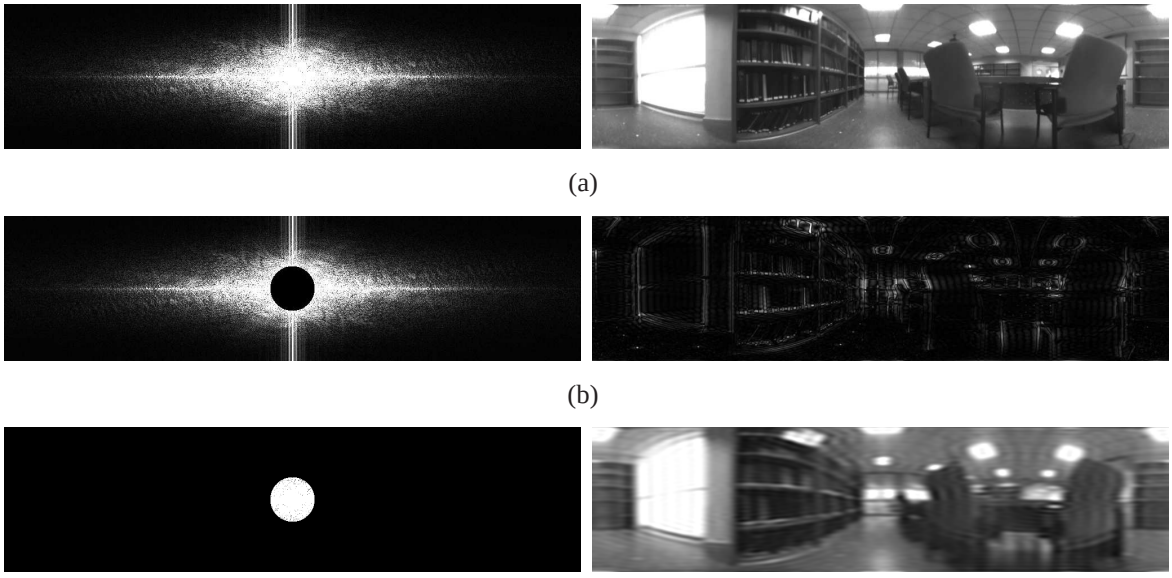
4.1.4 Transformada Esférica de Fourier (SFT)

Tal y como presentan Geyer y Daniilidis [67], una imagen puede ser representada en una esfera unitaria. En dicha esfera, los puntos corresponden con vectores unitarios que indican una dirección de cada pixel en el mundo real con respecto al centro de referencia de la cámara.

La posición de un punto en el espacio tridimensional $\mathbf{x} \in \mathbb{R}^3$ con coordenadas cartesianas expresadas con el vector $\mathbf{x} = (x_1, x_2, x_3)^T$ puede ser también descrito a partir de un sistema de referencia esférico por el vector $(r, \theta, \phi)^T$, siendo $r \in \mathbb{R}^+$ el radio, $\theta \in [0, \pi]$ el ángulo de colatitud y $\phi \in [0, 2\pi)$ el azimut. En la Figura 4.6 podemos ver la relación entre las coordenadas cartesianas y las esféricas. Matemáticamente, la relación entre ambos sistemas de coordenadas se puede expresar como:

$$(x_1, x_2, x_3)^T = (r\sin\theta\cos\phi, r\sin\theta\sin\phi, r\cos\theta)^T$$

En este caso, como la esfera es unitaria, el radio será fijo e igual a la unidad ($r = 1$). Denotaremos la superficie definida por la esfera unitaria bidimensional como $\mathbb{S}^2$. Cada punto



(a)

(b)

(c)

Figura 4.5: Transformada de Fourier 2D y representación en el dominio espacial de una imagen. Escena (a) original, (b) eliminando las bajas frecuencias y (c) eliminando las altas frecuencias.

contenido en esa superficie $P \in \mathbb{S}^2$ puede ser denotado con el vector correspondiente $(\theta,\phi)^T$, indicando la dirección del pixel respecto al centro de referencia de la cámara.

En [49], Driscoll y Healy muestran que las funciones esféricas armónicas Y_{lm} (también llamadas Funciones esféricas superficiales) forman una base ortonormal para $L^2(\mathbb{S}^2)$. $\{Y_{nk} | k \in \mathbb{N}_0, n = -k,\ldots,k\}$ son denominadas las bases estándares de los armónicos esféricos, o base de Fourier.

Por lo tanto, una función cualquiera de cuadrado integrable (es decir, una función cuya integral del cuadrado de su módulo definida sobre cierto intervalo converge) definida en la esfera $f \in L^2(\mathbb{S}^2)$ puede ser representada por su expansión armónica esférica como:

$$f(\theta,\phi) = \sum_{l=0}^{\infty} \sum_{m=-l}^{l} \widehat{f}_{lm} Y_{lm}(\theta,\phi), \tag{4.11}$$

con $l \in \mathbb{N}$ y $m \in \mathbb{Z}$, $|m| \leq l$. $\widehat{f}_{lm} \in \mathbb{C}$ son los coeficientes de la Transformada Esférica de Fourier, o SFT por sus siglas en inglés (*Spherical Fourier Transform*).

Y_{lm} representa la función armónica esférica de grado l y orden m definida como:

$$Y_{lm}(\theta,\phi) = \sqrt{\frac{2l+1}{4\pi}\frac{(l-m)!}{(l+m)!}} P_l^m(\cos\theta) e^{im\theta}, \tag{4.12}$$

donde $P_l^m(x)$ son los coeficientes de la función de Legendre asociada.

En este caso, debido a la naturaleza discreta de las imágenes digitales, debemos aplicar la Transformada Esférica Discreta de Fourier.

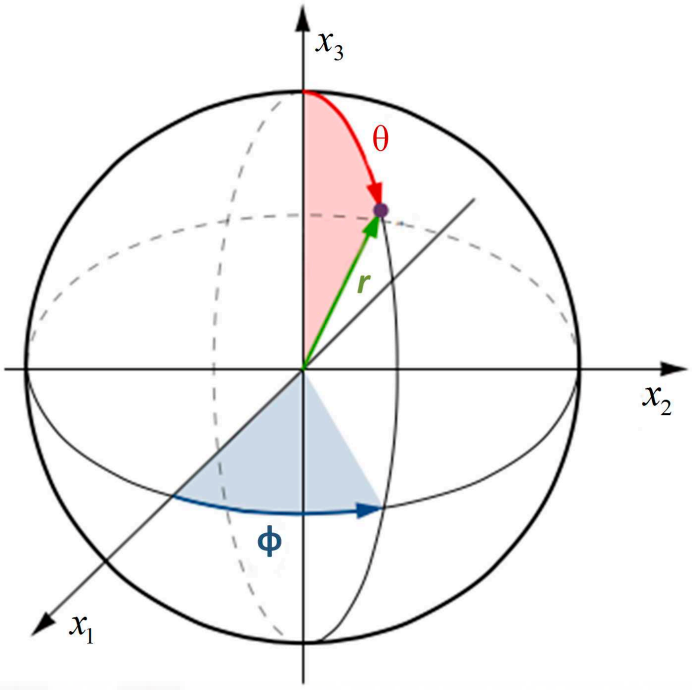


Figura 4.6: Sistema de referencia esférico representado en $\mathbb{R}^3$.

La información de entrada al algoritmo es una matriz que contiene el valor de los píxeles de la imagen seleccionados siguiendo una cuadrícula equiangular de puntos sobre la esfera. En la Figura 4.7 se muestran dos cuadrículas de puntos sobre la superficie esférica. La primera tiene un tamaño de 16x16 elementos, mientras que la segunda tiene 32x32.

Sea $B \in \mathbb{N}_0$ fijo, $N = 2^t$ la mayor potencia de 2 respecto a B, es decir, $t = \lceil \log_2 B \rceil$, y $\chi := (\theta_d, \phi_d)$ un conjunto de nodos sobre $\mathbb{S}^2$ que es el conjunto de muestras sobre la imagen, definido como el producto Cartesiano:

$$\chi = \{\theta_l | l = 0, \ldots, L-1\} \times \{\phi_j | j = 0, \ldots, J-1\} (J, L \in \mathbb{N}) \tag{4.13}$$

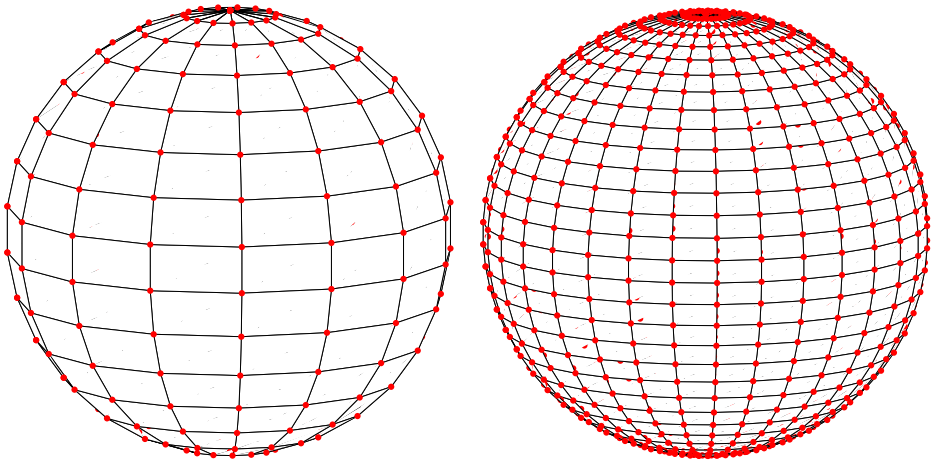
con colatitudes $\theta_l \in [0, \pi]$ y longitudes $\phi_j \in [0, 2\pi)$. $\mathfrak{I}_M$ es el conjunto de índices (l, m) admisibles:

$$\mathfrak{I}_M := \{(l, m) | l = 0, \cdots, B; \quad m = -l, \cdots, l\} \tag{4.14}$$

La función f muestreada puede ser desarrollada en la base de armónicos esféricos $\{Y_{l,m}\}_{(l,m)\in\mathfrak{I}_M}$, con lo que la ecuación 4.15 queda como:

$$f(\theta, \phi) = \sum_{(l,m)\in I_m} \widehat{f}_{lm} Y_{lm}(\theta, \phi) = \sum_{l=0}^{B} \sum_{m=-l}^{l} \widehat{f}_{lm} Y_{lm}(\theta, \phi). \tag{4.15}$$

B corresponde al ancho de banda de frecuencias de la función f en $\mathbb{S}^2$. De acuerdo con



(a) (b)

Figura 4.7: Cuadrícula de puntos equiangular sobre esfera unitaria. (a) 16×16 elementos, (b) 32×32 elementos.

el teorema de muestreo de Nyquist, para conseguir una reconstrucción perfecta de una rejilla de tamaño $2N \times 2N$ necesitaremos una limitación de ancho de banda igual a N.

Como ocurre con el resto de descriptores basados en la transformada de Fourier, es posible obtener un información que es invariante a rotaciones en la imagen con la Transformada Esférica de Fourier.

$Y_{lm}(\theta,\phi)$ para $-l \leq m \leq l$ genera un espacio vectorial que es invariante respecto al grupo rotaciones. Es decir, para una función definida en la superficie esférica con expresión vista en la ecuación 4.15, la magnitud de su proyección sobre el subespacio formado por Y_l se mantiene invariante ante rotaciones.

Tal y como expresan Huhle et al. [86], una rotación en una función esférica no provoca la mezcla de distintas bandas de frecuencia (l). Esto se traduce en que las normas de los distintos coeficientes de los subgrupos de frecuencias sean invariantes ante cualquier rotación 3D de la señal. En [193], este descriptor se recoge bajo las siglas SFD (Spherical Fourier Descriptor), definiéndose como $SFD = [e_1, \ldots, e_l, \ldots, e_B]$, donde

$$e_l = \sqrt{\sum_{m=-l}^{l} \widehat{f}_{lm}} \tag{4.16}$$

El SFD puede ser considerado como un espectro de potencia de cada banda de frecuencias.

En [99], Kazhdan et al. demuestran las propiedades de los armónicos esféricos (incluyendo la invariancia rotacional) e incluyen una comparación con otros descriptores aplicados a la caracterización de objetos 3D.

La implementación mostrada de la Transformada Esférica de Fourier, aplicada a una imagen muestreada con una cuadrícula de $M = N$ puntos, tiene un coste computacional de orden $\mathcal{O}(N^{3/2})$.

En [104], Kosteleck y Rockmore presentan la Transformada de Fourier sobre una función definida sobre el grupo de rotación SO(3) (SOFT). El grupo de rotaciones en $\mathbb{R}^3$ con respecto al origen SO(3) es un grupo de matrices ortogonales cuya determinante es igual a la unidad.

Las rotaciones se definirán usando la descomposición ZYZ de Euler. Dicha descomposición representa la rotación del sistema de coordenadas a través de 3 rotaciones elementales: una primera rotación sobre el ejes, una segunda sobre el eje y del sistema resultante, y una última sobre el eje z del sistema de referencia anterior. Siendo R_z y R_y rotaciones sobre los ejes z e y respectivamente:

$$R_z(\alpha) = \begin{pmatrix} \cos\alpha & -\sin\alpha & 0 \\ \sin\alpha & \cos\alpha & 0 \\ 0 & 0 & 1 \end{pmatrix} \qquad R_y(\beta) = \begin{pmatrix} \cos\beta & 0 & \sin\beta \\ 0 & 1 & 0 \\ -\sin\beta & 0 & \cos\beta \end{pmatrix} \tag{4.17}$$

con $0 \leq \alpha \leq 2\pi$ y $0 \leq \beta \leq \pi$. Usando estas dos matrices, cualquier rotación $R \in SO(3)$ puede ser escrita como:

$$R = R_z(\alpha)R_y(\beta)R_z(\gamma). \tag{4.18}$$

Para cada $R \in SO(3)$, se puede asociar un operador lineal $\Lambda(R)$ que actúa sobre una función $f(\theta,\gamma) = f(\omega)$ en $L^2(\mathbb{S}^2)$:

$$\Lambda(R)f(\omega) = f(R^{-1}\omega). \tag{4.19}$$

En ese mismo trabajo, Kosteleck y Rockmore demuestran que es posible encontrar el desfase relativo entre dos imágenes desde su representación en la esfera unitaria.

Siendo I_1 e I_2 dos imágenes capturadas en un mismo punto, nuestro interés es hallar la rotación r tal que alinea I_2 con I_1, es decir, $I_1 = \Lambda(r)I_2$.

La correlación entre las dos imágenes (I_1 e I_2) se puede expresar como:

$$C(R) = \int_{\mathbb{S}^2} I_1(\omega)\overline{\Lambda(R)I_2(\omega)}d \tag{4.20}$$

En lugar de evaluar la función expresada en la Ecuación 4.20 con todas las posibles rotaciones, podemos buscar la rotación R que maximiza la correlación a través de la Transformada de Fourier sobre $SO(3)$.

Sustituyendo la expansión de la Ecuación 4.15 de las funciones I_1 e I_2, la Ecuación 4.20 queda como:

$$C(R) = \int_{\mathbb{S}^2} \left[\sum_{l=0}^{B} \sum_{m=-l}^{l} \widehat{I}_{1_{lm}} Y_{lm}(\theta,\phi) \right] \overline{\left[\Lambda(R) \sum_{l'=0}^{B} \sum_{m'=-l'}^{l'} \widehat{I}_{2_{l'm'}} Y_{l'm'}(\theta,\phi) \right]} \tag{4.21}$$

Siendo $\widehat{I_1}$ e $\widehat{I_2}$ los coeficientes de la Transformada esférica de Fourier.

Ordenando los productos y teniendo en cuenta que para la familia de armónicos esféricos

$$\Lambda(R) Y_{lm}(\theta,\phi) = \sum_{k=-l}^{l} Y_{lk}(\theta,\phi) D^{l}_{km}(R) \tag{4.22}$$

con $D^{l}_{km}(R)$ la denotando la función Wigner-D [194], la Ecuación 4.21 puede reescribirse como:

$$C(R) = \sum_{l=0}^{B} \sum_{m=-l}^{l} \sum_{m'=-l}^{l} \widehat{I}_{1_{l-m}} \overline{\widehat{I}_{2_{l-m}}} (-1)^{m-m'} D^{l}_{mm'}(R). \tag{4.23}$$

Tomando la inversa de la SOFT del resultado obtenido en la Ecuación 4.23, podemos evaluar la correlación existente entre dos imágenes considerando una cierta rotación R. Siendo B la banda de frecuencias a la que se limita la Transformada Esférica de Fourier, podremos evaluar $C(R)$ en un conjunto de $2B \times 2B \times 2B$ ángulos de Euler.

Por tanto, la resolución angular está determinada por B, teniendo un error para α y γ igual a $\left(\frac{180}{2B}\right)$, y para β igual a $\left(\frac{90}{2B}\right)$.

Simuladas todas las rotaciones posibles, el valor máximo de la inversa SOFT de $C(R)$ denotará la máxima correlación entre señales, siendo R la rotación relativa entre ambas imágenes.

Para calcular los coeficientes de la SFT, en este trabajo se hace uso de la librería SSHT [126]. Esta librería tiene implementadas distintas rutinas para el cálculo de la SFT. Sus autores incluyen el algoritmo basado el trabajo de Driscoll y Healy (DH)[49], otro basado en la cuadratura de Gauss-Legendre (GL) [172] y su propia rutina para realizar el muestreo y obtener la transformada esférica de forma óptima (MW). En este trabajo, McEwen y Wiaux proponen una aproximación que involucra una extensión de la esfera bidimensional $\mathbb{S}^2$ a un toroide bidimensional (T^2), lo cual les permite hacer uso de la Transformada Rápida de Fourier (o FFT), para reducir el coste computacional del algoritmo.

Su teoría se deriva de dos trabajos previos de McEwen [125] (que sufre de inestabilidades para bandas de frecuencia menores a 32), y de [85]. También presentan una comparación del coste computacional de las tres propuestas. Su método (MW), emplea la mitad de tiempo que GL al calcular la SFT, aunque es un 25% más lento que DH. Cabe destacar que la aproximación que utilizan de DH es una versión mejorada con respecto a la inicial [79], basado en una aproximación "divide y vencerás"que reduce la complejidad a $\mathcal{O}(N \log^2 N)$. Sin embargo, los resultados numéricos recogidos en [126] muestran que tanto GL como DH

se vuelven inestables para límites de frecuencia entre $L = 1024$ y $L = 2048$. Por ello, en este trabajo se hará uso de la aproximación de McEwen y Wiaux.

Friedrich et al. detallan en [62] un algoritmo de localización para navegación robótica usando los armónicos esféricos. Incluyen la estimación de la orientación usando la SFT, centrándose en el caso especial en el que se produce únicamente rotación en el plano del suelo, es decir, rotación en el eje z. En este caso, el desfase entre imágenes se convierte en un caso simplificado unidimensional. La parte experimental se realiza con la adquisición de datos en un entorno artificial que se corresponde con un área de oficinas. En [63], amplían la parte experimental, incluyendo una base de datos real adquiridas con una cámara hemisférica de bajo coste consistente en una cámara cuya lente es una mirilla de puerta.

Makadia y Daniilidis abordan la estimación de la rotación 3D a partir de la SFT en los trabajos [120, 121, 122]. Al igual que en este trabajo, ellos parten de la imagen obtenida con un sistema catadióptrico de adquisición de imágenes, no una cámara esférica, por lo que al realizar la proyección de la información en la esfera unitaria, hay ángulos en los que no exige información.

Por otro lado, Schairer et al. presentan varios trabajos relacionados con la estimación de la orientación usando la SFT. En [168] se incluye una mejora en la precisión de la orientación basada en una normalización de la correlación entre los valores de la SFT al comparar las señales rotadas. En [169] se propone la introducción de un filtro de partículas a dicha tarea. Por último, en [170], sus autores desarrollan un sistema de navegación que combina la información odométrica con la localización usando la SFT sobre imágenes de muy baja resolución (32x32 píxeles o menos).

La Transformada Esférica de Fourier presenta una complejidad mayor que los otros descriptores basados en el dominio de la frecuencia, lo que conlleva un mayor coste computacional. Además, como necesita la proyección de la imagen sobre la esfera unitaria, requiere la calibración del sistema catadióptrico. Sin embargo, la posibilidad de estimar su orientación no sólo con rotaciones en el plano del suelo sino con cambios 3D en la orientación sobre un punto, lo convierten en un descriptor muy interesante en aplicaciones de navegación robótica.

4.2 Técnicas basadas en el Análisis de Componentes Principales (PCA)

En este apartado vamos a estudiar la aplicación del Análisis de Componentes Principales, más conocido por sus siglas en inglés PCA (Principal Component Analysis) sobre la información visual. Concretamente, primero se describe la técnica. Posteriormente se presenta una variante matemática que permite la reducción de su complejidad computacional. El siguiente punto incluye una modificación de PCA que permite incluir nueva información sin necesidad de volver a estimar la descomposición SVD (siglas en inglés de *Singular Value Decomposition* o *Descomposición en Valores Singulares*) de la base. Luego se expone una nueva variante del algoritmo que permite tratar la rotación de imágenes panorámicas usando el Análisis de Componentes Principales. Por último, se propone aplicar PCA sobre la información visual representada en el espacio de la frecuencia para aprovechar las ventajas de este dominio.

4.2.1 Análisis de Componentes Principales (PCA)

El Análisis de Componentes Principales es una técnica utilizada para reducir la dimensionalidad de un conjunto de datos. Intuitivamente la técnica sirve para determinar el número de factores subyacentes explicativos tras un conjunto de datos, que expliquen la variabilidad de dichos datos. Por lo tanto, se trata de una técnica estadística de síntesis de información, o reducción de dimensiones de las variables, con la menor pérdida de información posible, consiguiendo una compresión de los datos de partida.

Ante un banco de datos con muchas variables, el objetivo es reducir la dimensión de dichas variables, obteniendo una compresión de la información original.

Técnicamente, PCA busca la proyección de los datos en un nuevo espacio mediante el cual queden mejor representados en términos de mínimos cuadrados. Este método se emplea sobre todo en análisis exploratorio de datos y para construir modelos predictivos, basándose en la descomposición en autovalores de la matriz de covarianza, normalmente tras centrar los datos en la media de cada atributo.

Los orígenes de PCA se sitúan en el trabajo de Pearson [157]. Fue formulado con detalle por Hotelling [82] en 1933 y desde entonces es utilizado en numerosas aplicaciones. Cabe destacar el éxito del PCA en su aplicación a tareas de reconocimiento e identificación de caras humanas a partir de los trabajos de Turk y Pentland [187].

Kirby y Sirovich [5] fueron los pioneros en utilizar PCA como método de compresión de la información en una aplicación de reconocimiento visual mediante la apariencia.

Turk y Pentland [187] observaron que en tareas de descripción y reconocimiento facial basado en la apariencia, es posible utilizar un subespacio de dimensiones reducidas obtenido tras realizar una extracción de características PCA del espacio imagen original. Otra aportación importante de su trabajo fue lo que se conoce actualmente como método de Turk y Pentland (*Turk and Pentland's trick*), que permite disminuir el coste computacional del método PCA.

Murase y Nayar desarrollan en [135] un sistema de reconocimiento basado en la apariencia para reconocer e identificar la pose de objetos de forma libre usando PCA.

Tal y como se ha expuesto anteriormente, PCA construye una transformación lineal sobre el conjunto de datos iniciales para escoger un nuevo sistema de coordenadas. En el nuevo sistema, la mayor varianza de la información se corresponde con el primer eje (también llamado Primer Componente Principal), la segunda varianza de mayor valor se corresponde con el segundo eje, y así sucesivamente, ordenando los vectores principales por orden decreciente de varianza.

Para construir esta transformación lineal, debe obtenerse la matriz de covarianza (o matriz de coeficientes de correlación). Debido a la simetría de esta matriz, existe una base completa de vectores propios que forman un subespacio al que llevar los vectores iniciales de la información.

La transformación que lleva las antiguas coordenadas a la nueva base de proyección nos permite reducir la dimensionalidad de los datos. En el nuevo espacio, los primeros términos de la proyección recogen la mayor parte de la variabilidad de los datos, pues los vectores sobre los que están proyectados tienen asociada la mayor varianza del conjunto. Por tanto, con únicamente los primeros términos de la proyección de la imagen en el nuevo espacio, es posible describir la escena de forma que podamos identificarla de entre un conjunto de imágenes .

Con PCA, además de encontrar una transformación lineal a un espacio de dimensión menor, la señal se reconstruye con un error cuadrático medio mínimo. Gracias a este proceso, además de la compresión de la información, PCA favorece la reducción del ruido sobre el conjunto de datos inicial.

A continuación, se va a incluir el desarrollo matemático de PCA calculado por maximización de la varianza. Para calcular el nuevo subespacio, la información debe ordenarse en una matriz, siendo cada columna un elemento distinto de la base de datos. Partiendo de un conjunto de N imágenes con M píxeles cada una $\vec{x_j} \in \mathbb{R}^{Mx1}, j = 1,\ldots,N$, considerando $M \gg N$, con los cuales se construye la matriz $X = [\vec{x_0}|\vec{x_1}\ldots|\vec{x_{N-1}}] \in \mathbb{R}^{M\times N}$. A partir de esta matriz:

- Se calcula la media de los N vectores $\vec{x}_j$

$$\vec{\mu} = \frac{1}{N}\sum_{j=1}^{N}\vec{x}_j \tag{4.24}$$

- Se resta la media a cada uno de los vectores para obtener su valor centrado en el origen

$$\vec{\widehat{x}}_j = \vec{x}_j - \vec{\mu} \quad \text{para} \quad j = 1,\ldots,N \tag{4.25}$$

- Utilizando los vectores obtenidos en el paso anterior, se construye la matriz $\widehat{X} \in \mathbb{R}^{M\times N}$

$$\widehat{X} = [\,\vec{\widehat{x}}_1 \quad \vec{\widehat{x}}_2 \quad \ldots \quad \vec{\widehat{x}_N}\,] \tag{4.26}$$

- Calculamos la matriz de covarianza de la matriz $\widehat{X}$

$$C = \frac{1}{N} \quad \widehat{X}\cdot\widehat{X}^T \tag{4.27}$$

con $C \in \mathbb{R}^{MxM}$

La matriz de covarianza es cuadrada, simétrica y definida positiva, lo que la hace diagonalizable y permite su descomposición en valores propios no negativos. Además su rango no es máximo.

Esta última propiedad viene derivada de que que la matriz C es resultado de la multiplicación de la matriz $\widehat{X}$ por su traspuesta. Aunque la dimensión de C es $M \times M$, su rango es $N \ll M$, ya que $\widehat{X} \in \mathbb{R}^{M\times N}$. Por ello, el número de valores propios distintos de cero será N (λ_1, λ_2, ...,λ_N), más un último valor propio $\lambda_{N+1} = 0$ de multiplicidad algebraica $(M-N)$.

Esto se cumplirá siempre que los vectores iniciales $\vec{x}_j$ con $j = 1,2,\ldots,N$ sean linealmente independientes, o dicho de otro modo, que la matriz $\widehat{X}$ sea de rango máximo (N).

A partir de los $N+1$ valores propios, por la propiedad de simetría de la matriz C, se obtienen M vectores propios $\vec{u}_i \in \mathbb{R}^{M\times 1}$, con $i = 1,2,\ldots,M$, ortonormales y linealmente independientes entre ellos. Estos vectores forman la matriz cambio de base

$$U = [\vec{u}_1 \quad \vec{u}_2 \quad \ldots \quad \vec{u_M}], \tag{4.28}$$

siendo $U \in \mathbb{R}^{M\times M}$ ortonormal, con $U^{-1} = U^T$.

Los valores propios están asociados con la varianza de los datos. El mayor valor propio denota la mayor varianza de los datos, y el vector propio correspondiente indica la dirección de dicha varianza. Esto ocurrirá igual con el con el resto de valores y vectores propios. Por lo

tanto, la mayor variabilidad de los datos está reflejada en los primeros vectores propios. Dichos vectores suelen denominarse también *vectores principales*, o también *ejes* o *direcciones principales*.

Como $M \gg (N+1)$, existen valores propios con más de un vector propio asociado. Cada valor propio tiene tantos vectores asociados como dimensión de su núcleo, que se calcula como:

$$\dim ker[C - \lambda_j \cdot I] = \dim[C - \lambda_j \cdot I] - rg[C - \lambda_j \cdot I] \tag{4.29}$$

A partir de la ecuación 4.29, es posible demostrar la multiplicidad algebraica del valor propio $\lambda_N = 0$

$$\dim ker[C - \lambda_N \cdot I] = \dim ker[C - 0 \cdot I] = \dim[C] - rg[C] = M - N. \tag{4.30}$$

Todos los valores propios distintos de 0 tienen multiplicidad algebraica igual a uno (tienen asociado un vector propio). Además, de los M vectores propios, únicamente $N \ll M$ resultan útiles, ya que el resto no aporta información al estar asociados al valor propio cero (la varianza asociada es nula).

De esta forma, se obtiene un nuevo espacio vectorial distinto al inicial cuya base la forman los N vectores propios linealmente independientes asociados a valores propios mayores a cero $\vec{u}_j \in \mathbb{R}^{Mx1}$ con $j = 1, 2, \ldots, N$. La matriz cambio de base se forma con los vectores propios útiles ordenados por columnas:

$$U_N = [\ \vec{u}_1 \quad \vec{u}_2 \quad \ldots \quad \vec{u}_N\ ], \tag{4.31}$$

siendo $U_N \in \mathbb{R}^{M \times N}$.

A continuación, se proyecta la información original en el nuevo espacio:

$$\widehat{Y} = U_N^T \cdot \widehat{X} \tag{4.32}$$

donde

$$\widehat{Y} = [\quad \vec{\widehat{y}}_1 \quad \vec{\widehat{y}}_2 \quad \ldots \quad \vec{\widehat{y}}_N \quad] \tag{4.33}$$

con $\vec{\widehat{y}}_j \in \mathbb{R}^{N \times 1}$ la proyección del vector $\vec{\widehat{x}}_j$ en la nueva base.

Luego con el cambio de base realizado, se consigue expresar los N vectores de dimensión $M \times 1$ en N vectores de dimensión $N \times 1$, siendo $N \ll M$, logrando reducir la dimensionalidad de la información.

Si nos quedamos con los N vectores asociados a valores propios distintos de 0, no se produce pérdida de información.

4.2.2 Variante al desarrollo matemático de PCA

Existe una variación en el desarrollo del Análisis de Componentes Principales expuesto en la Sección 4.2.1 que permite obtener el mismo resultado de una forma computacionalmente más eficiente. El ahorro computacional se produce al trabajar con matrices de un tamaño más reducido al realizar la descomposición SVD. Esta variación está basada en el denominado *Método de Turk y Pentland* [187].

El cambio respecto al proceso explicado anteriormente parte de la obtención de la matriz de covarianza C (Ecuación 4.27). La matriz C es de dimensión $M \times M$, pero de rango N, al ser resultado del producto de una matriz de rango N por su traspuesta. La propuesta es calcular una matriz de covarianza distinta:

$$C' = \frac{1}{N} \quad \widehat{X}^T \cdot \widehat{X} \tag{4.34}$$

De esta forma, la matriz de covarianza C' tiene el mismo rango que dimensión, ya que $C' \in \mathbb{R}^{N \times N}$. La nueva matriz de covarianza también es cuadrada, simétrica y definida positiva. Además, su descomposición en valores propios tiene como resultado los N primeros valores propios que se obtienen de la matriz C (Ecuación 4.27). Es decir, los valores propios de C' son los mismos que los N primeros valores propios de C ($\lambda_1, \lambda_2, \ldots, \lambda_N$). Como $C \in \mathbb{R}^{M \times M}$ y $C' \in \mathbb{R}^{N \times N}$, la díagonalización de C' es considerablemente más rápida, pues $N \ll M$.

Sin embargo, la dimensión de los vectores propios $\vec{u'_j} \in \mathbb{R}^{N \times 1}$ obtenidos de la descomposición de C' no tienen la misma longitud que los elementos de entrada ($\vec{\widehat{x}_j} \in \mathbb{R}^{M \times 1}$). Por lo tanto, no pueden componer la matriz cambio de base de forma directa.

Para poder hacer uso de ellos, previamente hay que aplicar una transformación que convierta los N vectores $\vec{u'_j} \in \mathbb{R}^{N \times 1}$ de la descomposición de C' en un conjunto de N vectores de dimensión $M \times 1$ ($\vec{u}_j \in \mathbb{R}^{M \times 1}$).

La adaptación de la dimensión de los vectores se lleva a cabo mediante el *Método de Truk y Pentland*, que consiste en la multiplicación del conjunto de vectores propios $U' = [\, \vec{u'_1}\ \vec{u'_2} \ldots \vec{u'_N}\,]$ por la matriz de datos normalizada $\widehat{X}$:

$$U_N = \widehat{X} \cdot U' \tag{4.35}$$

con $U_N \in \mathbb{R}^{M \times N}$.

Las columnas normalizadas de U_N son, por lo tanto, vectores de dimensión $M \times 1$ que constituyen una base válida para la proyección de la base original en el nuevo espacio:

$$\widehat{Y} = (U_N)^T \cdot \widehat{X} \tag{4.36}$$

con $\widehat{Y} = [\, \vec{\widehat{y}_1}\ \vec{\widehat{y}_2}\ \ldots\ \vec{\widehat{y}_N}\,]$, donde $\vec{\widehat{y}_j} \in \mathbb{R}^{N \times 1}$, es la proyección del vector $\vec{\widehat{x}_j}$ en el espacio obtenido con PCA.

Por lo tanto, el espacio obtenido del Análisis de Componentes Principales usando esta variante matemática es el mismo que en el caso anterior. Sin embargo, en este segundo método se realiza la diagonalización de la matriz $C' \in \mathbb{R}^{N \times N}$ en lugar de la matriz $C \in \mathbb{R}^{M \times M}$.

En el caso de aplicaciones que usen información visual, M corresponde con el número de píxeles de la imagen, mientras que N es el número de escenas incluidas en la base, cumpliendo en la mayoría de aplicaciones la suposición de partida $M \ll N$.

En la práctica, se podrá reducir aún más la dimensionalidad de la base de proyección con una pérdida de información admisible para seguir llevando a cabo tareas de reconocimiento y asociación de imágenes. Utilizando los $k < N \ll M$ vectores con mayor valor propio asociado, se obtiene un subespecie de menor dimensión pero que sin embargo mantiene la mayor parte de la varianza original de los datos.

En trabajos como [135, 139, 156], es posible comprobar aplicaciones de navegación que utilizan subespacios de proyección con pérdida de información. De esta forma, se consigue una representación todavía más eficiente de los datos de partida.

4.2.3 Análisis de Componentes Principales de Modo Incremental (PCAI)

Los métodos de Análisis de Componentes Principales vistos anteriormente necesitan disponer de toda la información antes de proceder al cálculo de la nueva base de proyección. Por lo tanto, la creación del nuevo espacio se convierte obligatoriamente en un proceso *off-line*, ya que no es posible incluir nueva información a la base de forma gradual.

Si se desease incluir un nuevo vector de información $\vec{x}_j$, sería necesario volver a calcular la matriz de covarianza y la descomposición SVD de la matriz resultante. Esto implicaría almacenar la información original de partida durante todo el proceso, y un coste computacional que haría inviable la utilización de PCA en tareas tales como SLAM.

El Análisis de Componentes Principales en modo Incremental (PCAI) es una evolución de PCA que mantiene el fundamento teórico, características y ventajas de PCA, pero que permite incluir nueva información al conjunto de datos iniciales sin la necesidad de volver a realizar el cálculo de los autovectores. En [16, 117] es posible encontrar información detallada sobre el procedimiento para incluir nuevos vectores de información al conjunto original, y calcular el nuevo espacio de proyección.

Así pues, para aplicaciones en las que sea necesario añadir nueva información conforme vaya desarrollándose el proceso, será posible usar PCAI para reducir la dimensionalidad de los datos como técnica de compresión. Así, PCAI incrementará el número de vectores de información que forman parte de la base de datos.

4.2.4 PCA Rotacional

Las técnicas de compresión de la información basadas en PCA mostradas hasta este punto son algoritmos robustos para llevar a cabo el análisis de la información y la reducción de su dimensión.

Sin embargo, la proyección de una imagen y la proyección de una rotación de la misma escena tienen coordenadas completamente distintas en el nuevo espacio, resultando imposible reconocer una posición en el mapa si la imagen de entrada tiene una orientación diferente de la almacenada en el mapa.

Por lo tanto, si aplicamos el análisis directamente sobre la matriz que contiene los datos, la base contendrá únicamente información relativa a la orientación de la escena incluida en la matriz. Dicho de otra forma, PCA no presenta invariancia ante rotaciones de las imágenes.

Para solucionar este problema, Jogan y Leonardis [90, 91] proponen el *Subespacio de imágenes rotadas* (*Eigenspace of Spining-Images*). El algoritmo realiza una serie de rotaciones artificiales de cada una de las escenas incluídas en la base.

La matriz de información incluye pues N rotaciones en el plano del suelo de cada posición incluida en el mapa. Aplicaremos el Análisis de Componentes Principales sobre esta matriz. De esta forma, no solo se obtiene una representación capaz de resolver el problema de la orientación de la imagen, sino que la proyección en el nuevo espacio es más robusta en tareas de asociación de imágenes.

Cuando se trabaja con un conjunto de imágenes rotadas, la descomposición SVD de la matriz que contiene los datos presenta propiedades específicas que simplifican el cálculo del nuevo espacio de proyecciones.

Definimos $\vec{x_j} \in \mathbb{R}^{Mx1}$ como el vector que contiene los píxeles de una imagen panorámica, con j denotando la rotación artificial de la imagen original un ángulo igual a $\theta = j \cdot \frac{360}{N}$, y $R = [x_0|x_1|\ldots|x_{N-1}]$, la matriz cuyas columnas contienen en conjunto de rotaciones artificiales de la escena x_0.

La matriz de covarianza de R, definida como $Q = R^T \cdot R$, tiene la forma de matriz circular. Esta propiedad aparece independientemente del número de rotaciones incluidas de la imagen, siempre y cuando el desfase entre rotaciones consecutivas sea constante.

En la Figura 4.8 se muestra dos ejemplos de matrices circulares compuestas por (a) 32 rotaciones y (b) 128 rotaciones de una misma imagen.

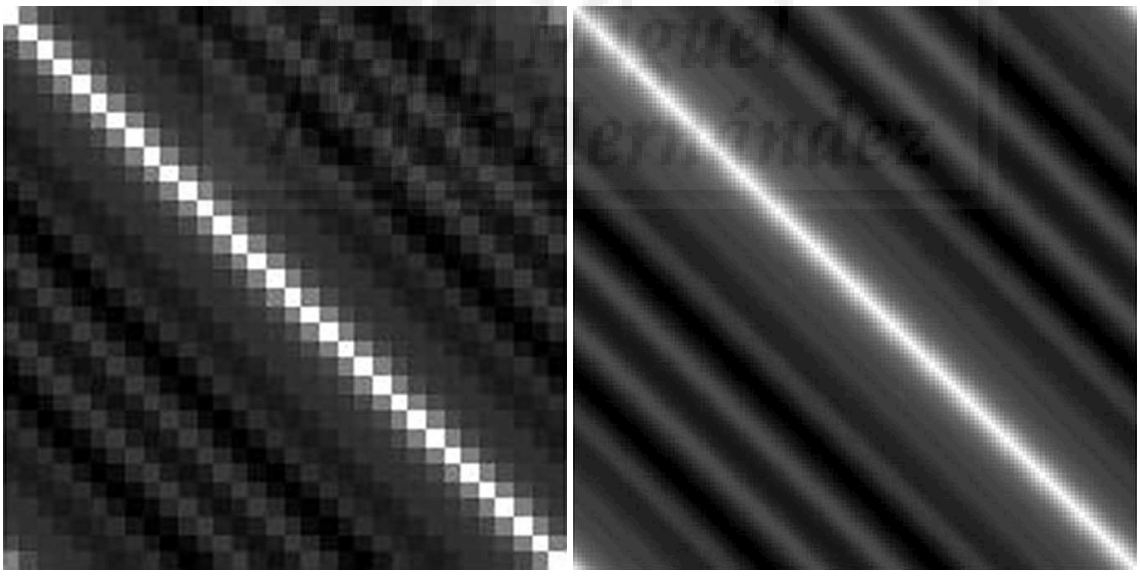
En [188], Ueonara y Kanade demuestran que los vectores propios de una matriz circular son independientes de su información, y corresponden a los vectores de la base de la Matriz de Fourier. Sin embargo, los valores propios asociados a dichos vectores sí que dependen de los datos de la matriz. Ordenando los vectores propios de forma decreciente con respecto a su

valor propio correspondiente, es posible de nuevo obtener una base en la que la proyección de la información contenga la mayor parte de la varianza de los datos en pocos términos.

El problema puede ser extendido a P localizaciones con N rotaciones por escena. Siendo $X = [R_1|R_2|\ldots|R_P]$ la matriz que contiene todas las imágenes de la base con sus correspondientes rotaciones, el producto interior $A = X^T \cdot X$ se compone de $P \times P$ bloques circulantes cuyo tamaño es $N \times N$. En la Figura 4.9 es posible ver una representación gráfica de una matriz de producto interior que incluye 128 rotaciones de 5 imágenes distintas.

Aplicando de nuevo las propiedades de las matrices circulantes, el cálculo de la descomposición en valores singulares de la matriz A se reduce a resolver N descomposiciones de orden P. Como el orden de P para este tipo de problemas es considerablemente menor que el de A, el Análisis de Componentes Principales es más efectivo en términos de coste computacional.

Seleccionando únicamente los vectores propios con mayor valor propio asociado se obtiene un subespacio que permite una representación reducida de la imagen, pero que recoge la mayor varianza de los datos, por lo que no pierde la capacidad de identificar las diferentes escenas.



(a) (b)

Figura 4.8: Matriz de covarianza de un conjunto de (a) 32 rotaciones y (b) 128 rotaciones de una misma imagen.

La representación que se obtiene de las imágenes en el nuevo subespacio está en el plano complejo. Se puede demostrar que los coeficientes de una imagen y sus diferentes rotaciones tienen el mismo módulo, cambiando únicamente su fase.

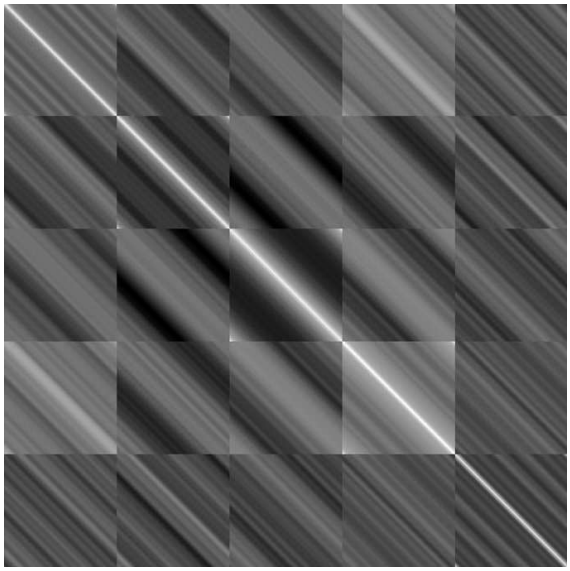


Figura 4.9: Producto interior de una matriz que incluye P=5 localizaciones y N=128 rotaciones.

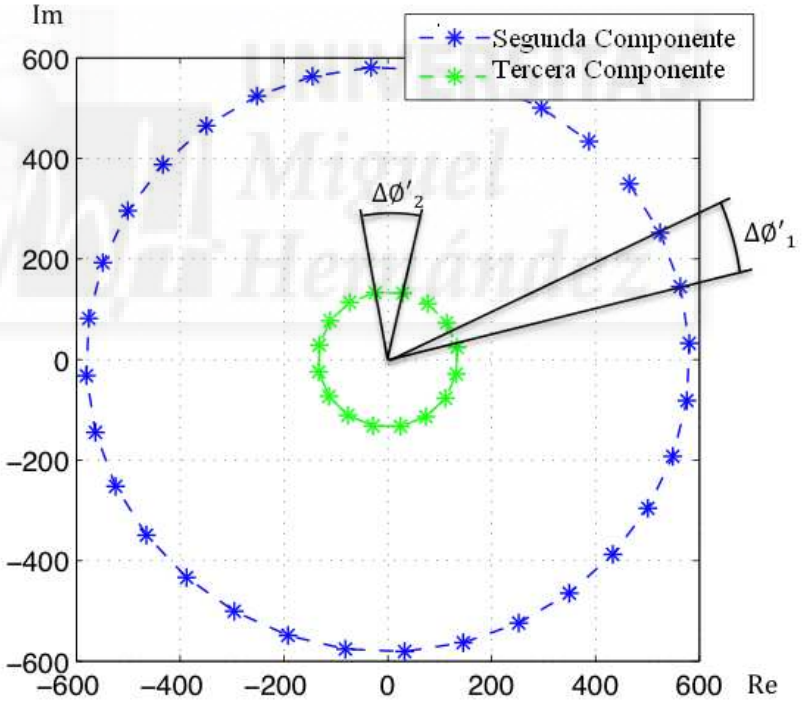


Figura 4.10: Proyecciones de 2 componentes del conjunto de rotaciones de una imagen utilizando PCA Rotacional.

En la Figura 4.10 se recoge la proyección en el nuevo espacio de dos componentes de un conjunto de rotaciones de una imagen en el plano complejo. En ella, se aprecia que el desfase de los coeficientes entre rotaciones consecutivas es constante. Por lo tanto, sólo se necesita almacenar la proyección de una orientación de cada escena, y el desfase de los diferentes

componentes entre rotaciones consecutivas. De esta forma, es posible simular artificialmente las proyecciones de las distintas rotaciones de cada escena.

El módulo de las proyecciones permite identificar una escena contenida en la base independientemente de su orientación. En tareas de navegación, esta información servirá para llevar a cabo la localización del robot. Una vez localizada la imagen más cercana en el mapa, se simulan artificialmente las proyecciones de las distintas rotaciones de esa imagen. Buscando la mejor correspondencia entre las proyecciones simuladas, y las de la imagen de entrada, podemos estimar la orientación

La resolución angular dependerá del número de rotaciones de cada posición que han sido incluidas en el mapa, tal y como se indica en la Ecuación 4.37.

$$Ang.\ Minimo(^\circ) = \frac{360}{N} \tag{4.37}$$

4.2.5 PCA sobre la Firma de Fourier

Tal y como se ha visto, PCA es una técnica que intrínsecamente no presenta invariancia ante rotaciones de las imágenes. En el punto anterior se ha descrito un método para introducir rotaciones artificiales en las imágenes de la base para lograr obtener información relativa a la fase en el nuevo espacio.

Sin embargo, también es posible obtener un subespacio de información invariante ante cambios de la orientación de las imágenes si aplicamos el Análisis de Componentes Principales sobre datos que ya tengan esa propiedad. Lo que se propone en este punto es aplicar la técnica PCA sobre la información visual en el espacio de la frecuencia, pues con ello se consigue comprimir información invariante a rotación, y por tanto, aunar las ventajas del análisis en frecuencias y la reducción de la dimensionalidad de datos.

El descriptor basado en frecuencias elegido es la Firma de Fourier (Capítulo 4.1.2). En la Ecuación 4.4 se puede comprobar que el módulo de la Transformada de Fourier se mantiene invariante ante cambios en la orientación de la imagen.

Para calcular la orientación, es necesario almacenar la matriz de fases de los componentes de Fourier sin ningún tipo de variación en su espacio proyección, ya que haría imposible aplicar el Teorema del Desplazamiento para hallar el desfase relativo entre dos secuencias. Por tanto, se aplicará el análisis PCA sobre la matriz de módulos únicamente.

En su aplicación a tareas de navegación robóticas, el descriptor final estará compuesto por las proyecciones de los módulos de los coeficientes de Fourier tras aplicar el análisis PCA, y por las fases de los coeficientes sin cambiar el espacio de proyección.

4.3 Histogramas de Orientación del Gradiente (HOG)

Los descriptores de Histrograma de Orientación del Gradiente, o HOG (*Histogram of Oriented Gradient*), son descriptores de características usados en visión por computador y en procesamiento de imágenes normalmente para la detección de objetos.

Esta técnica tiene en cuenta la orientación del gradiente en partes localizadas de una imagen. Es similar los de cálculo de la orientación de borde, como Canny [31], a descriptores de características invariantes de escala [68, 113, 115] y a otros relacionados con formas, pero difiere en que calcula una rejilla densa de celdas uniformemente repartidas y hace una normalización local del contraste para mejorar sus resultados.

La idea principal detrás de los descriptores de Histograma de Orientación del Gradiente es que la apariencia de un objeto local y su forma dentro de una imagen pueden ser descritas por la distribución de la intensidad de gradientes o dirección de bordes. La implementación de este descriptor puede lograrse dividiendo la imagen en pequeñas regiones conectadas, llamadas celdas, y compilando para cada celda un histograma de orientación del gradiente de los píxeles dentro de ella. La combinación de esos histogramas forma el descriptor de la imagen.

Para mejorar el descriptor, el contraste de los histogramas locales pueden ser normalizados calculando una medida de intensidad dentro de una región más grande de la imagen, llamada bloque. Usando ese valor medio, es posible normalizar todas las celdas dentro del bloque. Esta normalización mejora los resultados ante variaciones de iluminación u oscurecimientos en la escena.

Se pueden encontrar trabajos que usan los Histogramas de Orientación, como el propuesto por Freeman et al. [60]. Sin embargo, estos sólo tienen un funcionamiento aceptable cuando se combinaban con SIFT. Navneet Dalal y Bill Triggs introducen en [45] el descriptor HOG en el que se basa nuestra propuesta. En su trabajo, introducen distintas distribuciones espaciales de las celdas de división de la imagen, y diferentes métodos de normalización de los histogramas. El algoritmo presentado se centra en el reconocimiento de peatones, utilizando bloques de tamaño fijo en el que incluían pocos píxeles.

Más tarde, Zhu et al. presentan en [209] un algoritmo que consigue aumentar la efectividad y reduce el coste computacional del algoritmo presentado en [45] para el reconocimiento de personas mediante el uso, entre otras cosas, de bloques de tamaño dinámico.

A continuación se incluye la implementación del algoritmo para la creación de Histogramas de Orientación del Gradiente, la adaptación del descriptor a tareas de navegación robóticas, y por último, se presenta una modificación del descriptor original para añadir información relativa al color de la escena.

4.3.1 Implementación del Algoritmo

Todos los algoritmos basados en el cálculo de la Histograma de Orientación del Gradiente pueden dividirse en tres pasos:

1. Cálculo del Gradiente

   Primero es necesario calcular los valores del gradiente de orientación de la imagen. El método más común es aplicar la máscara de la derivada discreta centrada en el punto de una dimensión, tanto en dirección vertical como en horizontal. Concretamente, cada imagen entrante se filtra con estas máscaras:

   $$D_x = [-1 \quad 0 \quad 1] \qquad D_y = [-1 \quad 0 \quad 1]^T \tag{4.38}$$

   Aplicando la convolución de las máscaras D_x y D_y sobre la imagen $\mathcal{I}$ obtenemos las derivadas respecto al eje x ($\mathcal{I}_x$) y al eje y ($\mathcal{I}_y$) respectivamente:

   $$\mathcal{I}_x = \mathcal{I} * D_x \qquad \mathcal{I}_y = \mathcal{I} * D_y \tag{4.39}$$

   Una vez calculadas ambas derivadas, se puede obtener tanto la magnitud como la orientación del gradiente:

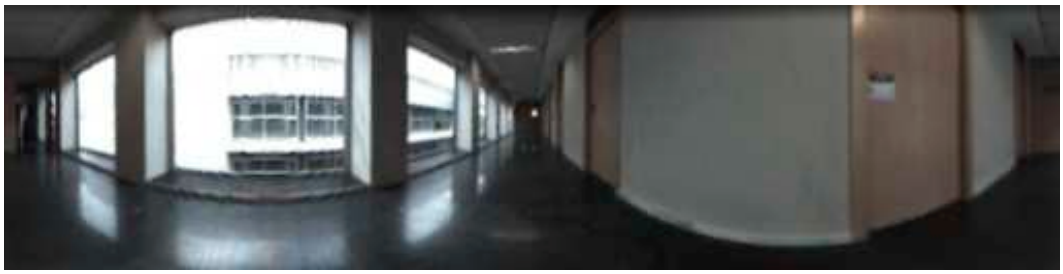
   - Magnitud del Gradiente:
     $$|G| = \sqrt{\mathcal{I}_x^2 + \mathcal{I}_y^2} \tag{4.40}$$
   - Orientación del Gradiente:
     $$\theta = arctan\frac{\mathcal{I}_x}{\mathcal{I}_y} \tag{4.41}$$

   En la Figura 4.12 se puede observar el resultado de derivar una imagen respecto al eje x, el eje y, y la magnitud de la derivada de una imagen.

2. Orientation Binning

   El segundo paso consiste en dividir la imagen en celdas, y calcular el histograma de orientación del gradiente de cada una de ellas. Las celdas utilizadas son rectangulares.

   Los histogramas pueden tener un rango entre 0^o y 180^o, o entre 0^o y 360^o, dependiendo si el gradiente tiene en cuenta el signo de la orientación o no. También debe definirse el número de divisiones (o *bins*) que contendrá el histograma. El número de *bins* determinará el rango de ángulos incluidos en cada división del histograma, y el tamaño final del descriptor. Por ejemplo, si utilizamos 8 divisiones con un rango angular de 180^o, se discretizarán los valores de la orientación cada $22{,}5^o$.



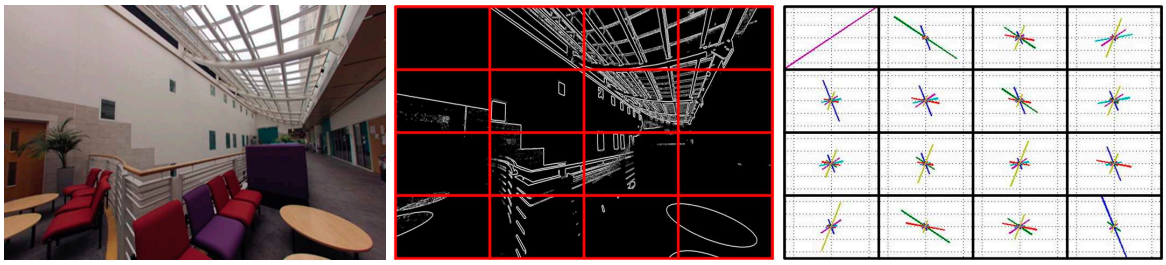
(a)

(b)

(c)

(d)

Figura 4.11: (a) Imagen original, (b) Derivada respecto al eje x, (c) Derivada respecto al eje y, y (d) Magnitud del gradiente.



(a) (b) (c)

Figura 4.12: (a) Imagen original, (b) División en celdas de la magnitud del gradiente de la imagen, y (c) representación de la dirección ponderada del gradiente de la imagen.

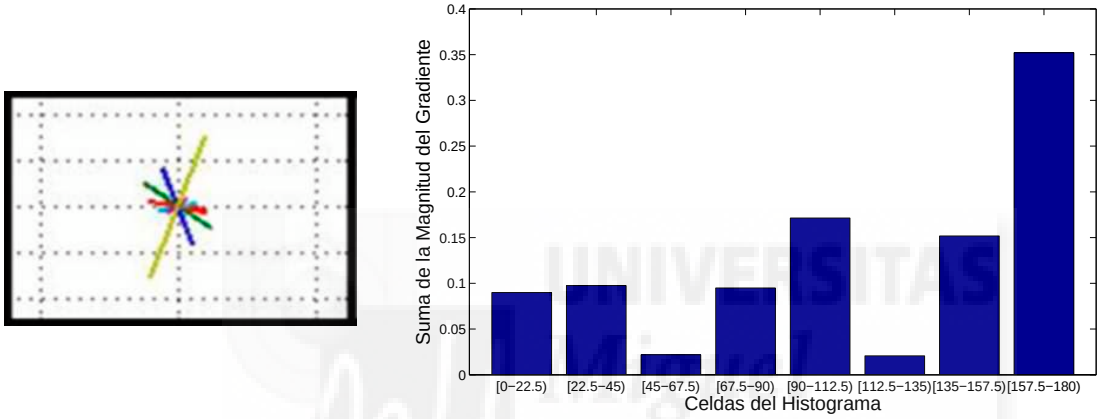


Figura 4.13: Dirección ponderada del gradiente e Histograma de Orientación asociado.

Para cada celda, se comprueba la orientación del gradiente de los píxeles incluidos para comprobar a qué división del histograma pertenece. Se suma el valor del módulo del gradiente de cada pixel a la división del histograma correspondiente. Con ello se recoge la información de la distribución espacial de la escena ponderada por el valor de su módulo.

La Figura 4.13 recoge un ejemplo de creación de histograma a partir del valor de la dirección del gradiente.

3. Creación y Normalización de Celdas

Para tener en cuenta los cambios de iluminación y contrastes, es posible normalizar la intensidad de los gradientes de forma local. Para ello, agrupamos las celdas en bloques más grandes espacialmente conectados. Esos bloques normalmente están superpuestos, lo que significa que cada celda contribuye más de una vez al descriptor final.

Existen dos geometrías principales de bloques: R-HOG, que son rectangulares, y C-HOG, que son circulares. R-HOG es el tipo de más común de bloque, y está determi-

nado por tres parámetros: el número de celdas por bloque, el número de píxeles por celda, y el número de canales en los histogramas de las celdas.

El gradiente de la imagen se ve muy afectado por cambios en la iluminación de la escena. La normalización de las celdas en bloques consigue reducir el efecto que los cambios en la iluminación tienen sobre el descriptor final de la imagen.

Existen distintos métodos para la normalización de los bloques. Nosotros utilizaremos bloques rectangulares de nueve celdas, que incluirán los 8-vecinos de la celda a normalizar. Si suponemos v un vector que contiene los 9 histogramas de un bloque, el valor normalizado del histograma de la celda central h se obtiene mediante la ecuación:

$$\overline{h} = \frac{h}{||v|| + e} \tag{4.42}$$

siendo $||v||$ la norma del vector que contiene todos los histogramas del bloque, y e una constante pequeña para evitar una posible división por cero, cuyo valor no afectará en los resultados.

4.3.2 Aplicación a tareas de Navegación

Se puede utilizar la técnica basada en HOG para obtener descriptores útiles en aplicaciones de navegación robótica. En concreto, vamos a presentar la adaptación del descriptor a imágenes panorámicas.

Por filas, todas las imágenes tomadas en un mismo punto tienen los mismos niveles de gris cualquiera que sea la orientación de la cámara al adquirir la imagen, y por lo tanto, el gradiente de la imagen por filas también conservará su valor independientemente de cualquier rotación de la escena. Por lo tanto, si calculamos el Histograma de Orientación del Gradiente sobre celdas con el mismo ancho de la imagen panorámica, el descriptor obtenido será invariante a rotaciones, permitiendo la localización del robot con independencia de su orientación.

El número de celdas empleado para crear el descriptor de la imagen será determinado experimentalmente. Su altura dependerá directamente del número de celdas horizontales empleado, ya que no existirá superposición entre celdas. En la Figura 4.14 se incluye un ejemplo de obtención del descriptor para localización sobre una imagen panorámica

Sin embargo, la aplicación de ventanas horizontales a la imagen para obtener un descriptor de la escena no proporciona información suficiente para estimar el desfase entre dos escenas capturadas en un mismo punto. Es necesario obtener información adicional. Para ello, se emplean celdas con la misma altura que la imagen separadas una cierta distancia D. Los histogramas de estas celdas verticales forman el descriptor relativo a la fase.

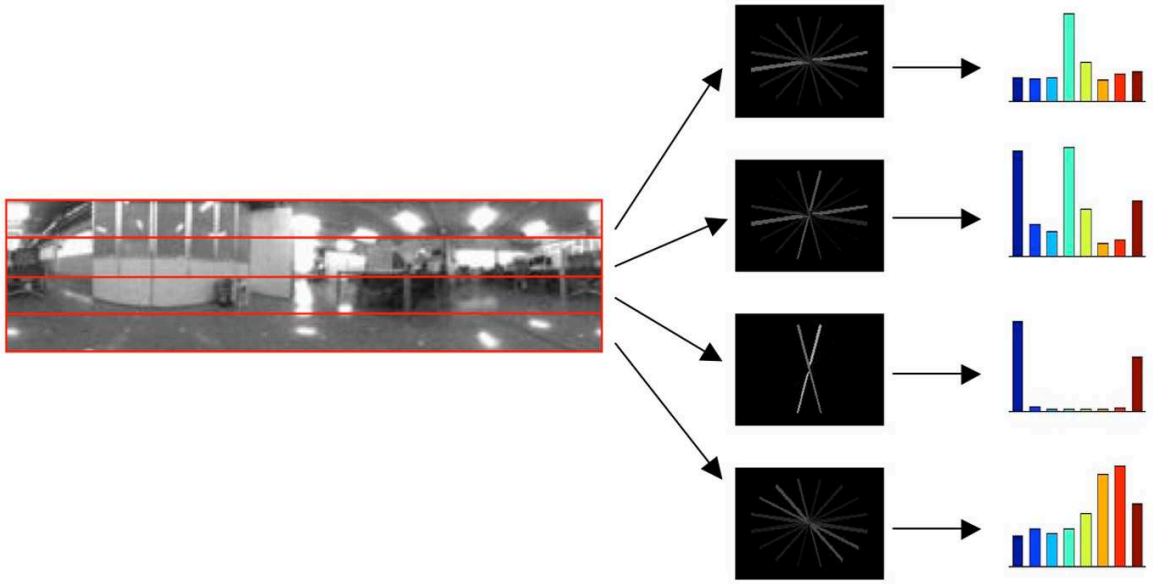


Figura 4.14: Descriptor HOG para localización sobre imagen panorámica.

Dadas dos imágenes capturadas en un mismo punto en el plano del suelo, primero se calcula el descriptor de fase de ambas imágenes. Posteriormente, se rota circularmente el orden de los histogramas del descriptor de una de las imágenes, y se vuelve a calcular la distancia entre los vectores obtenidos. Se realiza este proceso hasta haber realizado todas las rotaciones posibles.

Conceptualmente, rotar el orden de los histogramas equivale a obtener el descriptor de fase de la misma imagen desfasada D píxeles en el plano del suelo. En la Figura 4.15 se recoge la estimación del descriptor para el cálculo de la orientación y el proceso de rotación de los histogramas.

En este caso, la anchura de la celda sobre la que se calcula el histograma no tiene por qué coincidir con la distancia entre celdas contiguas, pudiendo existir superposición entre celdas. El ángulo mínimo de desfase que seremos capaces de detectar entre dos imágenes está discretizado, y es función de la distancia entre celdas verticales consecutivas:

$$\text{Desfase Mínimo}(^{o}) = \frac{D \cdot 360}{\text{Columnas Imagen}} \tag{4.43}$$

Tras realizar todas las comparaciones entre las rotaciones de los descriptores de fase de una imagen, y el descriptor de la segunda imagen, buscamos la rotación que presenta la mínima distancia. Siendo r la iteración en la rotación circular del orden de los histogramas en la que se encuentra la mínima distancia entre descriptores, el desfase entre imágenes se calcula como:

$$\text{Desfase entre imágenes}(^{o}) = r \cdot \frac{D \cdot 360}{\text{Columnas Imagen}} \tag{4.44}$$

Por tanto, la base estará compuesta por dos descriptores distintos. El primero es el formado por los histogramas de las ventanas horizontales, que será utilizado para llevar a cabo la localización. Una vez se ha asociado la imagen de entrada con la de la base, se emplean los descriptores formados por las ventanas verticales de ambas imágenes para calcular el desfase relativo entre imágenes, y con ello, estimar la orientación.

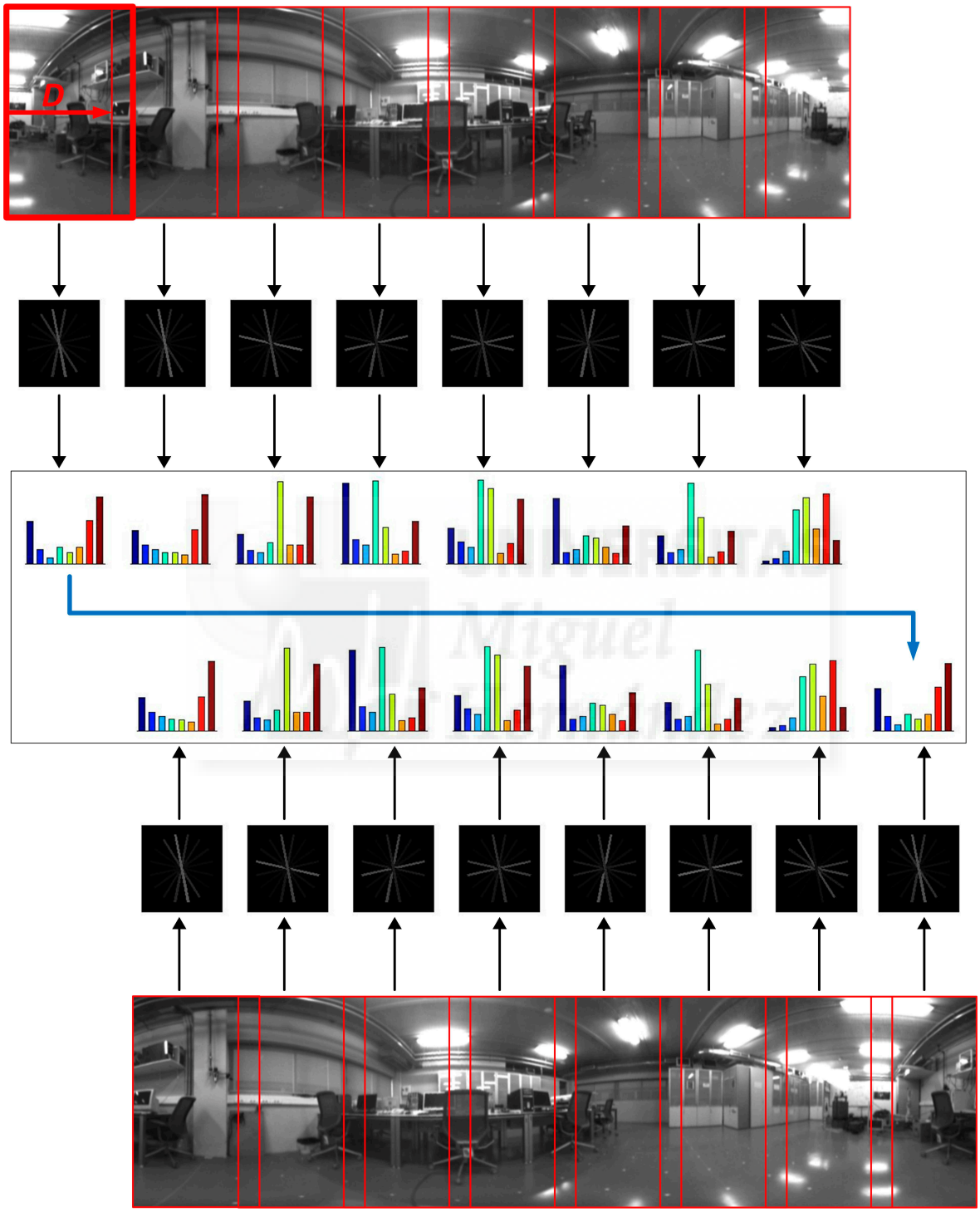


Figura 4.15: Descriptor HOG para estimación de la orientación sobre imagen panorámica, con ejemplo de rotación circular de los histogramas para el cálculo del desfase entre imágenes.

4.4 GIST

Otro concepto para la compresión de la información visual es el de GIST. Se puede definir como una representación abstracta de la escena que activa la memoria de representaciones de categorías de escenas (como por ejemplo montañas o ciudades). En [61], Friedman presenta el primer trabajo que tengamos conocimiento en el que se emplea este concepto.

Aude Oliva y Antonio Torralba desarrollan esta idea bajo el nombre *holistic representation of the spacial envelope* en [147] para conseguir crear un descriptor. También aparece en otros trabajos como [145, 148]. Los descriptores GIST, como su nombre indica, tratan de extraer de la imagen *lo esencial*, imitando el sistema de percepción humano y su habilidad de identificar rápidamente una escena mediante la identificación de regiones con color y texturas notables.

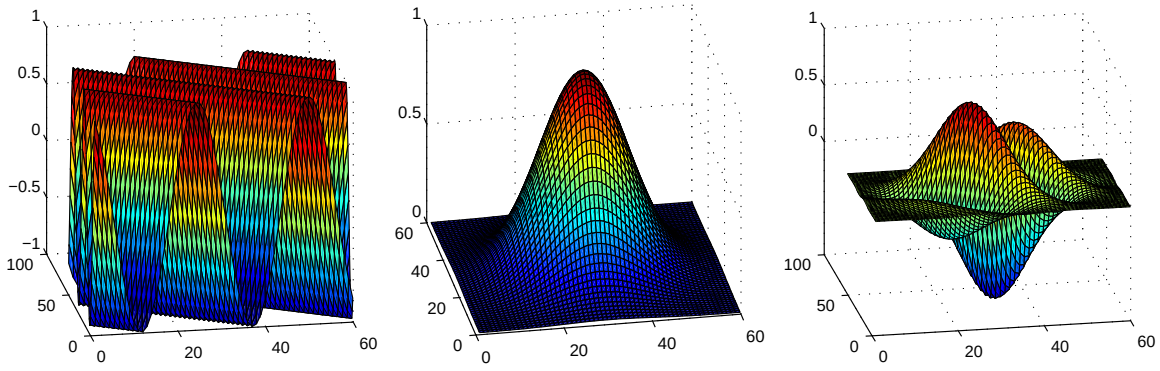
La idea de categorizar una escena a partir de estos términos proviene de estudios anteriores como [146, 149], basados en la capacidad humana de reconocimiento de información visual donde se demuestra que se puede clasificar una escena con imágenes borrosas en las que únicamente se aprecia la distribución espacial de la misma, es decir, la forma de la escena. Por lo tanto, es posible una representación de la imagen de naturaleza holística, sin necesidad de representar la forma de los objetos, sino simplemente su distribución u orientación característica.

Esta representación es lo que se conoce como el *gist* de una imagen, que incluye diversos niveles de procesamiento, desde características de bajo nivel (como puede ser el color o frecuencias espaciales), propiedades intermedias de la imagen (tales como superficies o volumen) e información de alto nivel (reconocimiento de objetos).

Las propiedades para describir la escena incluyen distintas frecuencias espaciales y escalas, color, y densidad de textura.

Bajo esta definición caben muchas interpretaciones de descriptores de la imagen, pues no hay una única manera de tratar de extraer *la esencia* de la imagen. Por ello, con el nombre de GIST se pueden desarrollar diversos descriptores. Todos ellos tienen en común que obtienen características de la escena trabajando con el conjunto global de la imagen, pero no lo hacen de la misma forma: algunos emplean únicamente filtros para extraer la orientación de los elementos de la escena, mientras que otros añaden la información que proporciona trabajar con diversas escalas de la imagen. También existen autores que aprovechan los distintos canales de color de la imagen para aumentar las características al definir un descriptor.

En este proyecto se van a estudiar dos métodos distintos de extracción de características basados en GIST:



(a) (b) (c)

Figura 4.16: (a) Sinusoide compleja, (b) Envoltura Gaussiana y (c) Filtro de Gabor resultante de la convolución de ambas funciones.

- El primero se denominará GIST-Gabor, pues centra la extracción de características en la aplicación de filtros de Gabor a la imagen.

- El segundo método aparece bajo el nombre de GIST-Color ya que, además de aplicar filtros de Gabor, extrae distintos canales de color de la imagen para añadir características al descriptor.

Como se puede ver, el punto en común de ambos es la utilización de filtros de Gabor para obtener información relacionada con la distribución espacial de las escenas. A continuación se describe la composición y funcionamiento de los filtros de Gabor, para en apartados posteriores detallar la formación de los dos descriptores basados en técnicas GIST.

4.4.1 Filtros de Gabor

El filtro de Gabor es un filtro lineal cuya respuesta de impulso es una función sinusoidal multiplicada por una función gaussiana, siendo casi un filtro paso banda.

La principal ventaja que se obtiene al introducir la envolvente gaussiana es que las funciones de Gabor están localizadas tanto en el dominio espacial como en el de la frecuencia.

La fórmula de una función compleja de Gabor en el dominio del espacio puede expresarse como:

$$g(x,y) = s(x,y) * \omega(x,y), \tag{4.45}$$

donde $s(x,y)$ es una sinusoide compleja, conocida como la portadora, y $\omega(x,y)$ una función gaussiana bidimensional, conocida como envoltura.

- La sinusoide compleja

  La función de la sinusoide compleja en coordenadas cartesianas se recoge en la siguiente ecuación:

$$s(x,y) = \exp(j(2\pi(u_0x + v_0y) + P)) \tag{4.46}$$

  donde (u_0, v_0) y P definen la frecuencia espacial, expresada como dos funciones reales separadas localizadas en la parte real e imaginaria del espacio complejo. Dichas partes pueden ser expresadas como:

$$Re[s(x,y)] = \cos(\exp(2\pi(u_0x + v_0y) + P) \tag{4.47}$$

$$Im[s(x,y)] = \sin(\exp(2\pi(u_0x + v_0y) + P). \tag{4.48}$$

  La frecuencia espacial de la sinusoide (u_0, v_0) también puede ser expresada en coordenadas polares a través de su magnitud (F_0) y dirección (ω_0):

$$F_0 = \sqrt{u_0^2 + v_0^2} \tag{4.49}$$

$$\omega_0 = \tan^{-1}\left(\frac{v_0}{u_0}\right). \tag{4.50}$$

  En coordenadas polares, las frecuencias espaciales se expresarían como:

$$u_0 = F_0 \cos\omega_0 \tag{4.51}$$

$$v_0 = F_0 \sin\omega_0. \tag{4.52}$$

  Usando esta representación, la sinusoide compleja se define como:

$$s(x,y) = \exp(j(2\pi F_0(x\cos\omega_0 + y\sin\omega_0) + P)). \tag{4.53}$$

- La Envoltura Gaussiana

  Matemáticamente, la envoltura Gaussiana del filtro de Gabor se expresa como:

$$\omega_r(x,y) = K\exp(-\pi(a^2(x-x_0)_r^2 + b^2(y-y_0)_r^2)) \tag{4.54}$$

  con (x_0, y_0) las coordenadas del máximo de la función, y a y b parámetros de escalamiento de la Gaussiana en cada eje. El subíndice r denota una operación de rotación de la envoltura:

$$(x-x_0)_r = (x-x_0)\cos\theta + (y-y_0)\sin\theta \tag{4.55}$$

$$(y-y_0)_r = -(x-x_0)\sin\theta + (y-y_0)\cos\theta \tag{4.56}$$

Conforme a y b aumentan de valor, disminuye la envoltura en el dominio espacial en el eje respectivo. La rotación definida en las ecuaciones 4.55 y 4.56 tiene sentido dextrógiro con respecto a θ. En la Figura 4.17 se incluye la representación gráfica de dos envolturas Gaussianas.

Por lo tanto, la función compleja del Filtro de Gabor, definida como convolución de la sinusoide y la envoltura Gaussiana, depende de los siguientes 9 parámetros:

- K Escala de la magnitud de la envoltura Gaussiana
- (a,b) Escala de la envoltura con respecto los ejes x e y
- θ Ángulo de rotación de la envoltura Gaussiana
- (x_0,y_0) Posición del máximo de la envoltura
- (x_0,y_0) Frecuencias espaciales de la sinusoide portadora
- P Fase de la sinusoide

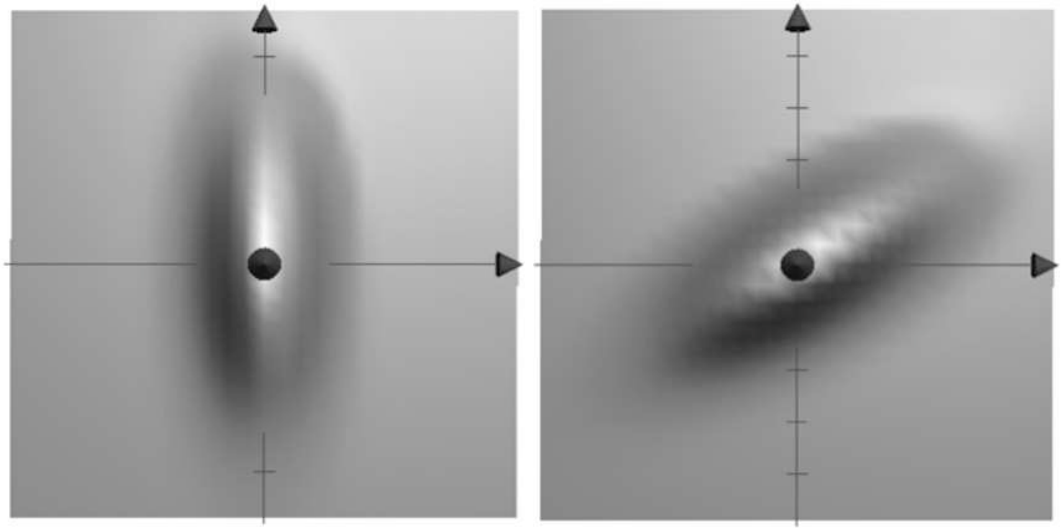
Aprovechando las propiedades de los filtros de Gabor con respecto al tratamiento de texturas en imágenes digitales, pueden ser utilizados en procesos de segmentación y/o compresión.

Una imagen puede considerarse como un mosaico de regiones con distintas texturas. Los filtros de Gabor permiten extraer un patrón de características a partir de las texturas. Por lo tanto, al aplicar los filtros de Gabor sobre una imagen se obtiene un conjunto de cualidades con los que identificar y clasificar dicha imagen. Aunque se pueden usar distintos métodos para la extracción de características a partir de la textura de la imagen, los filtros de Gabor pueden destacarse por ser óptimos a la hora de hacer interactuar las dimensiones de espacio y frecuencia. Además, por sus características pueden convertirse en detectores de líneas con dirección y escala seleccionables.

También se puede trabajar en el dominio frecuencial realizando la transformada de Fourier del filtro de Gabor. En el espacio de la frecuencia, una máscara de Gabor se corresponde con una función Gaussiana centrada en la frecuencia de la función sinusoidal. La ecuación del filtro de Gabor en el dominio de la frecuencia se puede expresar como:

$$\begin{aligned}\widehat{g}(u,v) = \frac{k}{ab}\exp(j(-2\pi(x_0(u-u_0)+y_0(v-v_0))+P))\cdot \\ \cdot\exp\left(-\pi\left(\frac{(u-u_0)_r^2}{a^2}+\frac{(v-v_0)_r^2}{b^2}\right)\right).\end{aligned} \tag{4.57}$$

En la Figura 4.19 se representa gráficamente una máscara de Gabor en el dominio de la frecuencia.



(a) (b)

Figura 4.17: Envolturas espaciales con $x_0 = y_0 = 0$, $a = 1/2$, $b = 1/4$ con ángulo (a) $\theta = 0^o$ y (b) $\theta = 45^o$

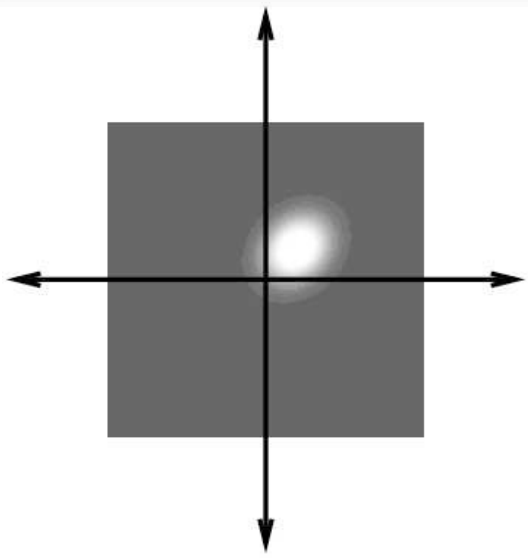


Figura 4.18: Representación bidimensional de un filtro de Gabor en el espacio de la frecuencia.

En [160], sus autores presentan una aplicación para la visualización de los filtros de Gabor y el efecto que tiene sobre las imágenes la variación de los distintos parámetros que forman parte de su ecuación.

4.4.2 Gist-Gabor

El descriptor que se desarrolla en este punto está dentro de la definición de GIST que propone Oliva y Torralba [147] para extraer un conjunto de características de baja dimensión de la imagen que permita su reconocimiento. En esta definición aparece el término *Espatial envelope* para designar una representación de las imágenes con muy pocos términos, consiguiendo concentrar información sobre la forma de las escenas de forma holística. Hacen referencia a conceptos como la rugosidad, la naturalidad, o la aspereza de la imagen, evitando información específica sobre objetos concretos.

Por tanto, el propósito con el que nace este concepto es el de crear un modelo computacional para el reconocimiento de escenas de forma que se usen características basadas en su apariencia global. Es necesario un procesamiento previo de la imagen que permita la extracción de las características a partir de las cuales clasificar la escena. Nuestro descriptor utiliza ese conjunto de características previas para describir la escena.

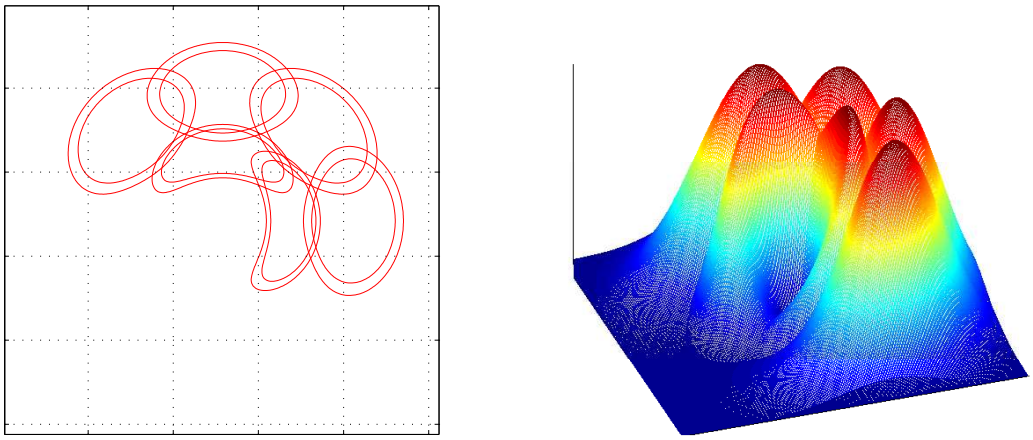
El filtrado de la imagen se realiza mediante una representación multiresolución de las máscaras de Gabor, es decir, variando las escalas de las máscaras además de su orientación.

El cálculo del descriptor de una imagen basado en GIST-Gabor se realiza en tres pasos principales:

1. Banco de Filtros de Gabor

   El primer paso consiste en crear el conjunto de filtros que se aplicarán a cada imagen. Se podrán seleccionar tanto el número de escalas como el número de orientaciones. Los filtros estarán repartidos equiangularmente por escala entre 0 y 180^o, por lo que definiendo el número de filtros por escala, quedarán determinadas sus orientaciones.

   La aplicación de los filtros se realizará en el dominio de la frecuencia, por lo que los filtros se calculan en el dominio de Fourier. La Figura 4.19 muestra gráficamente un banco de filtros de Gabor con dos escalas, la primera con 2 orientaciones, y la segunda con 4. La primera gráfica muestra el contorno de la máscara, mientras que la segunda es una representación tridimensional de la misma. Ambas están en el dominio de la frecuencia.



(a) (b)

Figura 4.19: Máscaras de Gabor en espacio frecuencia con distintas escalas espaciales y orientaciones. (a) Representación de los límites de las máscaras en 2D, y (b) representación de los valores de las máscaras en 3D.

2. Aplicación de las máscaras de Gabor a las Imágenes

   Tal y como se ha comentado anteriormente, el banco de Gabor está en el dominio de la frecuencia. Por ello, habrá que hacer la transformada de Fourier de la imagen para que esté en el mismo dominio. En concreto, se calcula la transformada bidimensional de Fourier.

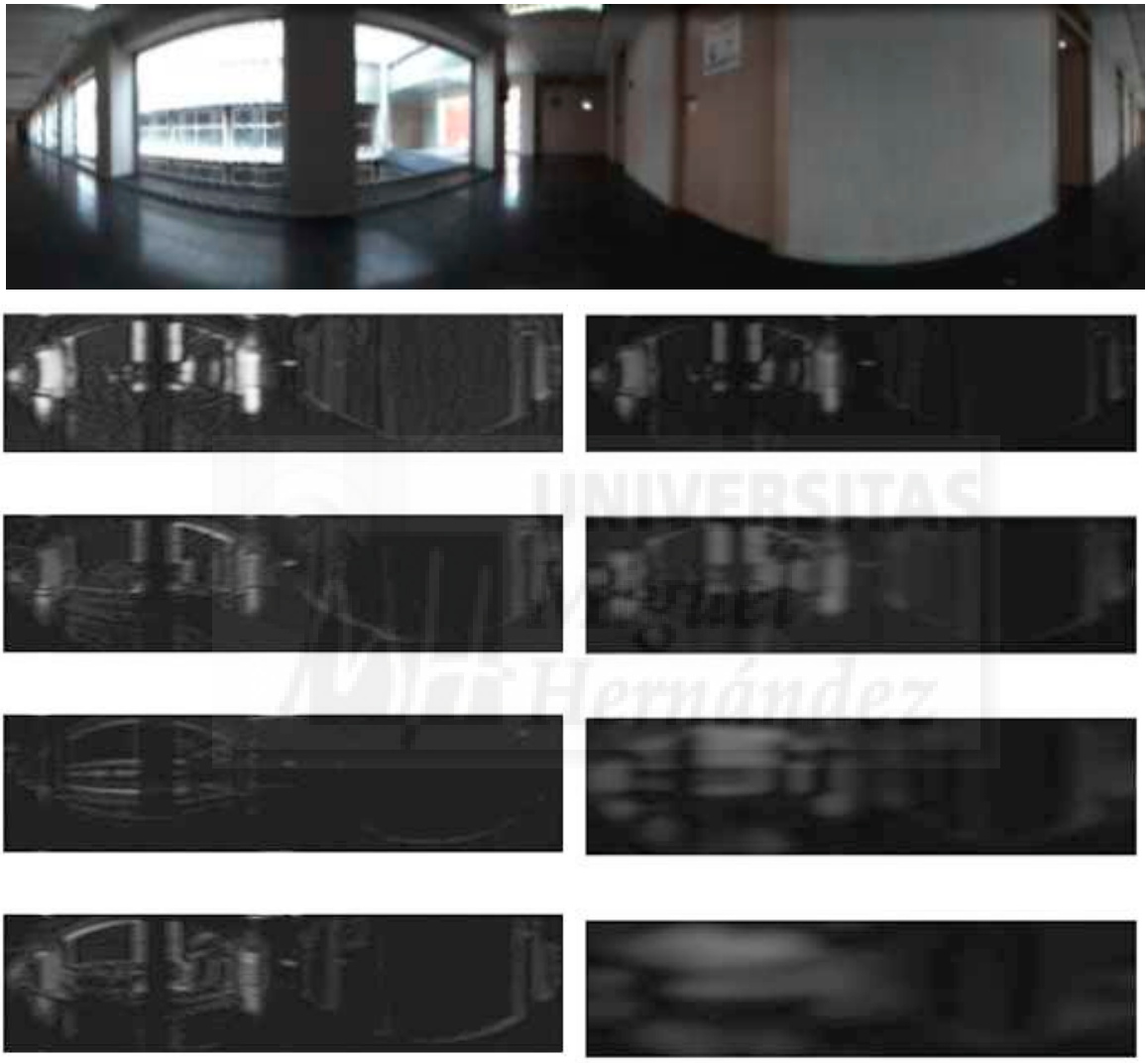
   Luego se realiza la convolución de las máscaras con la imagen. Hay que recordar que la transformada de Fourier de una convolución es el producto punto a punto de las transformadas. En otras palabras, la convolución en un dominio (por ejemplo el dominio espacial) es equivalente al producto punto a punto en el dominio espectral. Por lo tanto, al estar en el dominio de la frecuencia, esta operación se realiza mediante la multiplicación de ambas matrices elemento a elemento.

   Posteriormente se calcula la antitransformada del resultado, y se almacena su valor absoluto.

   La Figura 4.20 recoge un ejemplo del filtrado de una misma imagen por máscaras de Gabor usando distintas orientaciones y escalas.

3. Obtención del Descriptor

   El descriptor que se propone en este punto se aplicará sobre imágenes panorámicas. Una vez obtenidas todas las imágenes filtradas por los distintas máscaras, éstas se



(a) (b)

Figura 4.20: Filtrado de Gabor de una imagen con (a) diferentes orientaciones (0°, 45° , 90°, 135°) y (b) distintas escalas espaciales.

dividen en ventanas horizontales (con el mismo ancho de la imagen), y se calcula la media de los píxeles contenidos en cada ventana para obtener la información relativa a la localización.

Recordemos que, por filas, una imagen panorámica contiene los mismos píxeles cualquiera que sea su orientación. De las misma forma, la imagen filtrada contendrá la misma información por filas independientemente de la orientación de la escena. Por tanto, el nivel medio de brillo de los píxeles contenidos en las ventanas horizontales es invariante a rotación.

El descriptor final estará formado por los niveles medios de gris de las ventanas en las que se dividen las imágenes al aplicar los diferentes filtros de Gabor recogidos en un vector.

En la Figura 4.21(a) se muestra los componentes del descriptor de localización obtenidas a partir del filtrado de la imagen original por una máscara de Gabor (en este caso, con dirección de filtrado igual a 0^o). Esta operación se repite para los resultados obtenidos del filtrado de la imagen con todas las máscaras de Gabor, es decir, para todas las escalas y direcciones. El número de divisiones de la imagen, al igual que el número de máscaras a aplicar, es un parámetro a estudiar en la parte experimental. No existirá solapamiento entre las ventanas horizontales, por lo que el número de ventanas que se aplican sobre la imagen condicionará su altura.

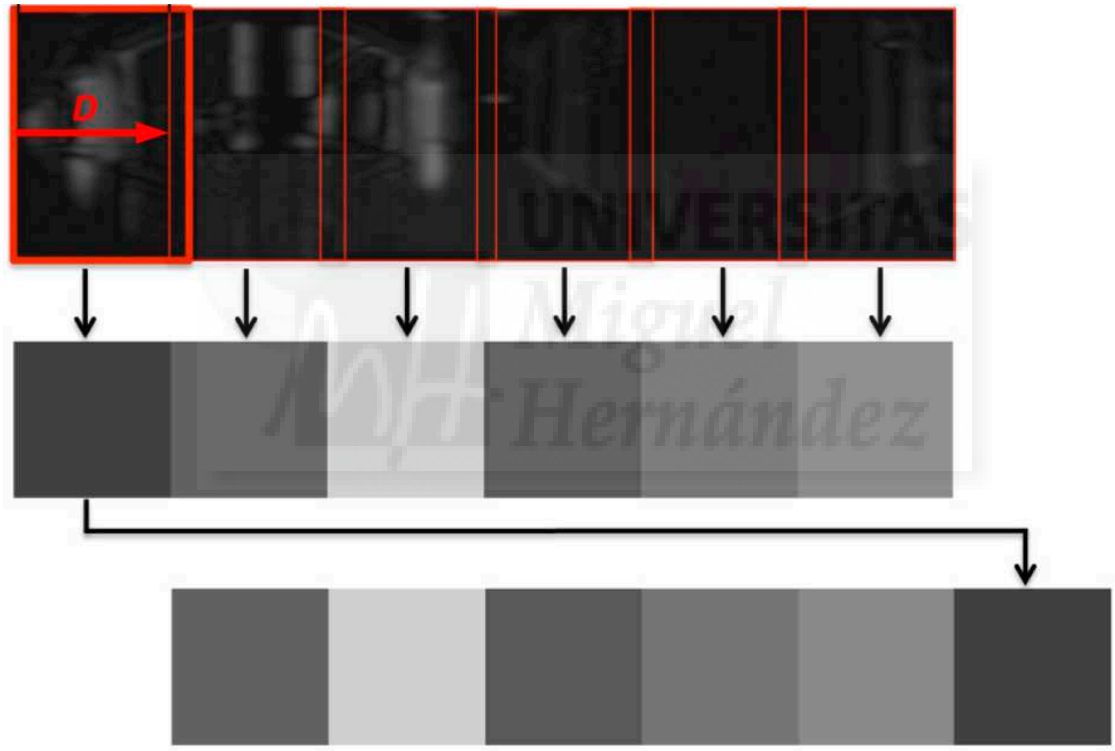
Tal y como ocurre con el descriptor basado en HOG (Sección 4.3.2), la información proporcionada por las ventanas horizontales permite llevar a cabo la localización, pero no estimar la orientación en tareas de navegación. Por ello, creamos un segundo descriptor de la imagen que usará ventanas verticales (con la misma altura de la imagen). Dichas ventanas estarán aplicadas cada D píxeles, siendo posible el solapamiento entre ventanas consecutivas. El descriptor de fase estará formado por el valor medio de las ventanas verticales.

El desfase relativo entre dos imágenes se estima mediante la comparación y rotación del orden de los valores del vector que forma el descriptor de orientación. En la Figura 4.21(b) se muestra la obtención de las características del descriptor a partir de las ventanas verticales, y la rotación del orden de los elementos.

Dados los descriptores de orientación de dos imágenes, se calcula la distancia entre ambos vectores (Distancia Euclídea), y se rota circularmente el orden de uno de ellos. Se repite este proceso hasta realizar todas las rotaciones posibles. La iteración en la que se encuentre la mínima distancia (r), indicará el desfase relativo entre imágenes.

El ángulo mínimo que seremos capaces de detectar entre dos imágenes vendrá determinado por la distancia entre ventanas verticales D (Ecuación 4.43), mientras que el desfase relativo entre dos escenas se obtiene mediante la Ecuación 4.44.

(a)



(b)

Figura 4.21: Obtención del descriptor GIST-Gabor a partir de una imagen filtrada por una máscara de Gabor.

Primero se utilizará el descriptor de localización para posicionar el móvil en el mapa mediante una asociación con el mapa. Tras ello, se utilizan los descriptores de orientación de la imagen entrante y la imagen del mapa seleccionada para hallar el desfase entre imágenes, y con ello, estimar la orientación durante la navegación del robot.

4.4.3 **Gist-Color**

En este apartado se presenta un descriptor de baja dimensión de una escena que recoge información relativa al color y textura de la misma. Por su carácter holístico, está bajo la definición de un descriptor GIST.

En [88], Itti et al. desarrollan un sistema de extracción de regiones destacadas de una escena a partir de un conjunto de características visuales como la intensidad, los canales de color o la dirección de los elementos de la escena mediante el filtrado de la escena con máscaras de Gabor. En [173], obtienen un algoritmo de clasificación de imágenes a partir de las mismas características. Siagian e Itti [174] obtienen un sistema de localización robótica basado en las características Gist, junto con una extracción de regiones destacadas en la escena para refinar la localización. En [33], este sistema se aplica en tareas de navegación en entornos reales.

Lo que se pretende con los llamados esquemas inspirados en características biológicas es clasificar las escenas imitando al proceso que tiene lugar en el córtex visual humano en las tareas de reconocimiento. Diversos estudios han demostrado que ciertas características biológicas se pueden utilizar en tareas de reconocimiento visual [84].

El modelo que aquí se propone combina las mismas características de color, intensidad y orientación de las escenas (extraídas con los filtros de Gabor), pero sin realizar segmentación de las imágenes, construyendo descriptores que aprovechen las propiedades de las imágenes panorámicas.

Como se ha comentado en los puntos anteriores, para extraer el *Gist* de una imagen se van a utilizar características holísticas.

Teniendo una imagen de entrada, primero la filtramos para obtener ciertos canales de características de bajo nivel en múltiples escalas espaciales.

Las distintas escalas espaciales se obtienen a partir de una pirámide Gaussiana de imágenes. Cada nivel de la pirámide se obtiene a partir de la reducción de la imagen del nivel anterior a la mitad de su tamaño, y la aplicación de un filtrado con una máscara Gaussiana para suavizarla.

En la Figura 4.22 es posible ver una Pirámide Gaussana formada por 8 escalas.

A continuación se detallan cada una de las características extraídas de la imagen original, y la construcción del descriptor.

1. Características de Color

   Los canales de color e intensidad se combinan para obtener tres pares de colores opuestos (tomados de la teoría de oposición de colores de Ewald Hering [80]).

Figura 4.22: Pirámide Gaussiana de una imagen formada por 8 escalas

Dicha teoría propone cuatro colores primarios: rojo (R), verde (V) , azul (B), y amarillo (Y). A partir de estos canales, se calcularán los pares de colores opuestos. Incluye también el canal de intensidad de la imagen. Los colores primarios se obtienen de la salida *rgb* de la imagen como:

$$R = r - \frac{(g+b)}{2} \tag{4.58}$$

$$G = g - \frac{(r+b)}{2} \tag{4.59}$$

$$B = b - \frac{(r+g)}{2} \tag{4.60}$$

$$Y = r + g - 2 \cdot (|r-g| + b) \tag{4.61}$$

$$I = \frac{(r+g+b)}{3} \tag{4.62}$$

Siguiendo con el modelo de características biológicas, se toma el funcionamiento de la retina para obtener los pares de colores opuestos. La retina usa un sistema de procesamiento de los colores basado en opuestos. Dicha teoría propone que la luz se separa en tres canales: rojo-verde (RG), azul-amarillo(BY) y brillo(I).

El canal de intensidad se obtiene directamente de la Ecuación 4.62. Los otros dos (RG y BY), se pueden hallar a partir de los obtenidos con las ecuaciones 5.1, 4.59, 5.3 y 4.61.

$$RG = |R - G| \tag{4.63}$$

$$BY = |B - Y| \tag{4.64}$$

Las características relacionadas con el color se obtienen utilizando operaciones *center-surround* a los tres canales. Estas operaciones emplean la comparación de los valores de una imagen en un punto con los de su alrededor. En la práctica, dicha comparación se realiza usando las distintas escalas de las imágenes.

En nuestro caso, el "centro" lo formarán las escalas de imágenes más nítidas (escalas más altas, denotadas con la letra *c*), mientras que los píxeles envolventes serán las escalas más pequeñas (*s*), que tienen menor resolución. La comparación entre escalas será denotado con el operador $\ominus$.

La ventaja es que, mediante las operaciones *center-surround*, se aporta información en distintas escalas y se añade invariancia a cambios de iluminación.

Para los dos canales de colores opuestos (*RG* y *BY*) y el de nivel de gris se construyen diferentes combinaciones de operaciones *center-surround*. Si se denotan los niveles

Tabla 4.1: Escalas de la pirámide Gaussiana de la imagen comparadas en las operaciones *center-surround*

c	s
2	3
2	4
2	5
2	6
3	4
3	6

de mayor resolución como s, y a los de menor como c, la comparación entre colores opuestos queda como:

$$RG(c,s) = |(R(c) - G(c)) \ominus (R(s) - G(s))| \tag{4.65}$$

$$BY(c,s) = |(B(c) - Y(c)) \ominus (B(s) - Y(s))| \tag{4.66}$$

$$I(c,s) = |I(c) \ominus I(s)| \tag{4.67}$$

Para realizar estas comparaciones, se reescala la imagen de menor tamaño (s) para que tenga las mismas dimensiones que la de mayor (c). De esa forma, aunque tendrá el mismo tamaño, no tendrá la misma resolución.

Considerando las escalas de la pirámide Gaussiana en orden inverso a su resolución (la escala 1 corresponde con la imagen original, es decir, la de mayor resolución), los pares $[c,s]$ utilizados en las operaciones *center-surround* son las que aparecen en la Tabla 4.1.

Cabe destacar que es posible que la resolución de la imagen no permita llegar hasta una pirámide de 6 niveles. En ese caso, el número de comparaciones *center-surround* que se realiza en la imagen estará condicionado por la escala máxima de la pirámide Gaussiana.

2. Características de Orientación

   Las características de color obtenidas hasta ahora son independientes de la distribución espacial de la escena, recogiendo únicamente información por celdas del color y la intensidad de la imagen.

   La información sobre la orientación se incorpora empleando filtros de Gabor aplicados a la imagen de entrada en niveles de gris. Se utilizarán cuatro orientaciones de filtrado:

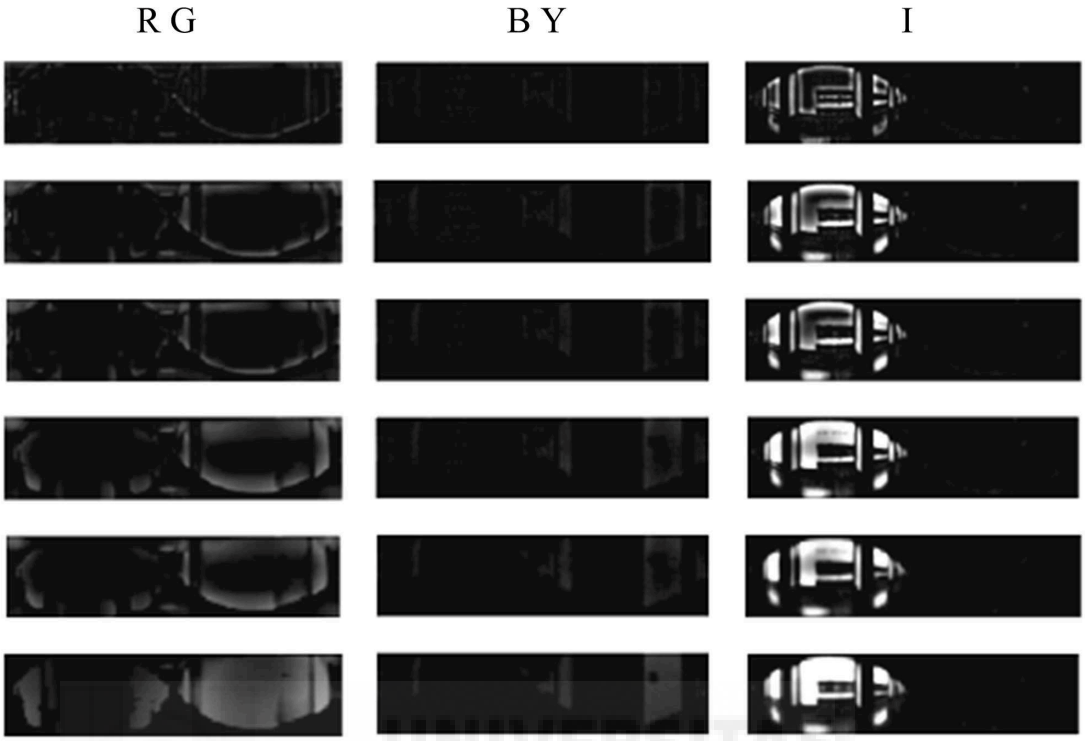


Figura 4.23: Resultado de las operaciones de comparación *center-surround* para distintas escalas sobre los canales RG, BY e I de una imagen.

$\theta_i = 0^o, 45^o, 90^o, 135^o$. Este filtrado se aplicará a las diferentes escalas de la pirámide de imágenes en el dominio espacial (no en el de la frecuencia).

3. Obtención del descriptor

   Al igual que con GIST-Gabor visto en la Sección 4.4.2, la información de los distintos canales se recoge en un descriptor mediante el cálculo del valor medio de los píxeles de celdas aplicadas a cada una de las escenas obtenidas.

   De nuevo, como el descriptor está destinado a la caracterización de imágenes panorámicas, la división de la imagen se hará de forma que en cada una tome todo el ancho de la imagen, obteniendo de esa forma invariancia ante rotaciones de la escena. Este descriptor será utilizado para la localización.

   En cuanto a la orientación, se usan celdas verticales (con la misma altura de la imagen), que se aplicarán cada *D* píxeles. Tanto la anchura de la imagen como la distancia de aplicación entre ventanas consecutivas son parámetros que se determinarán en la parte experimental, pudiendo existir superposición entre ventanas.

   La rotación entre dos imágenes se puede estimar mediante comparación y rotación circular de los descriptores de orientación de dichas imágenes.

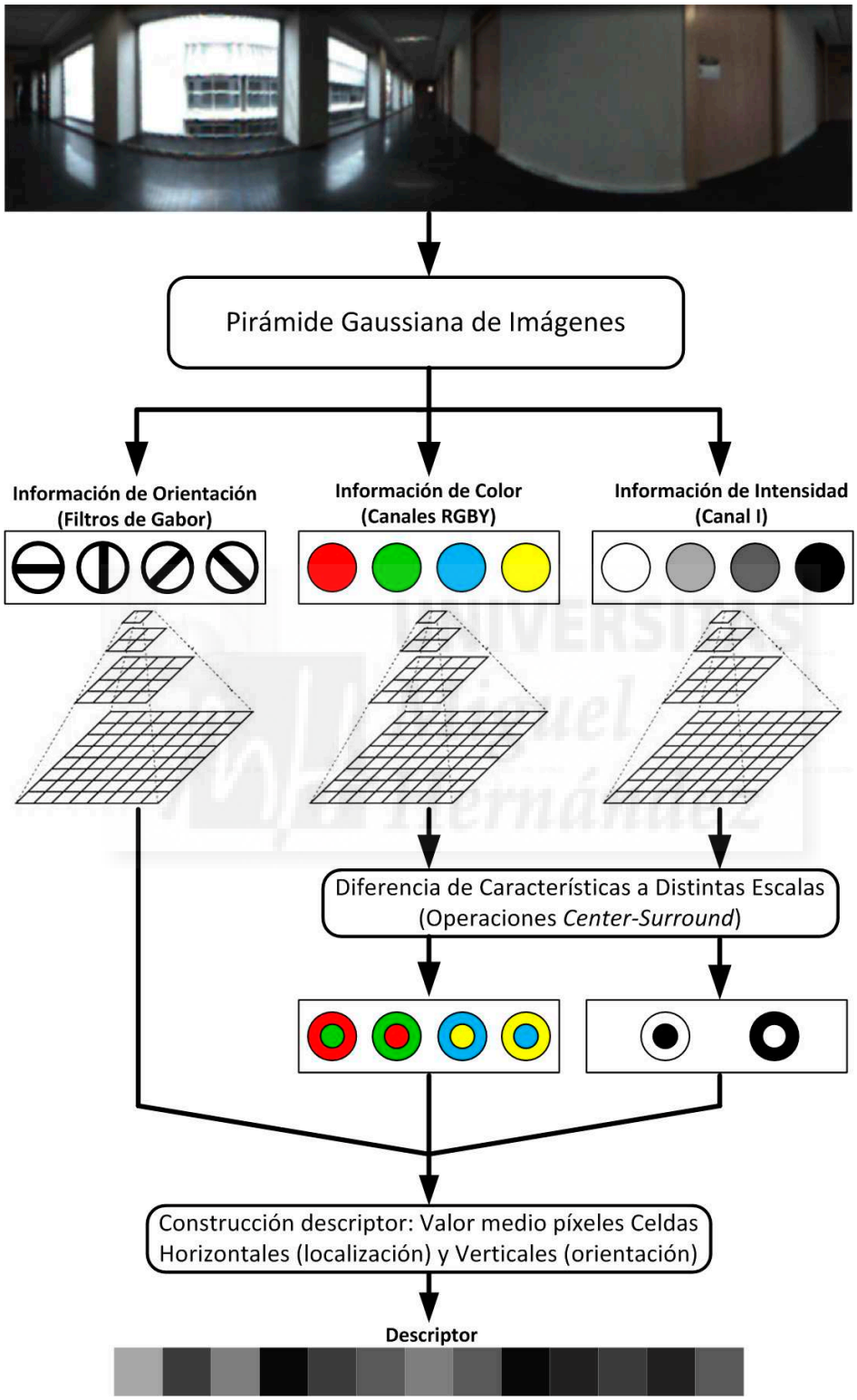


Figura 4.24: Esquema de características y operaciones para la obtención de GIST-Color.

La resolución angular está muestreada, con un paso que depende de la anchura de la imagen y D (Ecuación 4.43). El desfase entre imágenes dependerá de la iteración en la rotación circular del descriptor r en la que se encuentra la mínima distancia entre los dos descriptores de orientación (Ecuación 4.44).

El descriptor relativo a la orientación se formará únicamente con las escenas filtradas por las máscaras de Gabor, reduciendo de esta manera el tamaño del descriptor, y por tanto, los requerimientos de memoria. A pesar de ello, los resultados en el cálculo de la fase no se ven afectados, pues únicamente se está buscando la correlación entre dos imágenes, por lo que no se necesita una caracterización tan profunda de las imágenes.

El proceso de obtención del descriptor a partir de las distintas características aparece en la Figura 4.21.

En la Figura 4.24 se recoge el esquema de todo el proceso seguido para obtener el descriptor GIST-Color a partir de una imagen de entrada.

Por lo tanto, el descriptor GIST-Color recoge tanto información relativa al color de las escenas como de la distribución de los elementos en la misma a distintas escalas espaciales de una forma muy compacta.

UNIVERSITAS
Miguel
Hernández

CAPÍTULO

5

Análisis Comparativo de Técnicas de Apariencia Global sobre Escenas Panorámicas en Color.

En este capítulo se presenta una comparación de descriptores basados en apariencia global sobre imágenes panorámicas. En concreto, el estudio se centra en la precisión de estimación de la pose y requisitos computacionales de cada método. Para ello, se va a profundizar en la selección de parámetros de cada técnica, buscando un compromiso entre la precisión obtenida y el tiempo y memoria requeridos.

La comparación se realiza mediante la construcción de un mapa denso de imágenes y posterior localización de escenas de test dentro del mapa creado.

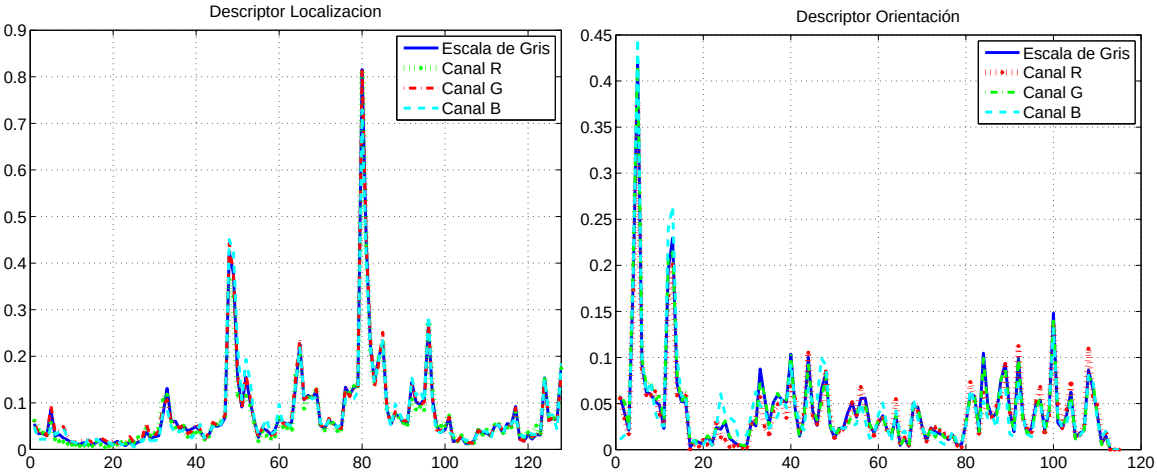
La base de imágenes utilizada en la parte experimental ha sido adquirida por el grupo de investigación durante el desarrollo de la tesis en un entorno real. Para ello, se ha empleado un sistema catadióptrico que adquiere imágenes omnidireccionales en color. La segunda sección del presente capítulo recoge las características de dicha base, así como ejemplos de la misma.

Como aportación, el estudio incluye la utilización de la información de color de las escenas para complementar los descriptores. En concreto, se analizará el empleo de distintos espacios de color e histogramas de los distintos canales de información de color.

Por último, se presenta una comparación del comportamiento de los descriptores al introducir ruido Gaussiano y oclusiones a las imágenes de test, que nos permitirá comprobar la robustez de las distintas técnicas ante estas situaciones.

5. ANÁLISIS COMPARATIVO DE TÉCNICAS DE APARIENCIA GLOBAL SOBRE ESCENAS PANORÁMICAS EN COLOR.

El estudio realizado en este capítulo será de utilidad en la selección del descriptor a utilizar en aplicaciones incluidas en capítulos posteriores.



(a) (b)

Figura 5.1: Descriptor HOG de una misma imagen en escala de gris, y sobre los canales R, G y B de la imagen en color para (a) Localización y (b) Orientación.

5.1 Consideración de la Información de Color.

Los descriptores incluidos en el Capítulo 4, a excepción de Color-GIST, extraen la información de la escena sobre un único canal. Por ello, en la gran mayoría de los trabajos incluidos en la bibliografía, los descriptores se aplican sobre imágenes en escala de grises.

Sin embargo, cuando la base está formada por imágenes en color, la información proporcionada por los distintos canales puede ser utilizada con el propósito de mejorar los descriptores con nuevos términos que permitan añadir nuevas características a la escena.

La idea más directa para aprovechar la información relativa al color de la escena es aplicar el mismo algoritmo para los tres canales RGB separadamente. Sin embargo, la alta correlación de los tres canales de información en color con respecto a la misma imagen en escala de grises puede hacer que los descriptores obtenidos mantengan al igual una elevada correlación. Si se da este caso, no se añadiría información útil para la caracterización de la escena, pues el descriptor obtenido incluiría información muy similar por triplicado.

En la Figura 5.1 es posible ver el valor del descriptor obtenido tras aplicar el descriptor HOG con un mismo número de celdas sobre una imagen en escala de gris y sobre los canales R, G y B de la misma imagen en color. Como se puede comprobar, existe una gran correlación entre los descriptores obtenidos. En este caso, crear un descriptor HOG con los canales RGB sería equivalente a formar un vector que incluyese tres veces la información del descriptor sobre la escala de grises.

Por ello, se plantean otras maneras de utilizar la información de las características de color en los descriptores. En [164], Sablak y Boult crean un descriptor mediante la obtención

de histogramas de los valores de la imagen en el espacio HSV [177]. En concreto, la caracterización de la imagen se lleva a cabo almacenando la posición de los máximos locales del histograma de los canales H, S y V de la imagen por separado.

Es posible encontrar numerosos trabajos en los que se propone trabajar en espacio HSV. Por ejemplo, Suhasini et al. [179] utilizan HSV en lugar del RGB para obtener un descriptor basado en la combinación de SIFT e ICH (*Inavariant Color Histogram*), presentando una clara mejora de la precisión en tareas de asociación de imágenes frente al mismo algoritmo aplicado sobre RGB. Junhua y Jing [95] muestran un clasificador de imágenes basado en la transformada Contourlet aplicada sobre el canal *Hue* del espacio HSV.

La información relativa al color de la escena también puede ser recogida a través del propio valor de los píxeles de cada canal en histogramas. Los histogramas contabilizan el número de píxeles dentro de un cierto rango de intensidad, por lo que representan la distribución de intensidad de una imagen digital, siendo además características independientes de la escala o resolución de la imagen.

Con el objeto de crear un descriptor útil en tareas de localización a partir de los histogramas de intensidad, especialmente cuando se trabaja con imágenes panorámicas, se realiza una división de la imagen en celdas horizontales, tal y como se hace en la construcción de los descriptores HOG y GIST (Figura 4.14). De esta forma, si la imagen es panorámica, se aprovecha que independientemente de la orientación, los niveles de intensidad incluidos de cada celda son los mismos, obteniendo características invariantes a rotación.

Para cada celda y canal de color de la escena, se crea un nuevo histograma con los valores de intensidad. Las divisiones de los histogramas serán equiespaciadas en todo el rango posible de valores de intensidad de cada canal. Al dividir por el número total de píxeles incluidos en cada celda, la suma de los valores del histograma será igual a la unidad. Este proceso se repite para los distintos canales del espacio HSV.

La transformación de la imagen del espacio RGB al HSV es un proceso computacionalmente ligero, al igual que la obtención de los histogramas de los distintos canales de color, por lo que la nueva información no conlleva un coste de tiempo de cálculo elevado. En cuanto al tamaño del descriptor, dependerá directamente del número de celdas por imagen y canales por histograma.

Esta información puede combinarse con los distintos descriptores incluidos en el Capítulo 4. En concreto, vamos a añadir los histogramas de los canales de color (HC) a distintos descriptores de Fourier (Sección 4.1), HOG (Sección 4.3) y GIST-Gabor (Sección 4.4.2). Cada descriptor estará pues compuesto por información relativa a la distribución espacial de la escena, y por la distribución del color en la imagen. GIST-Color (Sección 4.4.3) tam-

bién va a ser incluido en el estudio, pero debe destacarse que este descriptor ya considera la información de color de forma particular en su definición.

Si se considerase construir el descriptor final simplemente añadiendo la información de los distintos descriptores por separado, es posible que la diferencia en el número de componentes o en el valor del módulo de cada descriptor haga que la ponderación de uno de ellos en el descriptor compuesto sea excesivamente alto. Por ello, se normaliza cada descriptor por separado.

En la normalización de la información relativa al color, debe tenerse en cuenta el número de histogramas incluidos en el descriptor HC, así como el número de divisiones de la imagen.

Siendo h_H, h_S y h_V los histogramas de los correspondientes canales de color HSV de una celda horizontal de la imagen, se define h_{color} como el conjunto de esos histogramas. Tal y como se ha comentado anteriormente, cada uno de los histogramas se divide por el número de píxeles que abarca la celda, por lo que la suma de sus valores es igual a la unidad. Para normalizar h_{color}, debemos dividir por el número total de histogramas incluidos en él.

$$\widehat{h_{Color}} = \frac{[\, h_H \;\; h_S \;\; h_V \,]}{3} \tag{5.1}$$

El descriptor con las características de color (HC) incluye los conjuntos de histogramas h_{color} correspondientes a todas las celdas horizontales en que se divide la imagen. Siendo n el número de celdas de la imagen, el descriptor normalizado con las características de color se obtiene como:

$$HC = \frac{[\, \widehat{h_{Color_1}} \;\; \widehat{h_{Color_2}} \;\; \dots \;\; \widehat{h_{Color_n}} \,]}{n} \tag{5.2}$$

De esta forma, la suma de todos los elementos de HC será igual a 1.

Igualmente, debe realizarse la normalización de los descriptores relativos a la distribución espacial de la escena. Denotaremos ese descriptor como $D_{espacial}$ independientemente que esté basado en Fourier, HOG o GIST.

En el caso de los distintos descriptores basados en la transformada de Fourier, la normalización se hará mediante la división de cada transformada por el primer valor de la serie obtenida en dominio frecuencial, que se corresponde con el valor medio de la función. Cabe destacar que en el caso de la Firma de Fourier, este valor será distinto para cada fila de la imagen.

La normalización de los descriptores basados en HOG y GIST-Gabor se hará a través de la división de los elementos del descriptor entre la suma de todos sus valores.

En el descriptor final, será posible modificar la ponderación de los descriptores individuales normalizados. Éste puede expresarse como:

$$D_{compuesto} = [\, c_{espacial} \cdot D_{espacial} \quad c_{color} \cdot HC \,] \tag{5.3}$$

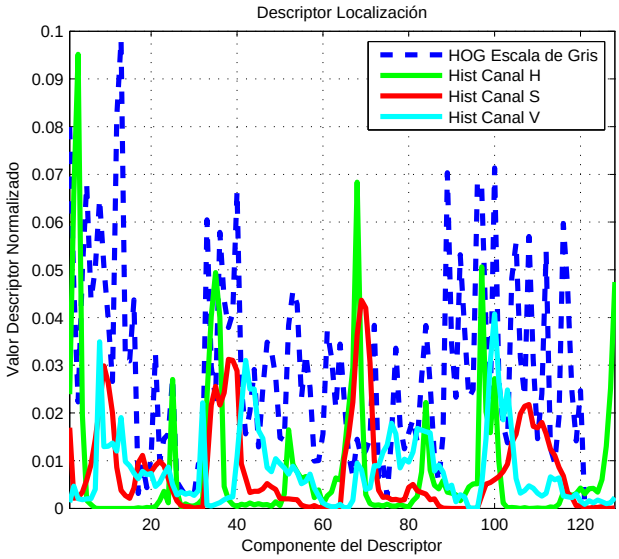


Figura 5.2: Descriptor HOG para una imagen en escala de gris, y cálculo de los histogramas de intensidad de los canales H, S, V con las mismas ventanas y de divisiones por histograma.

con $c_{espacial}$ y c_{Color} como constantes de ponderación. En los experimentos, ponderaremos los dos descriptores con el mismo valor: $c_{espacial} = c_{Color} = 0{,}5$.

Como ejemplo, la Figura 5.2 muestra gráficamente el valor del descriptor HOG sobre una imagen en escala de grises, y los histogramas de color de los canales HSV. El número de celdas horizontales y las divisiones por histograma son las mismas que en el descriptor HOG para obtener un vector de la misma longitud. En la práctica, no es necesario que los histogramas de color tengan la misma longitud que los descriptores de la distribución espacial.

Como aportación de este trabajo, se va a llevar a cabo una comparación del comportamiento de distintos descriptores de apariencia global desarrollados en el Capítulo 4 sobre una base de imágenes panorámicas, centrándonos en la utilización de la información de las escenas de color.

Para ello, aplicaremos cada una de las técnicas sobre las imágenes en escala de grises, sobre los canales RGB separadamente, sobre los canales HSV, sobre los seis canales RGB y HSV por independiente formando un solo descriptor, y añadiendo los histogramas de los niveles de brillo de los canales de color (HC) a los distintos descriptores, tal y como se ha expuesto en esta sección.

5.2 Base de Imágenes.

La base de imágenes que se describe en esta sección ha sido adquirida en distintas salas, despachos y elementos comunes de la segunda planta del primer módulo del edificio Innova de la Universidad Miguel Hernández.

En la Figura 5.3 se puede ver un plano general con todas las estancias de las que se disponen escenas. Concretamente, se incluye un pasillo (1), tres despachos con distintas distribuciones (2,3,4), una biblioteca o sala de reuniones (5) y el salón de grados (5).

Las imágenes han sido adquiridas con un conjunto catadióptrico formado por la cámara DMK-21BF03 [183], cuyas especificaciones se incluyen en la Tabla 3.2, y el espejo Eizoh Wide70 [52]. Por lo tanto, las escenas son omnidireccionales en color, con una resolución de 640×480 píxeles.

La distribución de las imágenes sigue una cuadrícula de 40×40 cm entre elementos. Las condiciones de iluminación durante las capturas son reales. La existencia de grandes ventanales en las distintas estancias dificultó la captura de las imágenes, siendo necesario reducir la ganancia de las imágenes notablemente para evitar la saturación de la escena. Debido a ello, el histograma de la mayorías de las imágenes está muy concentrado en la zona baja del espectro.

En la Tabla 5.1 aparece el número de imágenes por estancia de la base.

Zona	**Número de Imágenes**
(1) Pasillo	212
(2) Despacho 1	35
(3) Despacho 2	72
(4) Despacho 3	84
(5) Biblioteca	169
(6) Salón de Grados	300
Total	**872**

Tabla 5.1: Número de imágenes de la base de imágenes por área.

Para completar la base y realizar experimentos de asociación con escenas distintas a las que conforman el mapa, se han capturado imágenes de test repartidas por todas las estancias de la base. Estas imágenes también han sido adquiridas bajo condiciones de iluminación reales.

La localización de las escenas de test no sigue un patrón concreto, pudiendo coincidir con la posición de una imagen del mapa o no. En el caso que no coincida, se pueden dar tres

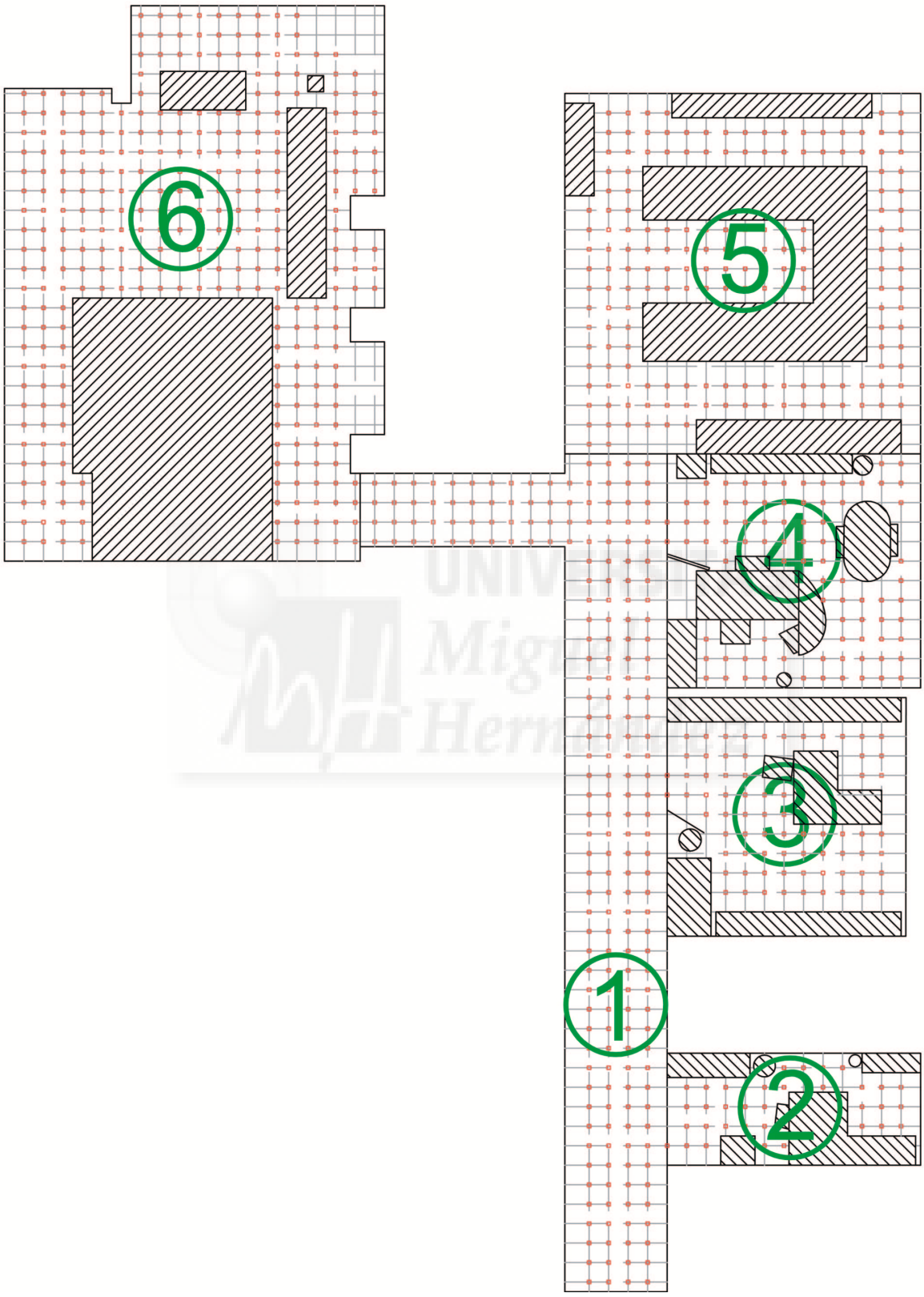


Figura 5.3: Plano de distribución de las estancias incluidas en la base de imágenes.

Zona	Número de Imágenes	Rotaciones	Total
(1) Pasillo	12	x16	192
(2) Despacho 1	9	x16	144
(3) Despacho 2	10	x16	160
(4) Despacho 3	13	x16	208
(5) Biblioteca	16	x16	256
(6) Salón de Grados	17	x16	272
Total	**77**	**x16**	**1232**

Tabla 5.2: Número de imágenes de test por área de la base experimental.

casos distintos: que esté más próxima a una sola imagen del mapa, que se encuentre entre dos puntos del mapa, o que sea equidistante a cuatro imágenes de la cuadrícula.

Las imágenes de prueba no han sido adquiridas al mismo tiempo que las que forman el mapa de las estancias. Aunque se ha intentado conseguir unas condiciones de apariencia similares, el uso frecuente de las estancias y los cambios de iluminación ha dificultado esta tarea, existiendo, por ejemplo, alteraciones en la posición de algunos elementos de las estancias.

Este hecho dificulta la asociación entre las imágenes de test y las del mapa, aunque también nos permite comprobar la robustez de los descriptores.

De cada imagen de test se han capturado 16 orientaciones distintas, con una diferencia de $22{,}5^{o}$ entre rotaciones consecutivas. La Tabla 5.2 muestra el número de imágenes de test por área.

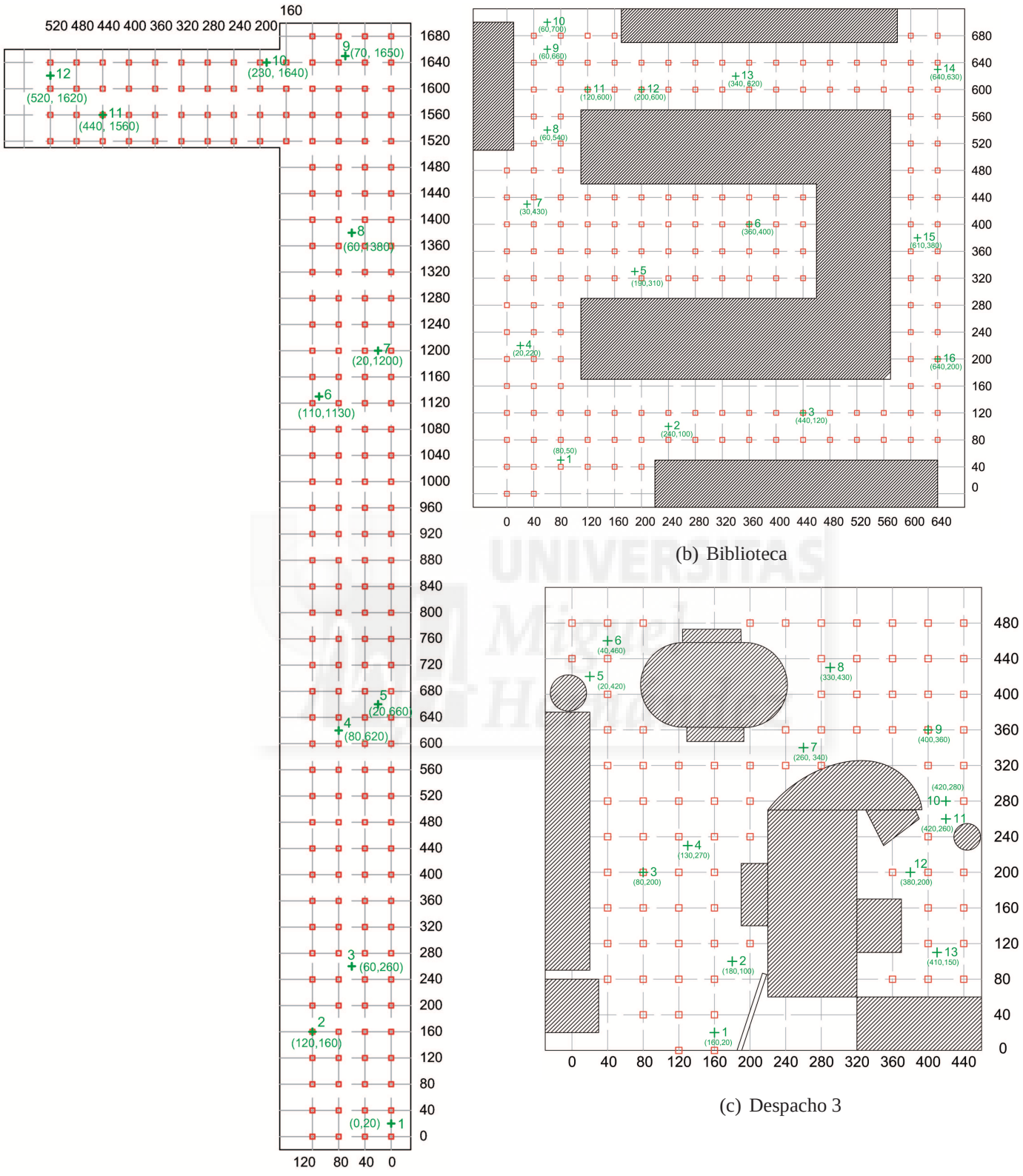
En las Figuras 5.4 y 5.5 se presentan planos detallado de cada área. Los puntos marcados en rojo pertenecen a las imágenes del mapa, mientras que los marcados en verde corresponden a las imágenes de test.

A continuación se detallan los elementos y distribución de cada una de las estancias.

En la Figura 5.4 (a) se incluye el plano del Pasillo. La longitud total del pasillo es de 21 metros, con una anchura aproximada de 180 cm. Tiene elementos poco diferenciadores, ya que se repite el mismo esquema de puertas y ventanales con una distribución espacial muy similar. La hora de captura varía entre las 9h y las 12h.

La Figura 5.5 (b) muestra el Despacho 1. Es la estancia más pequeña incluida en la base. Los elementos más destacables son dos mesas, dos estanterías, tres sillas y distintos cuadros en las paredes. La hora de adquisición de las imágenes comprende de 12h a 14h.

En el Despacho 2, cuyo plano se incluye en la Figura 5.5 (c) pueden destacarse dos grandes estanterías, una mesa central y dos archivadores en el fondo de la estancia. Debido a la gran superficie de ventanas con iluminación directa, se disminuyó notablemente la ganancia

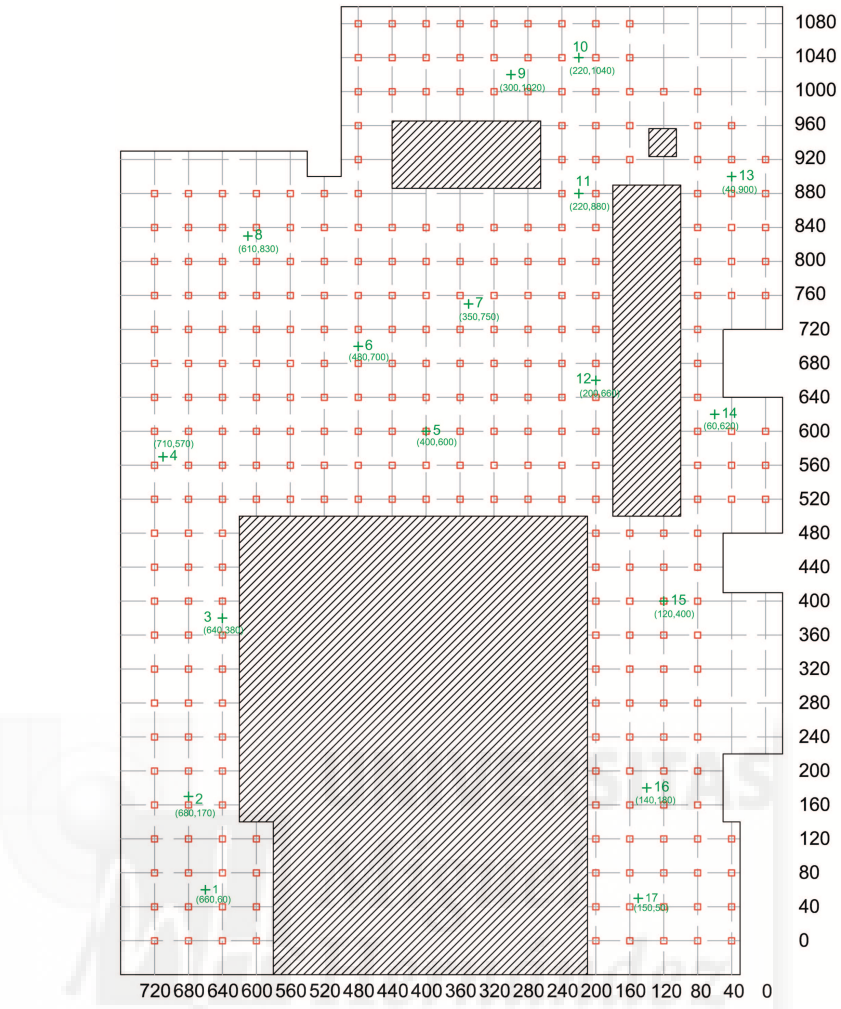


(a) Pasillo

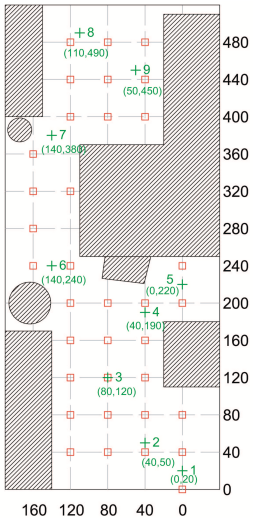
(b) Biblioteca

(c) Despacho 3

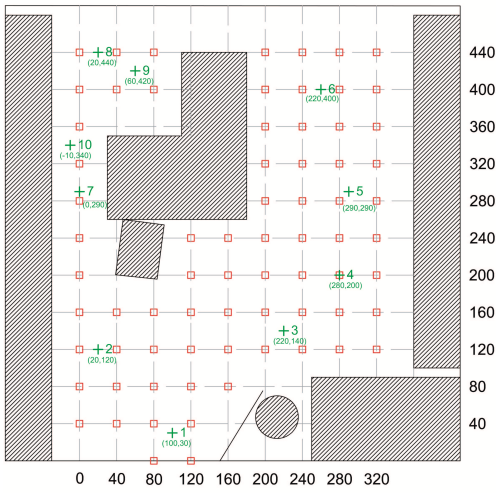
Figura 5.4: Detalle de la posición de las imágenes del mapa (rojo) y de test (verde) para las zonas (a) 1, (b) 5 y (c) 4 de la base de imágenes.



(a) Salón de Grados



(b) Despacho 1



(c) Despacho 2

Figura 5.5: Detalle de la posición de las imágenes del mapa (rojo) y de test (verde) para las zonas (a) 6, (b) 2 y (c) 3 de la base de imágenes.

del sensor visual al adquirir las escenas. Las imágenes se capturaron de 12:30 a 13:30 para evitar la entrada de sol directo.

El Despacho 3 (Figura 5.4 (c)) tiene como elementos destacables dos mesas, un armario, distintas sillas y dos cajoneras. El intervalo de captura varía de las 16h a las 18h.

La Biblioteca, recogida en la Figura 5.4 (b), tiene ventanales en distintas direcciones, por lo que ha sido inevitable la entrada directa de luz solar. Tiene numerosos elementos que se repiten a lo largo de toda la estancia, como estanterías, mesas y sillas con la misma apariencia. La adquisición de las imágenes se ha realizado entre las 16h y las 19h.

El Salón de Grados, cuyo plano aparece en la Figura 5.4 (a), es la estancia de mayor superficie. En ella se dan condiciones de iluminación muy distintas dependiendo de la zona de la sala, con áreas cercanas a ventanales y otras a las cuales no llega la luz natural. El elemento más distintivo es un grupo central de butacas, que no se presenta en ninguna otra sala. También aparecen un par de mesas y distintas sillas. La hora de captura varía de las 12h a las 19h.

La mayoría de descriptores se basan en la transformada panorámica de la imagen omnidireccional. Por ello, obtenemos la proyección cilíndrica de las imágenes omnidireccionales. Las escenas panorámicas se obtienen mediante el cambio de sistemas de coordenadas cilíndrico de la imagen omnidireccional a cartesiano. El tamaño final de estas imágenes es de 128x512 píxeles. La Figura 5.6 incluye ejemplos de escenas de cada estancia.

En la Sección 3.2.2 puede obtenerse más información sobre la obtención de la escena panorámica a partir de la omnidireccional.

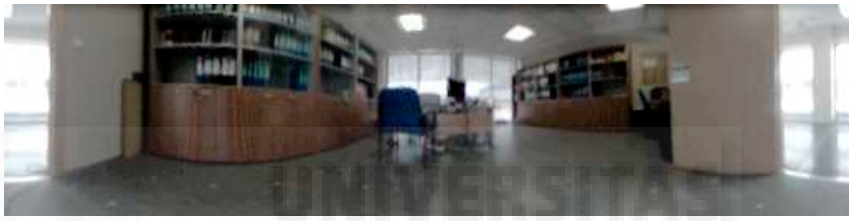
Conviene destacar que en la mayor parte de las escenas de la base no existen elementos que sean claramente diferenciadores del resto de imágenes. Al ser un ambiente de oficinas, existen numerosos elementos que se repiten en las distintas estancias, o con características muy similares. Como ejemplo, tenemos las sillas que aparecen en los distintos despachos y en la biblioteca, las estanterías repartidas por las diferentes estancias, o las puertas. Además, todas las estancias están pintadas con el mismo color, y el mobiliario es muy similar.

Por esta razón, puede darse con facilidad *aliasing* entre distintas escenas de la base. El *aliasing* es un efecto que provoca que una señal continua muestreada de forma digital no pueda ser reconstruida de forma unívoca a partir de su señal digital. En nuestro caso, esto se traduce en que los descriptores no tienen la capacidad de asociar una imagen de test a otra imagen del mapa de forma diferenciada con el resto de imágenes, debido a la existencia de distintas posiciones con apariencia muy similares.

Como ejemplo, en la Figura 5.7 vemos dos escenas pertenecientes al Pasillo adquiridas en distintas localizaciones, con una distancia de 240 cm entre ellas. Visualmente podemos percatarnos que las escenas tienen una apariencia muy similar. Ante una imagen de test

(a) Pasillo

(b) Despacho 1



(c) Despacho 2

(d) Despacho 3

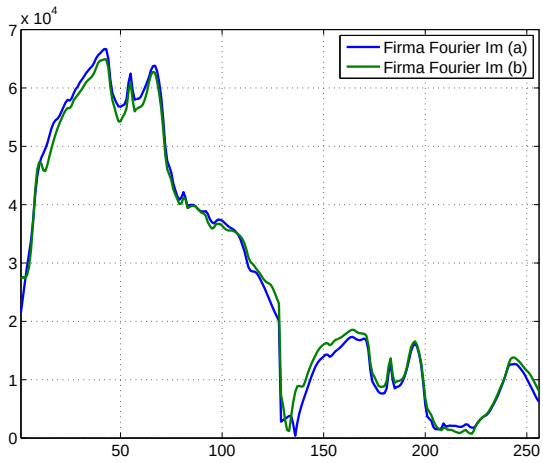
(e) Biblioteca

(f) Salón de Grados

Figura 5.6: Imágenes de ejemplo de cada estancia incluida en la base.

(a) Pasillo X=40, Y=280.

(b) Pasillo X=40, Y=520.



(c)

Figura 5.7: Ejemplo de *aliasing* visual.

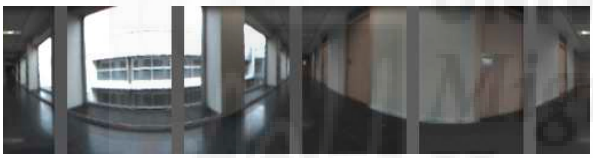
adquirida en las proximidad de un punto cercano a cualquiera de las dos imágenes, sería difícil determinar cuál es la imagen más cercana, pues la distancia entre los descriptores sería muy similar para ambos casos. En la Figura 5.7 (c) se representa la Firma de Fourier de ambas imágenes. Puede comprobarse como los descriptores son muy similares a pesar de tratarse de escenas separadas espacialmente.

Para comprobar la robustez de los distintos descriptores, introduciremos en las imágenes de test oclusiones y ruido. En concreto, las oclusiones se harán mediante cuatro bandas verticales con distintas anchuras que ocultarán distintos porcentajes de la imagen. Respecto al ruido, añadiremos artificialmente ruido Gaussiano de media 0 y distintas varianzas a cada canal de la imagen panorámica.

La Figura 5.8 recoge una imagen panorámica a la cual se ha añadido distintas oclusiones y ruido. Las oclusiones van del 5 % del total de la imagen, al 40 %, mientras que la varianza del ruido Gaussiano varía de una $\sigma = 0{,}0025$, a $\sigma = 0{,}0500$.

(a) Oclusión = 5%

(e) Ruido $\sigma = 0{,}0025$



(b) Oclusión = 10%

(f) Ruido $\sigma = 0{,}0050$

(c) Oclusión = 25%

(g) Ruido $\sigma = 0{,}0100$

(d) Oclusión = 40%

(h) Ruido $\sigma = 0{,}0200$

Figura 5.8: Imágenes de ejemplo incluyendo oclusiones y ruido Gaussiano.

5.3 Experimentos y Resultados

Los experimentos realizados se centran en el análisis y comparación de las distintas técnicas de descripción global de la información visual usando la información de color sobre escenas panorámicas.

Para ello, se estudia la precisión de estimación de la pose de las distintas imágenes de test sobre el mapa denso mostrado en la sección anterior. También se incluye una comparación del coste computacional de cada uno de los descriptores.

Primero, se lleva a cabo un estudio sobre las imágenes en escala de grises (menos GIST-Color, que se realiza sobre RGB), lo que nos permite ajustar los parámetros de cada descriptor, y hacer una primera comparación de los requisitos computacionales y la precisión de cada descriptor en tareas de localización.

La comparación se completa mediante el uso de la información de color tal y como se describe en la Sección 5.1, y la introducción de ruido y oclusiones en las imágenes de test.

Los algoritmos y las distintas simulaciones han sido desarrollados sobre el software Matlab R2012b. Los experimentos han sido realizados con un ordenador equipado con dos procesadores Quad-Core con velocidad de 2.8GHz, y 10 GB de memoria RAM.

5.3.1 Variables de los Descriptores

En el Capítulo 4 se describen detalladamente cada una de las técnicas de compresión de imágenes basadas en la apariencia global que van a incluirse en la comparación.

A continuación se incluye un resumen de los principales parámetros que se pueden modificar de cada uno de ellos.

En los descriptores basados en la Transformada de Fourier, es posible seleccionar el número de elementos de la transformada en el espacio frecuencial. En el caso de Fourier 1D, se elige el número de componentes de la transformada del vector formado por el valor medio de las columnas de la imagen. En Fourier 2D, se determina el tamaño de la submatriz que recoge las frecuencias más bajas de la transformada de la imagen. Con respecto a la Firma de Fourier, seleccionamos el número de elementos que se preservan de cada fila.

Para todos los descriptores de Fourier, se elige separadamente el número de elementos del módulo de la transformada, que nos permitirá localizar la posición del robot en el mapa, y de la fase, que nos proporciona información para conocer su orientación.

La técnica que aplica el análisis PCA sobre la Firma de Fourier queda determinada por el número de elementos del módulo de la transformada por fila, y del número de vectores que forman la base de proyección tras el análisis PCA para realizar la localización. Respecto

al cálculo de la orientación, como no se aplica el análisis PCA a la fase de la transformada, sigue dependiendo únicamente del número de elementos que seleccionamos por fila de fase.

En PCA Rotaciones, el descriptor depende del número de rotaciones que se realizan por imagen, y de los vectores principales seleccionados tras la descomposición de la base.

El descriptor de Histogramas de Orientación del Gradiente queda determinado por el número de celdas horizontales (con la misma anchura de la escena) en que se divide la imagen para llevar a cabo la localización, y por la anchura de las celdas verticales (con la misma altura de la escena) además de la distancia entre ellas para el cálculo de la orientación.

Nótese que entre las celdas horizontales no se va a producir solapamiento, por lo que el número de celdas horizontales determinará su altura. Sin embargo, en las celdas verticales sí puede producirse solapamiento si la distancia de aplicación de las celdas es menor que su anchura. El número de celdas verticales dependerá directamente de la anchura de la imagen, y de la distancia entre celdas consecutivas. Cabe destacar que en el descriptor también puede modificarse el número de divisiones que se realiza de cada histograma. Sin embargo, los resultados experimentales muestran que con 8 divisiones (o *bins*) por histograma es suficiente, ya que aumentarlo no supone una mejora de la precisión en la estimación de la pose, y sin embargo, su disminución sí provoca un empeoramiento de los resultados. Por ello, se fija el número de canales por histograma a 8.

Los parámetros de GIST-Gabor están compuestos por el número de máscaras con los que se filtra la imagen, y las celdas en que se dividen las imágenes filtradas. Para la localización, se usan como máximo dos escalas espaciales de filtrado de Gabor para limitar el coste computacional del descriptor. Por tanto, el descriptor de localización depende del número de máscaras usado en las dos escalas espaciales, y el número de celdas horizontales usados para dividir cada imagen filtrada. La dirección de filtrado de las máscaras de Gabor es función del número de máscaras utilizadas, pues están repartidas de forma equiangular entre 0 y 180^o. Para el cálculo del desfase, se utilizan únicamente imágenes filtradas con máscaras de Gabor de la primera escala espacial, con un límite de 4 máscaras.Así pues, el descriptor para la estimación de la orientación tiene como variables la anchura de las celdas verticales empleadas, y la distancia entre ellas.

GIST-Color utiliza un número fijo de filtros de Gabor, tal y como aparece en la Sección 4.4.3, con 4 orientaciones distintas. Las escalas espaciales del filtrado vienen determinadas por el número de escalas de la pirámide Gaussiana de la imagen de entrada. Ante una nueva imagen de entrada, se crean seis reducciones del tamaño original de la imagen de entrada. Para el filtrado de Gabor, se utilizan únicamente los tres primeros niveles de la pirámide. En cuanto a las características de color, las comparaciones entre canales opuestos de color utilizan los 6 niveles de la pirámide, tal y como recoge en la Tabla 4.1. Los parámetros

modificables de este descriptor son relativos a las divisiones que se realizan a los resultados de los filtrados de Gabor y las comparaciones de los canales de color. Para la estimación de la posición, se elegirán el número de celdas (o bloques) empleados para dividir las información de orientación y el número de celdas en que se divide la información relativa al color. El descriptor de orientación se formará únicamente con la información de orientación, es decir, con las imágenes filtradas por las máscaras de Gabor. Sus parámetros son el número de celdas verticales usadas, y la distancia de aplicación entre ellas.

En la Tabla 5.3 se recogen los distintos parámetros para cada descriptor.

5.3.2 **Análisis Comparativo**

A continuación se incluyen los resultados de los distintos experimentos llevados a cabo.

Hay que destacar que para las simulaciones de PCA Rotaciones no se ha utilizado la base de imágenes completa. Los requerimientos computacionales del algoritmo en el cálculo del mapa obligan a reducir la cantidad de escenas incluidas en el mapa. Por ello, aparecerá marcado con un asterisco en las gráficas comparativas. Para los experimentos de este descriptor se han utilizado únicamente tres zonas de la base, que corresponden a los tres despachos, es decir, las zonas 2, 3 y 4 (Tabla 5.2). En total, el mapa reducido se compondrá de 191 imágenes, con 32 localizaciones distintas de test, lo que supone 512 imágenes de test diferentes considerando las rotaciones.

Primero se muestra una comparación de los distintos descriptores aplicados sobre las escenas en escala de grises (a excepción de GIST-Color, que se estudia sobre RGB). Esta primera comparación sirve para seleccionar los parámetros de cada método y conocer los requerimientos de cada descriptor.

Posteriormente, se presentan los resultados de los distintos descriptores cuando se añade la información relativa al color.

Por último, se incluye el estudio del comportamiento de las distintas técnicas ante ruido Gaussiano y oclusiones de las imágenes de test.

5.3.2.1 **Selección de Parámetros**

Para realizar la selección de los parámetros de cada método, vamos a considerar únicamente el espacio de color original del descriptor. Este espacio de color corresponde a la escala de grises, menos GIST-Color, que es RGB.

En la selección de los diferentes parámetros se ha preponderado la precisión en la localización sobre los requerimientos computacionales. Sin embargo, el tamaño del mapa y el tiempo de cómputo han sido también considerados, limitando el número de componentes de los descriptores si la mejora en la precisión no era apreciable.

Fourier 1D	Posición	N	Elementos del módulo de la Transformada de Fourier.
	Orientación	N_{rot}	Elementos de fase de la Transformada de Fourier.
Fourier 2D	Posición	N	Tamaño de la submatriz de módulos de la Transformada 2D de Fourier ($N \times N$).
	Orientación	N_{rot}	Tamaño de la submatriz de fases de la Transformada 2D de Fourier ($N_{rot} \times N_{rot}$)
Firma de Fourier	Posición	N	Elementos del módulo de la Transformada de Fourier de cada fila.
	Orientación	N_{rot}	Elementos de fase de la Transformada de Fourier de cada fila.
Fourier sobre PCA	Posición	N	Elementos del módulo de la Transformada de Fourier de cada fila.
		V_{PCA}	Número de vectores principales seleccionados tras el análisis PCA.
	Orientación	N_{rot}	Número de elementos de fase de la Transformada de Fourier de cada fila.
PCA Rotaciones	Posición y Orientación	R_{im}	Rotaciones artificiales equiangulares de la imagen.
		V_{PCA}	Número de vectores principales seleccionados tras el análisis PCA.
HOG	Posición	C_H	Número de celdas horizontales.
	Orientación	S_V	Anchura de las celdas verticales (píxeles).
		D_V	Distancia entre celdas verticales consecutivas (píxeles).
GIST-Gabor	Posición	$Masc_1$	Número de filtros de Gabor de primera escala.
		$Masc_2$	Número de filtros de Gabor de segunda escala.
		C_H	Número de celdas horizontales.
	Orientación	S_V	Anchura de las celdas verticales (píxeles).
		D_V	Distancia entre celdas verticales consecutivas (píxeles).
GIST-Color	Posición	B_G	Número de celdas (o bloques) horizontales de las imágenes de Gabor.
		B_C	Número de celdas (o bloques) horizontales de los canales de colores opuestos.
	Orientación	S_V	Anchura de las celdas verticales (píxeles).
		D_V	Distancia entre celdas verticales consecutivas (píxeles).

Tabla 5.3: Parámetros de cada descriptor.

Los resultados de localización se dividen en la estimación de la posición en el mapa, y el cálculo de la orientación.

La estimación de la posición se realizará mediante la asociación de cada imagen de test con la imagen más cercana del mapa. La precisión en el cálculo de la posición se presenta de dos formas: mediante curvas *Recall-Precision*, y mediante la distancia métrica entre la imagen de test y la imagen del mapa con menor distancia imagen, es decir, con menor distancia entre los respectivos descriptores.

Las curvas *Recall-Precision* [69, 159] evalúan el comportamiento de los descriptores en su tarea de asociación de características. Los conceptos de *Recall* y *Precision* se definen en las Ecuaciones 5.6 y 5.5 respectivamente.

$$recall = \frac{\text{num. de correspondencias correctas seleccionadas}}{\text{num. total de correspondencias correctas}} \tag{5.4}$$

$$precision = \frac{\text{num. de correspondencias correctas seleccionadas}}{\text{num. de correspondencias seleccionadas}} \tag{5.5}$$

Tal y como están definidos, *recall* es un índice de la capacidad del descriptor para encontrar todas las asociaciones correctas, mientras que *precision* mide la capacidad de obtener asociaciones correctas cuando el número de asociaciones seleccionadas varía. Su valor está entre 0 (que indicaría que no se ha seleccionado ninguna correspondencia correctamente), y 1 (que denota que el descriptor ha sido capaz de encontrar todas las correspondencias correctas).

El proceso para la estimación de las curvas *Recall-Precision* es el siguiente:

Primero, se calcula la distancia imagen entre la imagen de test y las imágenes que componen el mapa. Siendo $D^T = [\, d_1^T, d_2^T, \ldots, d_n^T \,]$ el descriptor de la imagen de test compuesto por n componentes, y $D^i = [\, d_1^i, d_2^i, \ldots, d_n^i \,]$ el descriptor de la i-ésima imagen que compone el mapa, la distancia imagen se define como la distancia Euclídea de ambos descriptores:

$$d_E^{T,i}(D^T, D^i) = \sqrt{\sum_{j=1}^{n} (d_j^T - d_j^i)^2}, \quad i = 1, \cdots, M, \tag{5.6}$$

con M igual al número de imágenes que componen el mapa.

La asociación de entre la imagen de test y la más cercana del mapa se lleva a cabo mediante el criterio de la mínima distancia imagen. Por lo tanto, una vez calculadas todas las la distancia Euclídeas entre la escena de test y las del mapa, se selecciona el caso la imagen con la mínima distancia.

Para determinar si la asociación es correcta, se comprueba si la imagen del mapa asociada a la mínima distancia Euclídea es la posición más cercana métricamente a la imagen de test.

Tras repetir este proceso para todas las imágenes de test, se obtiene un vector de dos columnas, que incluye la mínima Distancia Euclídea de cada experimento, y el resultado de la asociación, que toma valores 1 o 0 dependiendo de si es correcta o falsa.

A continuación, se ordena la lista de asociaciones de todos los experimentos de forma ascendente respecto a la distancia Euclídea, y se obtienen los valores de *Recall* y *Precision* definidas anteriormente. Debe tenerse en cuenta que para cada experimento se obtendrá un valor distinto de *Recall* y *Precision*, que reflejará la capacidad del algoritmo a asociar correctamente dado un cierto umbral. Dicho umbral se corresponde en nuestro caso a la d_E del experimento considerado. Debido a que las asociaciones han sido ordenadas por orden ascendente de distancia imagen, a medida que aumenta *recall*, también aumenta valor del umbral.

En nuestro caso, no usaremos ningún umbral máximo para considerar una asociación correcta. Esto significa que por cada experimento se obtiene una correspondencia correcta, que es el caso asociado a una menor distancia imagen, también conocido como vecino más cercano. Por ello, el número total de correspondencias correctas es igual al número de total experimentos.

Además, los valores de *recall* y de *precision* del último experimento coinciden, pues el número de correspondencias seleccionadas y el número total de correspondencias correctas en ese caso es el mismo.

Para completar los resultados, consideraremos tres casos distintos para crear las gráficas *Recall-Precision*, que considerarán si se realiza la asociación correcta con sólo la imagen del mapa con mínima distancia Euclídea o vecino más cercano (denotado por sus siglas en inglés N.N., o *Nearest Neighbour*), considerando los dos vecinos más cercanos (S.N.N.), o los tres vecinos más cercanos (T.N.N), es decir, los tres casos con menor distancia imagen para cada experimento.

De las gráficas obtenidas, no debemos tener sólo en cuenta el valor final. La propia distribución de la curva nos informa de la robustez del descriptor ante falsos positivos dentro de un cierto umbral para la distancia imagen. Así pues, será preferible que los valores se mantengan en todo momento con una precisión alta cualquiera que sea el valor de *recall*, pues indicará que se introducen menos falsos positivos dentro de ese umbral.

Para ilustrar este ejemplo, en la Figura 5.9 se pueden ver dos ejemplos de curvas *Recall* − *Precision* distintas. Ambas tienen un valor final de precisión similar, siendo mayor el de *R-P 2*. Sin embargo, el comportamiento de las curvas es distinto. En el caso del segundo ejemplo, tenemos un mayor número de falsos positivos entre los aciertos asociados con umbrales más altos. Si fijamos el umbral de la correspondencia asociada a $recall = 0{,}3$, se puede apreciar que la precisión de *R-P 1* es del 100 %, mientras que *R-P 2* se situaría en torno al 82 %.

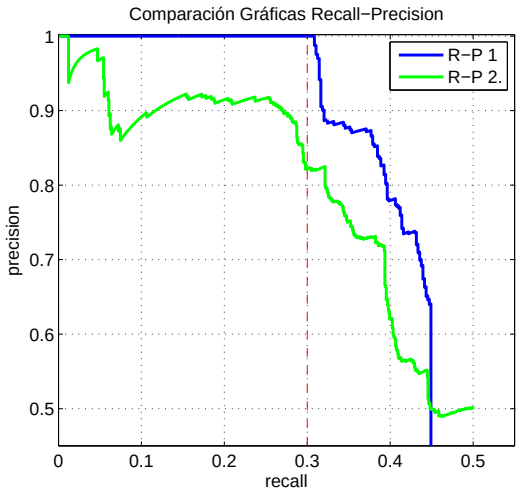


Figura 5.9: Comparación de dos curvas Recall-Precision.

Dicho de otro modo, fijando los respectivos umbrales, seríamos capaces de obtener el 30% de asociaciones correctas con un 100% de probabilidad en el caso del primer descriptor, y con un 82% de probabilidad para el segundo descriptor.

Los parámetros seleccionados experimentalmente se recogen en la Tabla 5.4, y los resultados *Recall-Precision* para los distintos descriptores se muestran en la Figura 5.10.

En ella podemos apreciar que los resultados de HOG y GIST-Color (Figuras 5.10 (f) y 5.10 (h)) superan al resto de descriptores. Además, el comportamiento sobre los falsos positivos es muy similar. A continuación, se puede destacar el comportamiento de PCA-Rotaciones, especialmente debido a su baja tasa de falsos positivos hasta un valor de *recall* relativamente alto. La precisión final de la Firma de Fourier (Figura 5.10 (b)), Fourier 2D (Figura 5.10 (c)) y GIST-Gabor (Figura 5.10 (g)) es muy similar, aunque destaca la alta precisión de este último descriptor hasta un *recall* del 40%. Por su lado, Fourier 1D y el descriptor basado en la Firma de Fourier+PCA (Figuras 5.10 (a) y 5.10 (d)) son lo que presentan una tasa más baja de asociaciones correctas.

Por otro lado, en la Figura 5.11 se presenta la distancia métrica de la imagen de test al vecino más cercano en distancia imagen. Esta gráfica de barras nos permite estudiar si los descriptores, aunque no encuentren exactamente la imagen más cercana en el mapa, realizan una asociación cercana a la correcta, o no. Se puede apreciar que los resultados siguen una tendencia muy similar a la mostrada por las curvas *recall-precision*, siendo HOG y GIST-Color los descriptores que mejores resultados presentan, pudiendo destacar el primero de ellos.

Estos resultados también muestran que, aunque la Firma de Fourier, Fourier 2D y GIST-Gabor presentan unos resultados de precisión finales muy similares en el análisis *recall-*

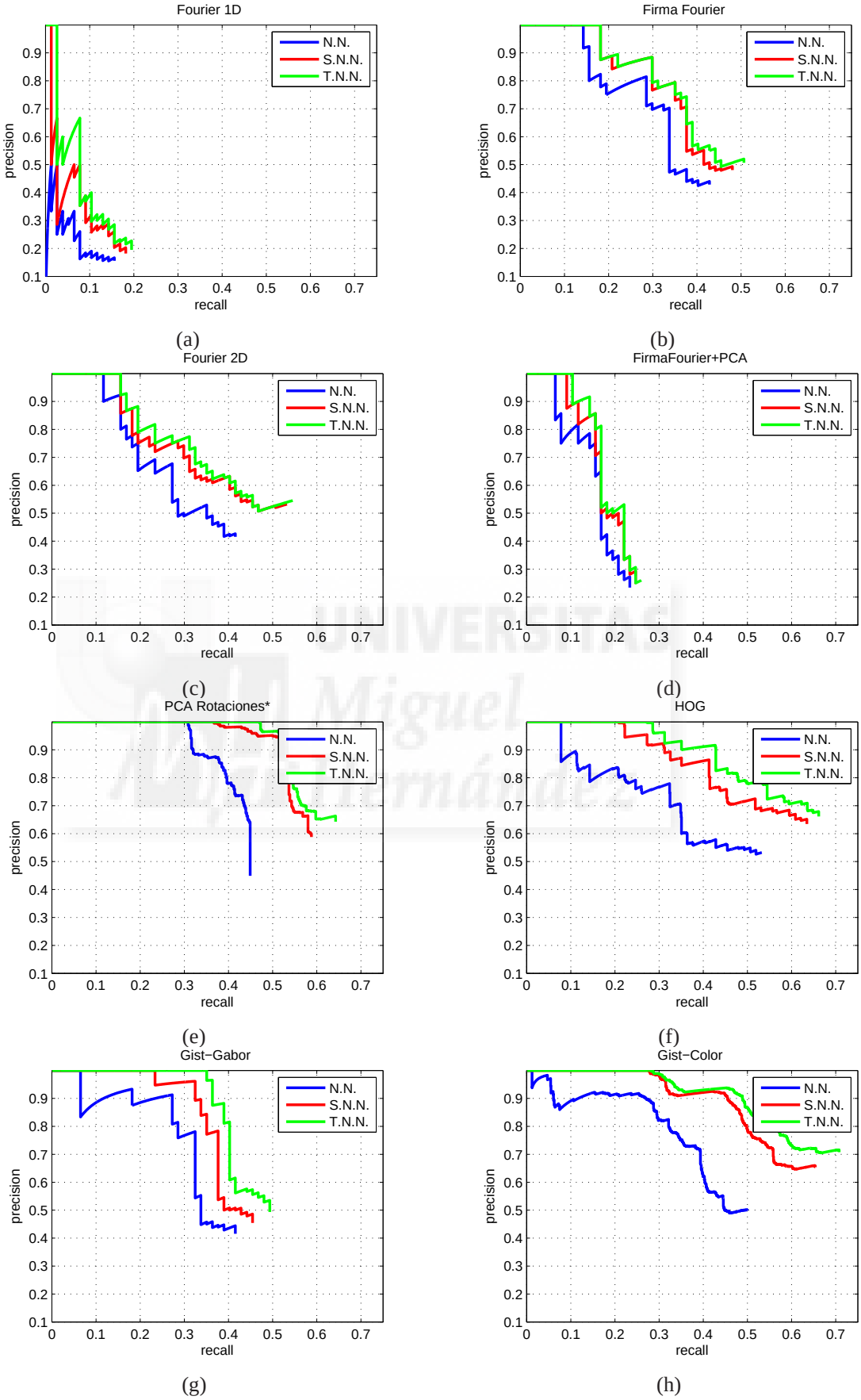


Figura 5.10: Gráficas *Recall-Precision* incluyendo el vecino más cercano (N.N.), los dos más cercanos (S.N.N.) o los tres vecinos más cercanos (T.N.N.).

Fourier 1D	Posición	N	32
	Orientación	N_{rot}	4
Fourier 2D	Posición	N	64
	Orientación	N_{rot}	8
Firma de Fourier	Posición	N	32
	Orientación	N_{rot}	16
Fourier sobre PCA	Posición	N	32
		V_{PCA}	872
	Orientación	N_{rot}	16
PCA Rotaciones	Posición y Orientación	R_{im}	16
		V_{PCA}	100
HOG	Posición	C_H	64
	Orientación	S_V	16
		D_V	4
GIST-Gabor	Posición	$Masc_1$	4
		$Masc_2$	8
		C_H	64
	Orientación	S_V	64
		D_V	32
GIST-Color	Posición	B_G	8
		B_C	32
	Orientación	S_V	8
		D_V	16

Tabla 5.4: Parámetros seleccionados para cada descriptor.

precision, el descriptor basado en GIST supera notablemente a los descriptores basados en Fourier en cuanto a proximidad métrica de la asociación, obteniendo unos resultados similares a PCA Rotaciones.

En cuanto a la estimación de la fase, en la Figura 5.12 se estudia el desfase entre la orientación real y la orientación estimada de las imágenes de test. Para valorar la robustez del descriptor en el cálculo de la fase, vamos a considerar los resultados de las asociaciones cuya distancia métrica en el mapa sea igual o menor a 40cm. De esa forma, evitamos estudiar el desfase entre imágenes cuya localización haya sido errónea.

Los algoritmos más precisos en la estimación de fase son PCA Rotaciones, GIST-Gabor, GIST-Color y HOG. Sin embargo, conviene recordar que estos son descriptores cuya estimación de fase está muestreada, bien sea por el número de rotaciones de la imagen incluidas en el mapa (en el caso de PCA Rotaciones), o por el número de ventanas verticales utilizadas.

La exactitud de fase de la Firma de Fourier y Fourier 2D son similares. Nótese que la

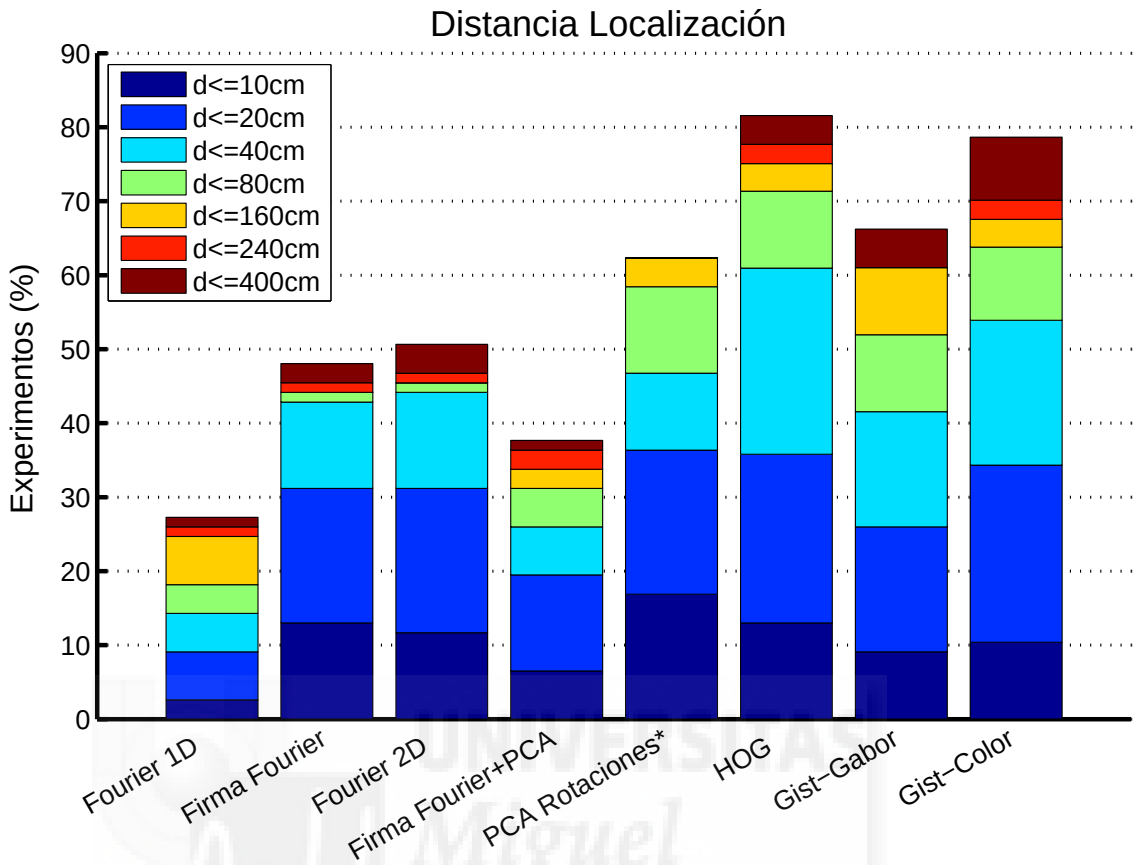


Figura 5.11: Distancia métrica entre el vecino más cercano del mapa y la posición estimada de las imágenes de test.

estimación de fase para la Firma de Fourier y para Firma de Fourier+PCA utiliza la misma información y algoritmo. La pequeña diferencia en los resultados de ambos descriptores es consecuencia de las asociaciones consideradas, ya que son diferentes para ambos descriptores.

Fourier 1D presenta la peor precisión en estimación de fase. Aún así, consigue que el 80 % de los experimentos tenga un error igual o menor a 10^{o} utilizando únicamente 4 términos por escena.

En estos experimentos, el mapa está constituido por todos los descriptores de las imágenes de las diferentes estancias, a excepción de las de test. La Figura 5.13 recoge el tamaño del mapa usando los distintos descriptores, dividiendo la medición en la cantidad de memoria para almacenar la información de estimación de posición y orientación por separado. El descriptor más compacto por diferencia es Fourier 1D, seguido por HOG y los descriptores GIST. Hay que destacar que para mejorar la precisión en la estimación de la fase de HOG, GIST y PCA Rotaciones, el aumento del tamaño del descriptor sería notable, pues es

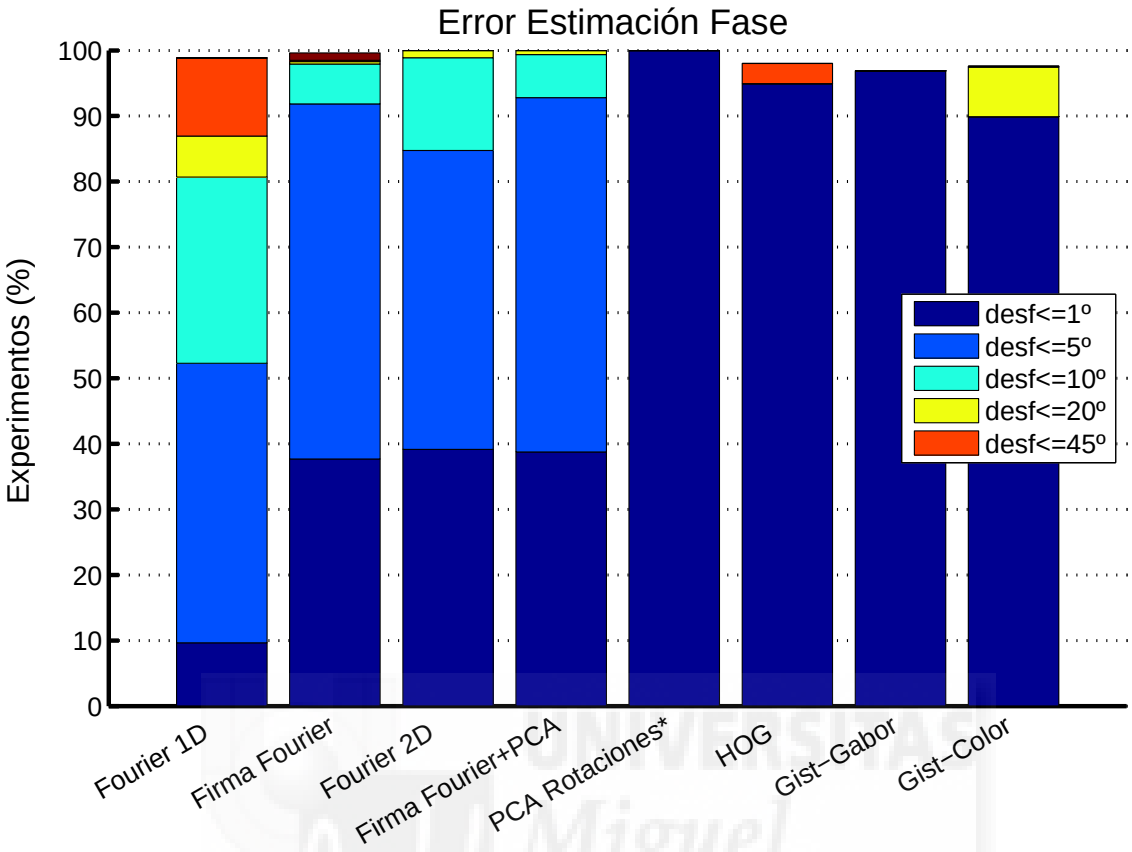


Figura 5.12: Desfase entre la la orientación real del robot y la obtenida experimentalmente.

proporcional al número de celdas verticales aplicadas sobre la imagen.

Respecto PCA Rotaciones, la medición de memoria incluye la base de proyección con los vectores seleccionados, la proyección del mapa original sobre el nuevo espacio, y la diferencia de fases entre proyecciones consecutivas. Como puede verse en la gráfica, la información de fases es despreciable frente al resto.

Además, los requisitos computacionales durante la obtención de la nueva base de proyección son todavía mayores. Considerando únicamente la matriz que contiene las imágenes (de tamaño 128x512) y todas sus rotaciones (un total de 16), y teniendo en cuenta que cada escalar de tipo double ocupa 8 bytes, la memoria necesaria para almacenar las 872 imágenes del mapa completo sería de $128 \times 512 \times 16 \times 872 \times 8 = 7{,}314{,}866{,}176$ bytes $= 6{,}81$ Gbytes. Por esta razón, nos vemos obligados a limitar el número de imágenes que contiene el mapa.

Por último, podemos ver la reducción de memoria que se logra al aplicar PCA sobre la firma de Fourier, pasando la información relativa a la localización de 29Mbytes a 4Mbytes.

La Figura 5.14 muestra las mediciones de tiempo empleado para crear el mapa, y para llevar a cabo la estimación de la pose del robot, incluyendo tanto posición como orientación.

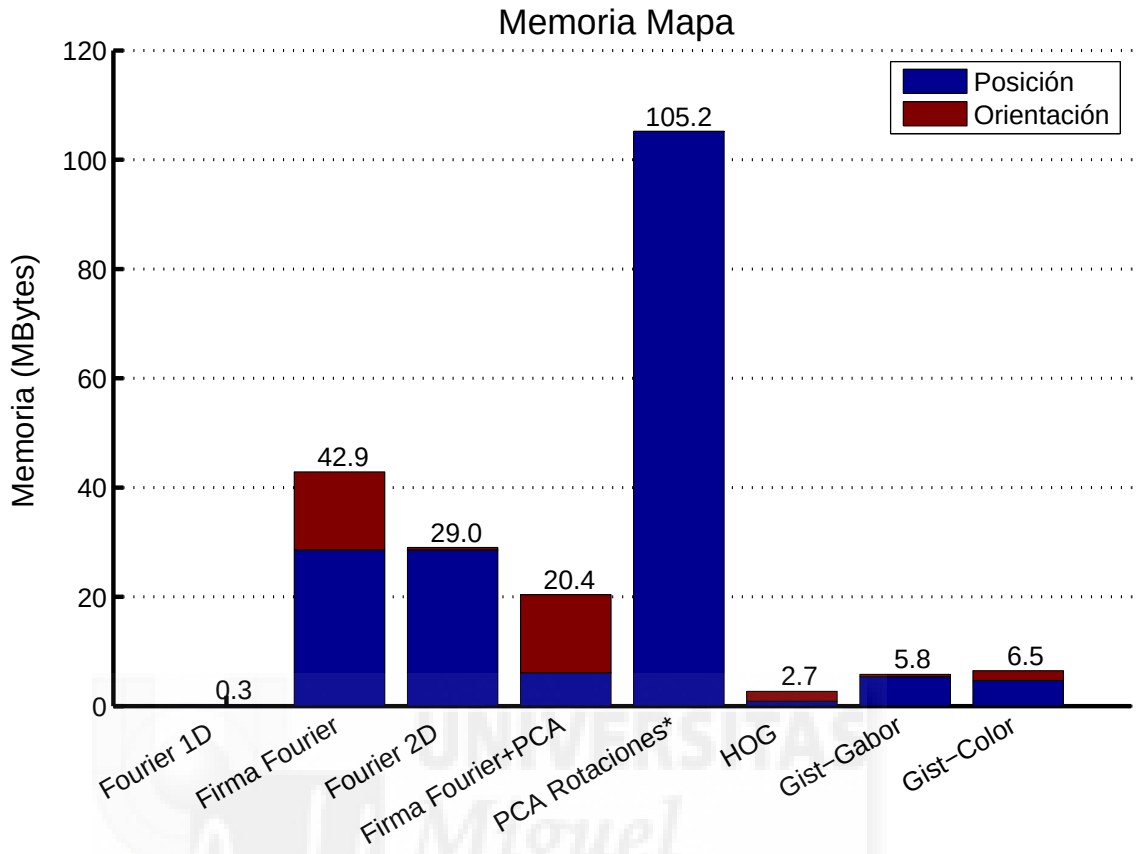
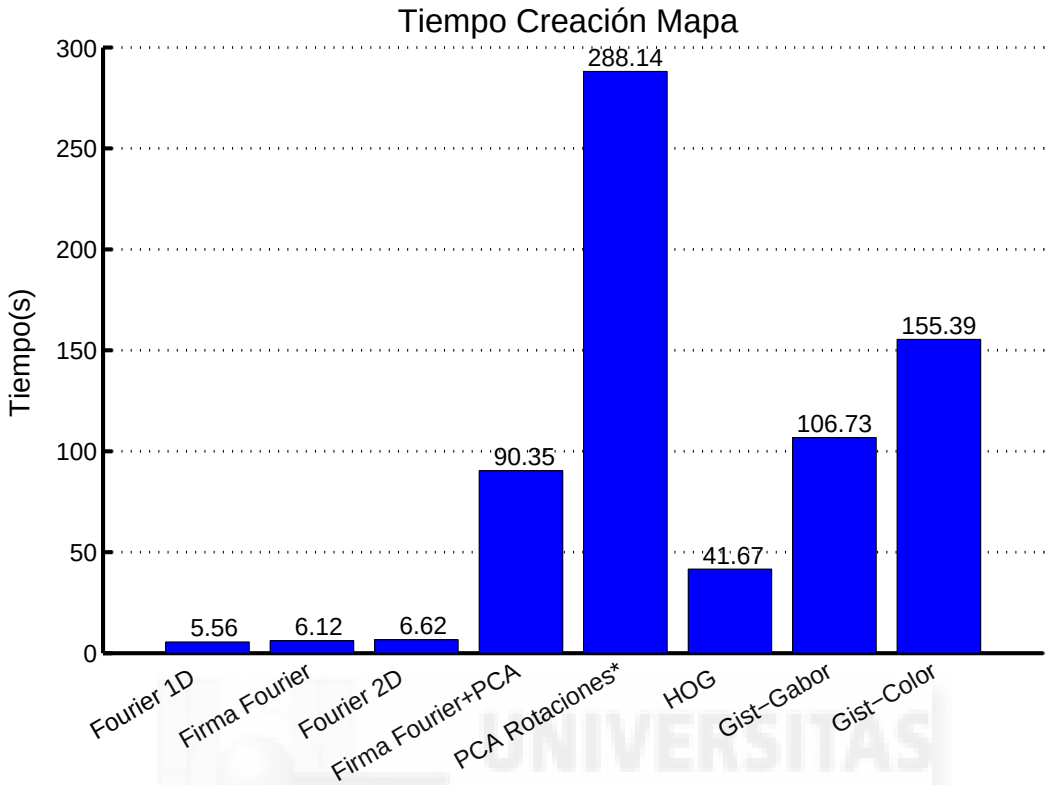


Figura 5.13: Memoria necesaria para almacenar el mapa.

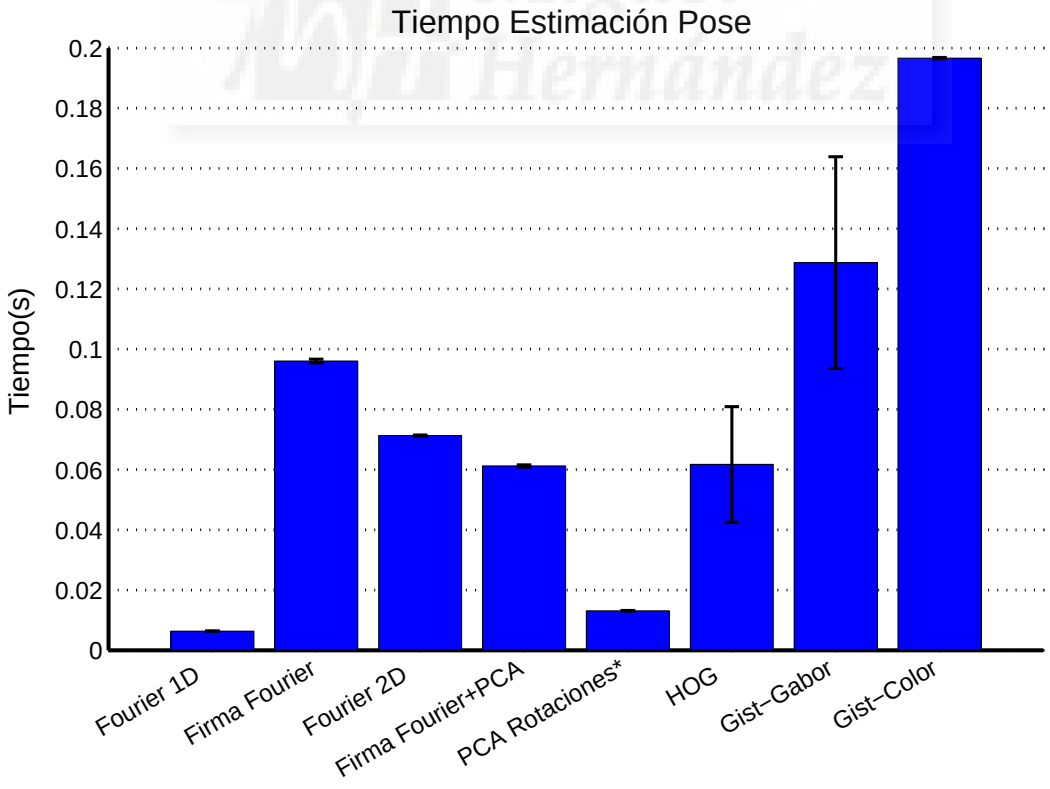
En la creación del mapa (Figura 5.14 (a)) se pueden destacar las técnicas basadas en Fourier como las más eficientes, exceptuando Fourier+PCA, pues el Análisis de Componentes Principales es un proceso computacionalmente pesado, multiplicando por 15 el tiempo necesario sobre la Firma de Fourier. La complejidad computacional de PCA también puede verse en PCA Rotaciones, ya que es, con diferencia, el algoritmo que más tiempo emplea en la construcción del mapa. Los descriptores HOG y GIST también requieren más tiempo que los basados en Fourier para construir el mapa, especialmente GIST-Color.

El tiempo de estimación de la pose (Figura 5.14 (b)) incluye el tiempo de cálculo del descriptor, estimación de la posición, y de la orientación.

HOG y los descriptores GIST presentan una distribución de tiempos similar a la obtenida durante la creación del mapa. Sin embargo, las técnicas basadas en Fourier muestran un tiempo de estimación de pose relativamente alto comparado con el tiempo de creación del mapa. Esto se debe a que la estimación de la orientación es computacionalmente complejo, lo que conlleva un aumento de tiempo en la estimación de la pose, especialmente en la Firma de Fourier. Sin embargo, como Fourier 1D únicamente emplea 3 componentes de fase, no se



(a)



(b)

Figura 5.14: Tiempo para (a) creación del mapa y (b) estimación de la pose.

ve tan afectado por este hecho.

Por contra, PCA Rotaciones pasa a ser uno de los algoritmos más rápidos Esto se debe a que para la estimación de la pose, ya no se realiza ningún análisis de vectores principales, sino que únicamente se proyecta la imagen de test sobre la base obtenida durante la construcción del mapa.

5.3.2.2 **Espacios de Color**

A continuación, se presentan los resultados de precisión en la estimación de la pose y requisitos computacionales de cada algoritmo usando la información de color.

Para ello, se ha aplicado cada descriptor a los canales RGB, HSV, a RGB y HSV conjuntamente, y también se estudia añadir la información color a través del Histograma de Color (HC), tal y como se ha descrito en la Sección 5.1.

Cabe destacar que la información del Histograma de Color ha sido añadida de forma distinta en cada método que utiliza PCA. En la Firma de Fourier+PCA se crea un descriptor que combina la Firma de Fourier y HC, y a continuación se aplica PCA. Sin embargo, esto no ha sido posible en el caso de PCA Rotaciones.

En el cálculo de la base de proyección de PCA Rotaciones, la matriz a partir de la cual se forma el producto interior utilizado para realizar la descomposición de los vectores principales debe estar compuesta por un vector y distintas rotaciones del mismo (Sección 4.2.4). Si añadimos el descriptor HC a las distintas rotaciones de la imagen, la base obtenida no presentaría rotaciones correctas de los vectores de información. Tampoco sería correcto rotar directamente el vector resultante de añadir HC a los píxeles de la imagen, pues no sería equivalente a una rotación real de la imagen.

Por lo tanto, no puede incluirse la información del Histograma de Color al realizar el análisis PCA de la base. Por ello, se obtiene primero las proyecciones de las imágenes en el nuevo espacio de proyección, y se añaden a dichas proyecciones el vector de HC.

Respecto a GIST-Color, se ha obtenido un descriptor para el espacio BN. Para ello, la comparación de los canales de color se reduce a la comparación multiescala de la imagen en escala de gris.

Se mantienen todos los parámetros de los distintos descriptores presentados en la Tabla 5.4. Para los Histogramas de Color, se utilizan 8 o 16 divisiones de las escenas panorámicas en celdas horizontales según descriptor, con 32 bins por histograma.

En esta comparación no se incluye el estudio de la estimación de fase. Independientemente del descriptor, la información relativa a la fase de la escena utiliza únicamente el espacio de color BN aunque para la localización se incluya algún otro espacio de color. Por ello, la precisión en el cálculo de la orientación no cambia al introducir la información de color.

En la Figura 5.15 podemos ver los resultados de precisión en la localización para usando la información de color. Para poder realizar una comparativa de forma sencilla, la información presentada se reduce a la precisión final del tercer vecino más cercano de cada método en los diferentes espacios de color.

De forma general, los descriptores presentan un aumento de precisión usando la información de color de la escena, exceptuando a Fourier 2D y PCA rotaciones sobre RGB+HSV, y GIST-Gabor cuando se añade HC. Es especialmente notable la mejora que se produce en Fourier 1D cuando se utiliza HSV, ya que triplica su precisión, pasando de 19% a un porcentaje del 57%. Sobre la imagen en escala de grises, este descriptor no presenta la suficiente capacidad de asociar correctamente las imágenes de la base y las de test como para considerarlo en aplicaciones reales de navegación. Sin embargo, esto cambia al aplicarlo sobre HSV.

También se puede destacar la alta precisión de HOG al introducir el Histograma de Color.

En el caso del espacio de color RGB, los resultados no muestran mejoras significativas en ningún descriptor, debido a la alta correlación entre los canales tres canales R, G y B. La excepción es GIST-Color. Como ya se ha comentado anteriormente, este descriptor está especialmente ideado para el espacio RGB, pues incluye la comparación de colores opuestos de Hering (Sección 4.4.3).

El Histograma de Color mejora los resultados de todos los descriptores respecto al espacio BN, siendo la mejor opción para introducir la información de color en los descriptores Fourier 2D, HOG y GIST-Gabor.

La aplicación de los descriptores sobre HSV presenta iguales o mejores resultados que hacerlo sobre el espacio RGB. La mejora es especialmente significativa en Fourier 1D, la Firma de Fourier y en PCA Rotaciones.

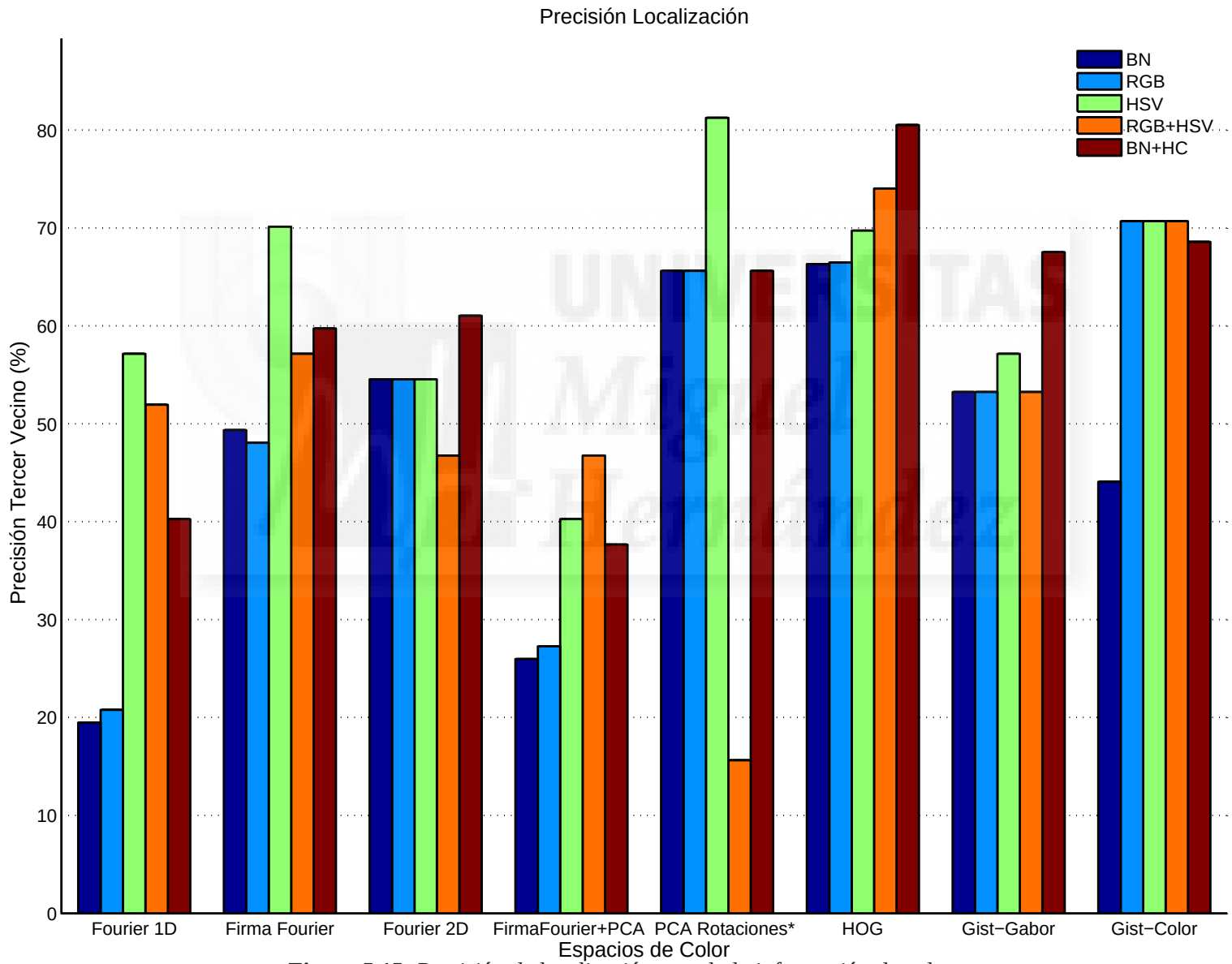


Figura 5.15: Precisión de localización usando la información de color.

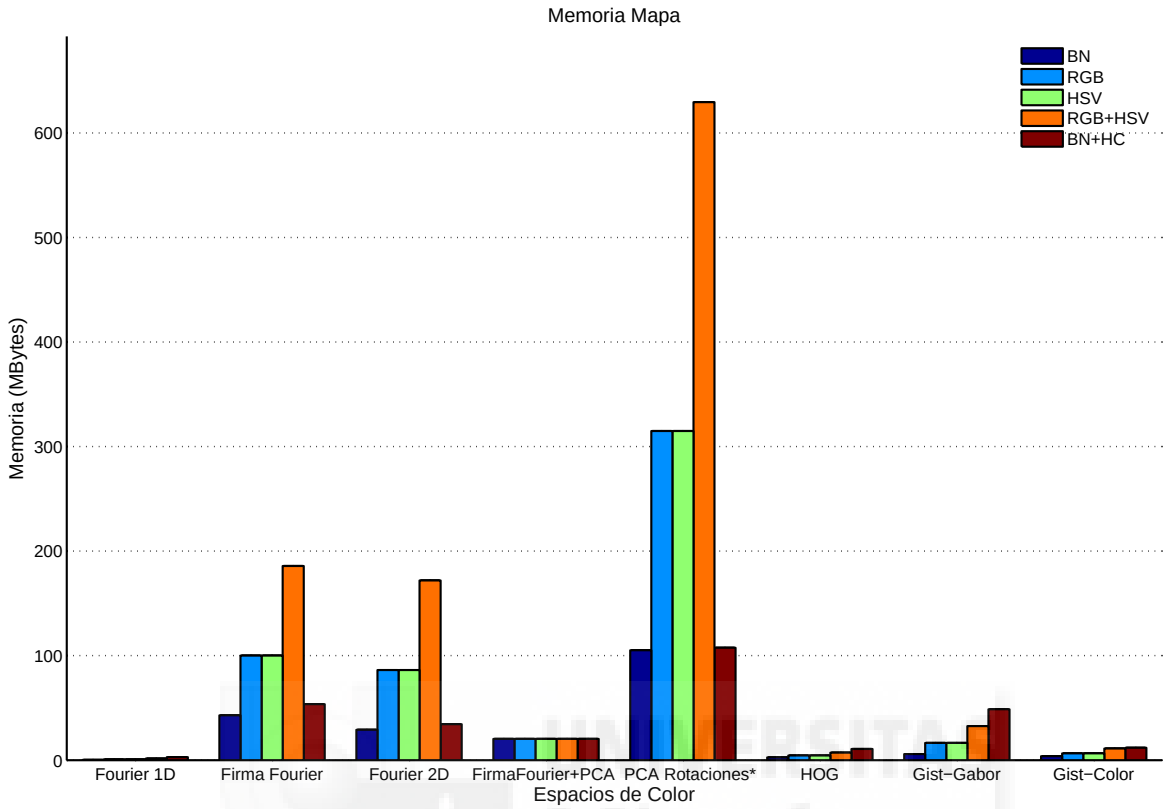


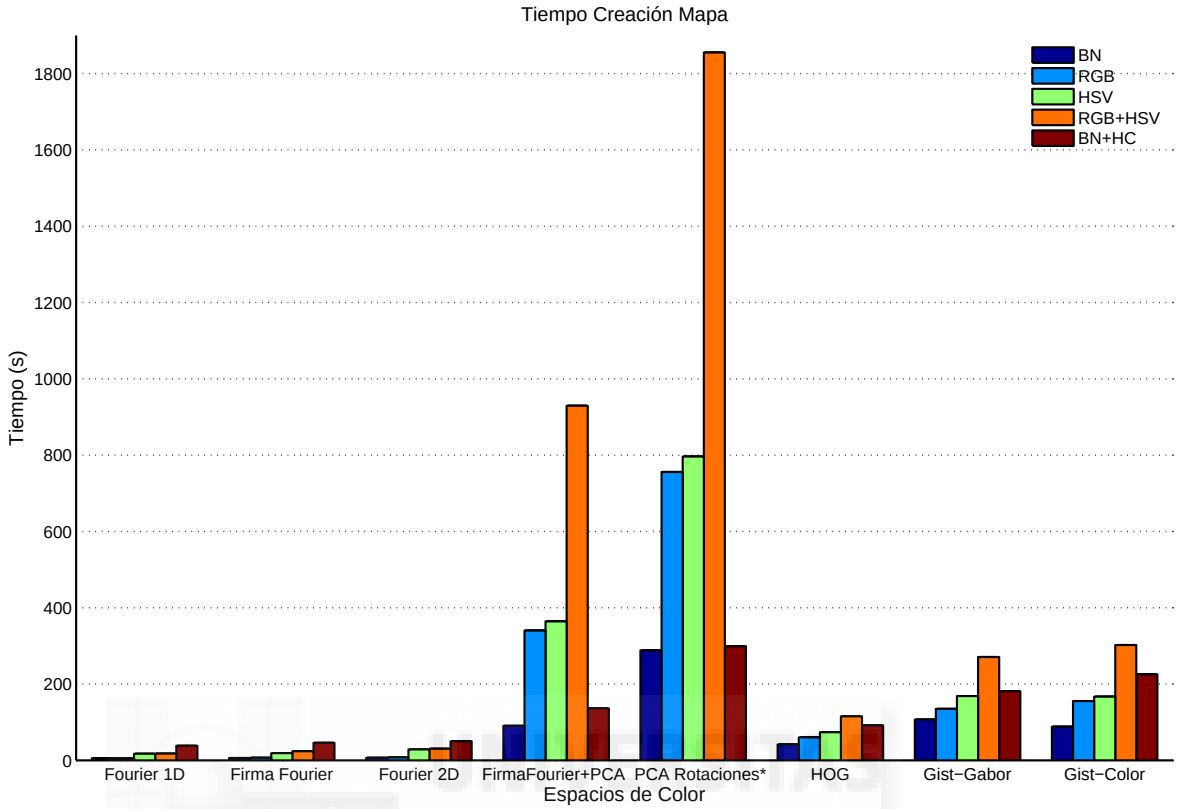
Figura 5.16: Memoria necesaria para almacenar el mapa usando la información de color

La memoria necesaria para almacenar el mapa con cada técnica se presenta en la gráfica de barras de la Figura 5.16. En general, la utilización de los espacios de color RGB o HSV **triplica** la memoria del mapa, mientras que el uso conjunto de ambos espacios de color multiplica por 6 su tamaño. En cuanto al histograma de color, añade una cantidad de memoria fija al mapa. El aumento relativo de memoria al añadir HC depende del descriptor

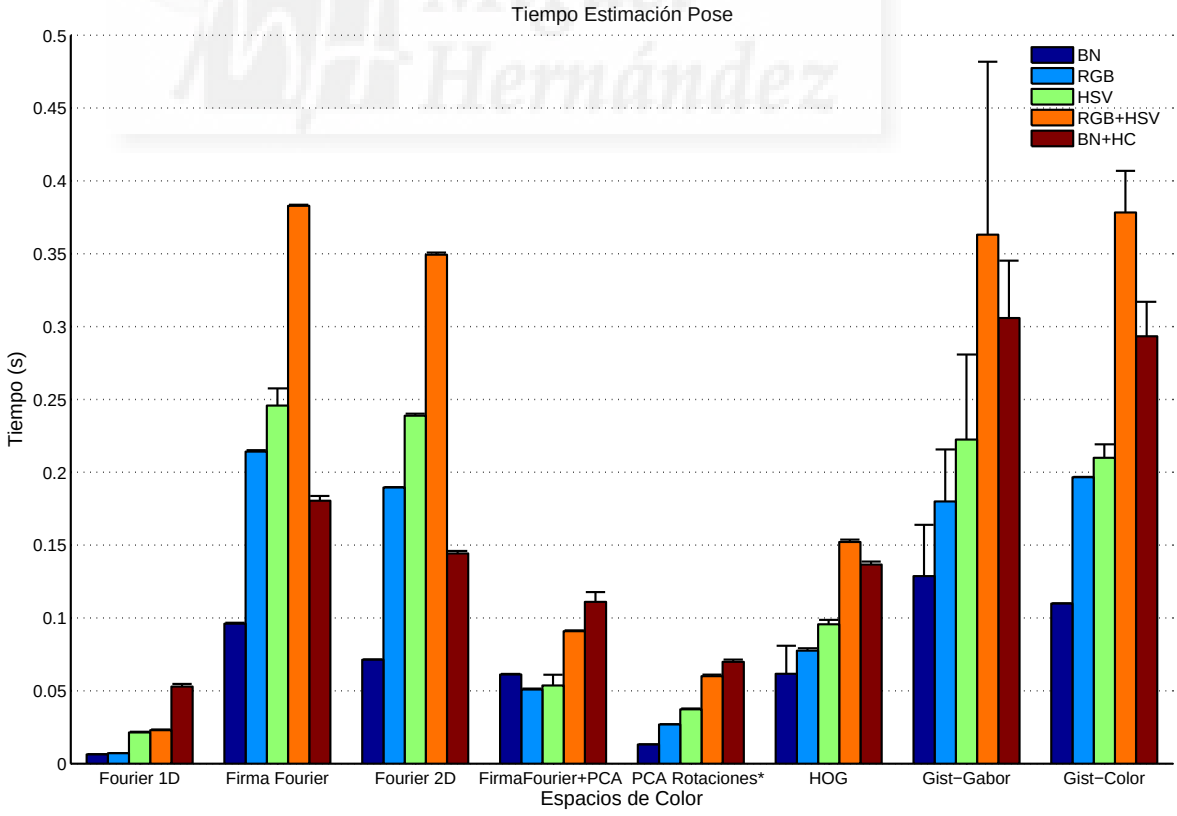
Nótese que Fourier+PCA es el único descriptor que no aumenta el tamaño del mapa al usar los distintos espacios de color. Esto es debido a que se aplica el Análisis PCA a toda la información que compone la base, y seleccionamos el mismo número de vectores principales para todos los casos.

La Figura 5.17 recoge el tiempo necesario para crear el mapa (Figura 5.17 (a)) y para estimar la pose de una imagen dentro del mapa (Figura 5.17 (b)).

En la construcción del mapa, la Firma de Fourier+PCA y PCA Rotaciones son los algoritmos que más tiempo necesitan, destacando el segundo. Pone de nuevo de manifiesto que el Análisis de Componentes Principales es un proceso computacionalmente muy costoso, sobre todo cuando aumenta el tamaño de la información a analizar, es decir, cuando usamos RGB+HSV. Por su lado, las necesidades de tiempo de las técnicas GIST superan a las basadas en Fourier y a HOG.



(a)



(b)

Figura 5.17: Tiempo usando la información de color para (a) creación del mapa y (b) estimación de la pose.

En la estimación de la pose, los métodos que usan PCA pasan a ser destacables por su bajo tiempo consumido. Excepto Fourier 1D, las técnicas basadas en Fourier igualan en tiempo a las basadas en GIST. Por su lado, HOG es un descriptor con un consumo relativamente bajo de tiempo de cálculo.

Comparando espacios de color, a medida que se incluyen más canales de color, el tiempo aumenta. Cuando aplicamos los descriptores sobre HSV, aunque el número de canales es el mismo que RGB, el tiempo es ligeramente mayor debido a la necesidad de transformar el espacio de color de la imagen de entrada.

La estimación del descriptor HC lleva un tiempo asociado de varía de 0.05 a 0.1 segundos aproximadamente según utilicemos 8 o 16 celdas horizontales por escena. Ese tiempo se añade al empleado por los descriptores sobre el espacio en escala de grises.

En el caso de Fourier 1D, la Firma de Fourier+PCA y PCA Rotaciones, añadir el Histograma de Color requiere más tiempo que emplear los otros espacios de color en la estimación de la pose. Para la Firma de Fourier y Fourier 2D, este tiempo queda por debajo del empleado al usar HSV o RGB, mientras que para HOG y GIST, este tiempo es inferior al necesario con RGB+HSV.

De forma general, la pérdida de precisión relativa cuando se introduce la información de color es mayor que cuando utilizamos el espacio BN. Sin embargo, los espacios HSV y BN+HC siguen siendo los que mejores resultados presentan para los distintos descriptores, exceptuando la Firma de Fourier+PCA, que obtiene los mejores resultados con la combinación de los espacios RGB+HSV.

5.3.2.3 **Comportamiento ante Ruido y Oclusiones**

Para completar el estudio, se incluyen los resultados de precisión en la localización y estimación del desfase cuando la imagen se ve afectada por ruido Gaussiano y distintas oclusiones. En la Figura 5.8 se incluye el ejemplo de una imagen con las distintas oclusiones y varianzas del ruido introducido en las imágenes de test.

La Figura 5.18 recoge la precisión de localización del tercer vecino para cada descriptor usando los distintos espacios de color y porcentaje de oclusión.

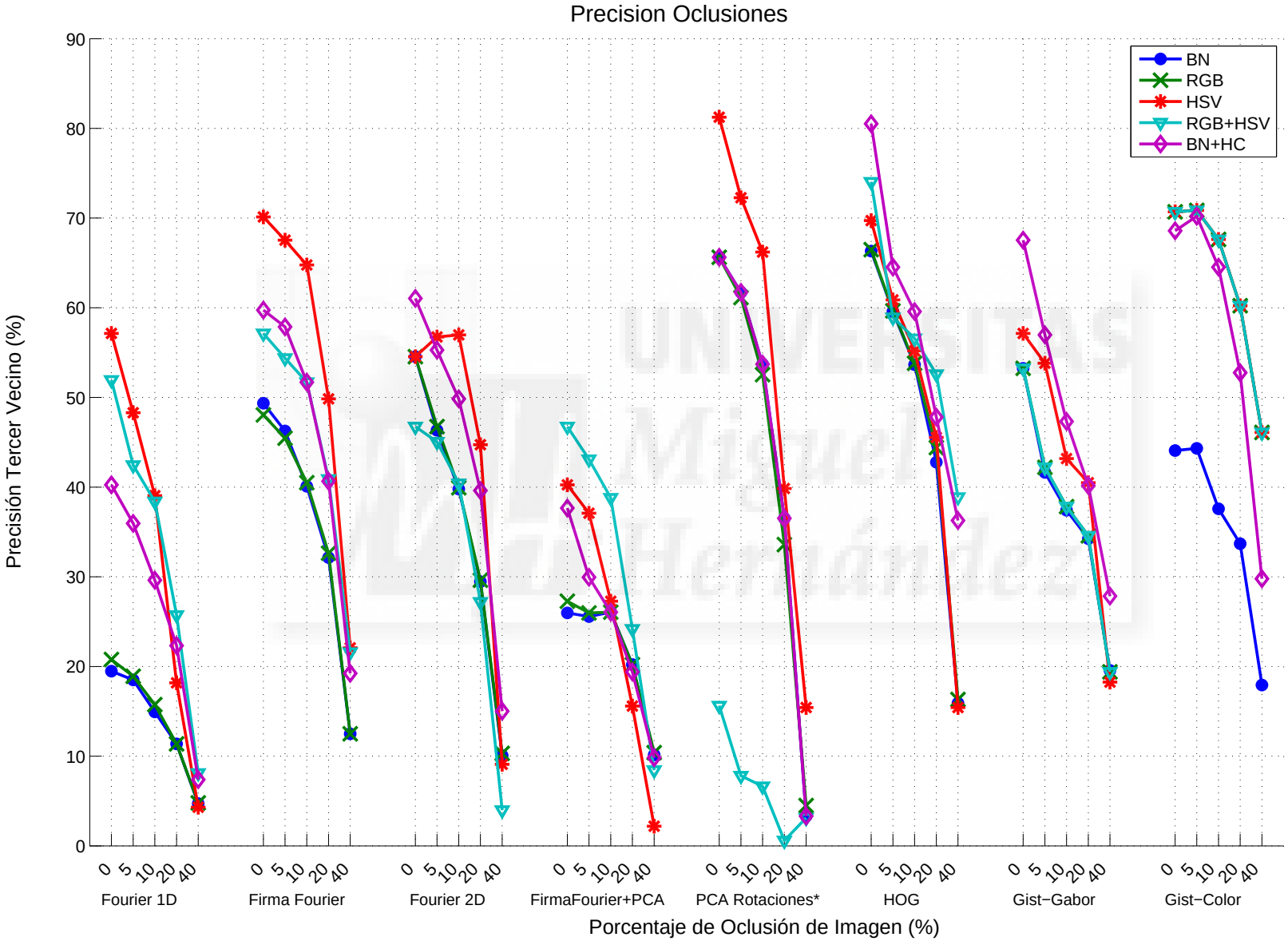


Figura 5.18: Precisión de localización usando la información de color ante oclusiones en las imágenes de test.

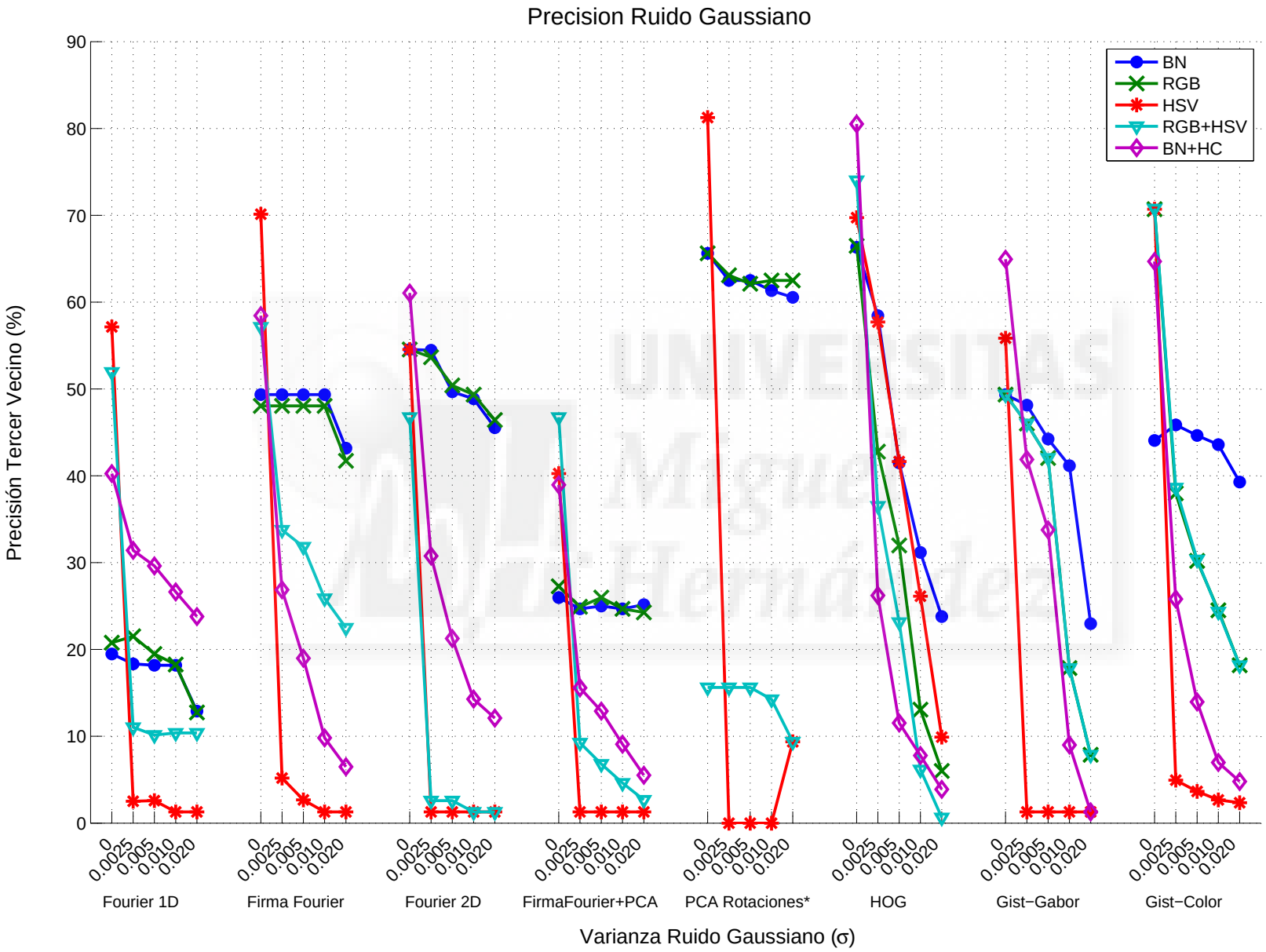


Figura 5.19: Precisión de localización usando la información de color ante ruido Gaussiano en las imagenes de test.

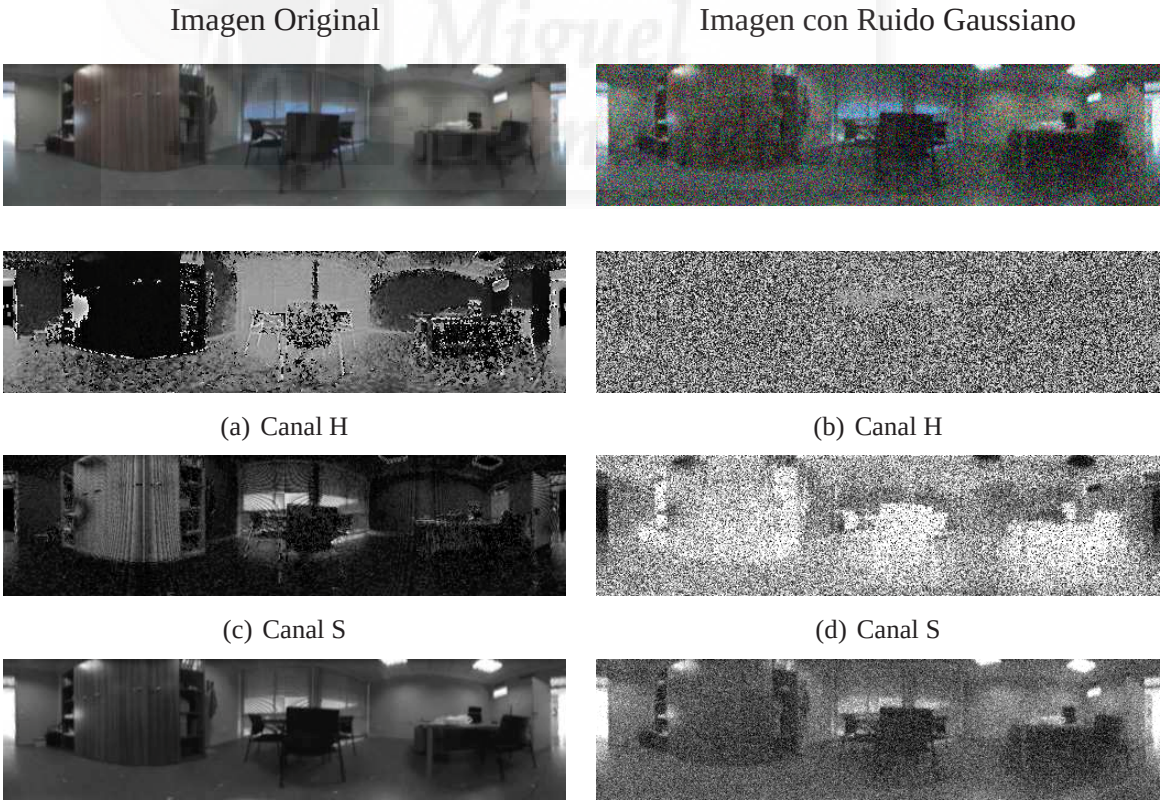
Los métodos de apariencia global que se ven más afectados por las oclusiones son las técnicas basadas en la Transformada de Fourier y PCA rotaciones, siendo éste último el más sensible.

HOG y los descriptores GIST son menos sensibles a la oclusión de la imagen, destacando la precisión de HOG sobre BN+HC, y GIST-Colot sobre RGB, HSV y la combinación RGB+HSV.

Por otro lado, la Figura 5.19 presenta los resultados de localización para el tercer vecino más cercano cuando las imágenes de test se ven afectadas por ruido Gaussiano.

En los resultados, podemos apreciar como el espacio HSV es especialmente sensible al ruido Gaussiano. Únicamente HOG muestra un comportamiento aceptable cuando se usa este espacio de color.

En las imágenes de test, el ruido Gaussiano se añade a los canales R, G y B de la escena original por separado. En la Figura 5.20 se muestran los canales H,S y V de una imagen de test sin ruido y con ruido Gaussiano con media nula y $\sigma = 0{,}0200$. Se puede aprecial claramente como los canales H y S están especialmente afectados por el ruido, siendo prácticamente imposible reconocer la imagen original.



(a) Canal H (b) Canal H

(c) Canal S (d) Canal S

(e) Canal V (f) Canal V

Figura 5.20: Imágenes de ejemplo incluyendo oclusiones y ruido Gaussiano.

El Histograma de Color muestra una reducción de su capacidad de descripción de la imagen al introducir ruido. Comparando BN y BN+HC, el Histograma de Color sólo mejora los resultados de localización de Fourier 1D. En el resto de descriptores, añadir HC supone reducir la precisión frente a BN.

Por lo tanto, los espacios de color que presentan una menor sensibilidad frente a ruido Gaussiano son la escala de grises (BN), y RGB, mientras que por descriptores, la Firma de Fourier, Fourier 2D PCA Rotaciones y la Firma de Fourier son lo que muestran mayor precisión.

Por último, se incluyen los resultados de la estimación de fase cuando las imágenes de test están afectadas por oclusiones (Figura 5.21 (a)) y por ruido Gaussiano (Figura 5.21 (b)). De nuevo, el error de fase se calcula únicamente sobre aquellas asociaciones de imágenes cuya distancia métrica máxima entre la posición de test y del mapa es igual o menor a 40cm. Además, como los métodos de cálculo de desfases son independientes al espacio de color, utilizamos el espacio BN.

Frente a las oclusiones, el aumento del error de fase es más notable en los descriptores basados en la Transformada de Fourier que en el resto de técnicas. Los resultados muestran que, para una oclusión del 40%, el error es el doble que para la imagen original. Es especialmente significativo el error de Fourier 1D, cuya varianza de error llega a los 45º.

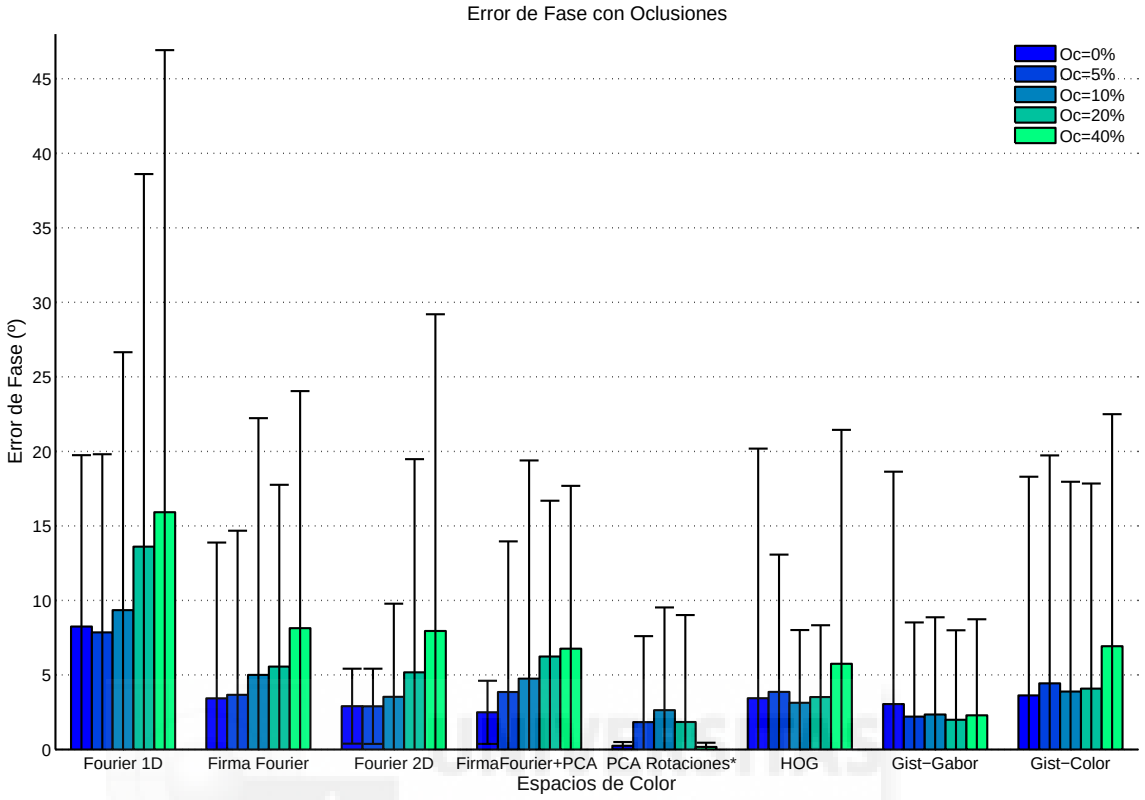
Por contra, PCA rotaciones es el descriptor que más precision muestra en la estimación de fase.

El error medio de GIST-Gabor queda por debajo de los 3º cualquiera que sea la oclusión de la imagen, mientras que HOG y GIST-Color presentan una precisión muy similar, con un error medio menor a 8º.

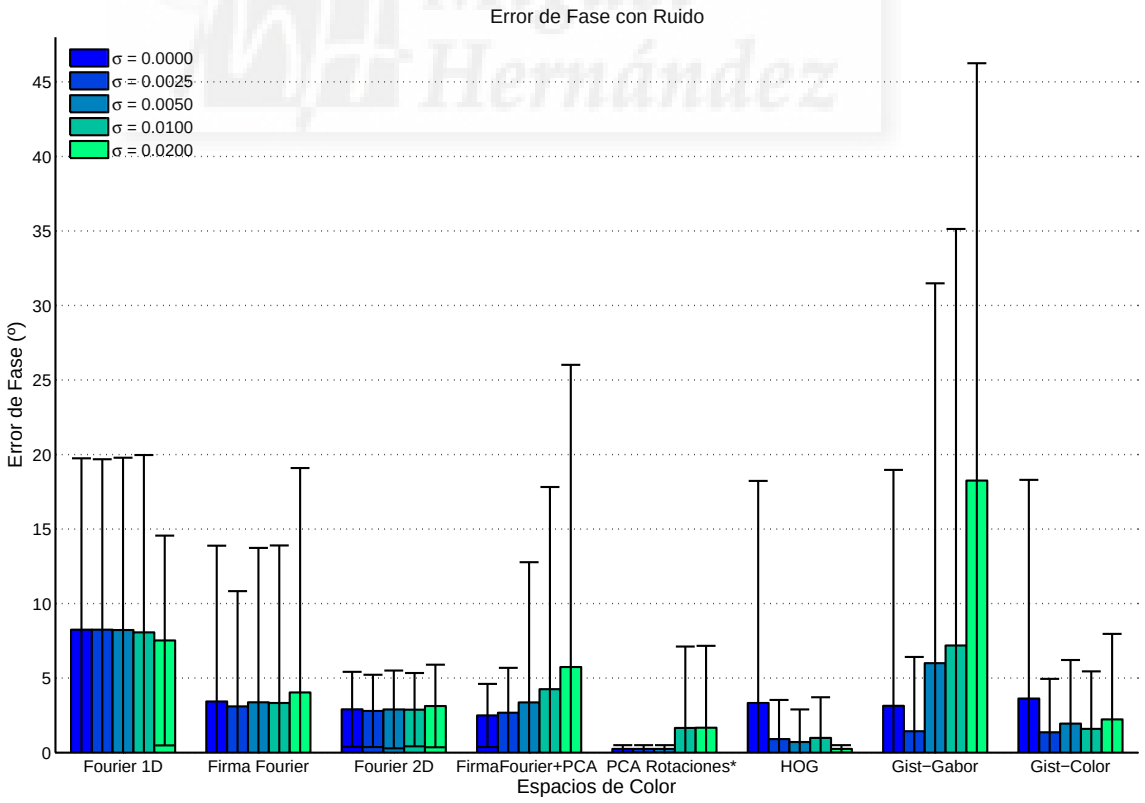
Cuando se introduce ruido Gaussiano en las escenas de test, la técnica más afectada en la estimación de fase es la de GIST-Gabor, y la Firma de Fourier+PCA. El error de fase de GIST-Gabor es especialmente elevado, multiplicando por 6 el error medio cuando la varianza del ruido es 0.0200.

Las técnicas basadas en Fourier (a excepción de Fourier+PCA) presentan poca variación al introducir el ruido en las escenas de test, manteniendo un error muy similar cualquiera que sea la el ruido introducido.

Por su parte, HOG y GIST-Color presentan un mejor comportamiento en la estimación de fase ante ruido que ante oclusiones.



(a)



(b)

Figura 5.21: Error en la estimación de fase ante (a) oclusiones y (b) ruido Gaussiano en las imágenes de test.

5.4 Conclusiones

En este capítulo se ha presentado la comparación de distintos descriptores basados en la apariencia global sobre imágenes panorámicas. Dicha comparación se ha centrado en el coste computacional y el requerimiento de memoria de los diferentes descriptores en la creación de un mapa denso, además del tiempo requerido y precisión en la estimación de la pose de un robot dentro de ese mapa.

El estudio se extiende con la introducción de la información de color usando distintos espacios y los Histogramas de Color. Además, también se comprueba la robustez de cada descriptor ante oclusiones y ruido de las imágenes de entrada.

Las principales conclusiones se presentan a continuación:

Precisión en Localización y Requerimientos Computacionales

- La utilización de la información de color mejora la precisión en la localización de forma general en todos los descriptores.
- El uso de HSV presenta mejores resultados que RGB, excepto cuando las imágenes están afectadas por ruido Gaussiano.
- Añadir el Histograma de Color mejora el porcentaje de localizaciones correctas frente al descriptor calculado sobre la imagen en escala de grises.
- A excepción de la Firma de Fourier y Fourier 2D, añadir la información utilizando el Histograma de Color tiene un mayor coste computacional que estimar los descriptores sobre RGB o HSV.
- Aunque HSV utiliza el mismo número de canales que RGB, la necesidad de calcular los nuevos canales de color se traduce en un aumento del tiempo para obtener el descriptor.
- El uso de RGB+HSV no supone una mejora notable de la localización frente a la utilización de un solo espacio de color (RGB o HSV), mientras que los requisitos de tiempo y memoria sí aumentan de forma significativa.
- Aunque PCA Rotaciones presenta una alta precisión en la estimación de la posición y orientación, los requerimientos computacionales lo hacen desaconsejable en la construcción de mapas densos. Además, junto al descriptor de Firma de Fourier+PCA, son los únicos descriptores que no permiten trabajar de forma incremental, ya que añadir una nueva imagen al mapa supone aplicar de nuevo el Análisis de Componentes Principales, y por tanto, volver a realizar la descomposición de la base entera.

- HOG presenta un buen compromiso entre precisión y requerimientos computacionales, especialmente cuando se completa el descriptor con el Histograma de Color. Igualmente, es un descriptor muy compacto.

- Fourier 1D es, con diferencia, el descriptor más compacto sobre cualquier espacio de color usado. Aunque la precisión de localización es baja cuando se utiliza la imagen en escala de grises y el espacio de color RGB, aumenta hasta al 58 % al emplear HSV. Se convierte en un descriptor interesante cuando existan limitaciones importantes de tiempo, y sobre todo, de memoria.

- La Firma de Fourier y Fourier 2D tienen un coste computacional muy reducido al construir el mapa. Sin embargo, el tamaño del mapa es alto comparado con el resto de descriptores, y también el tiempo necesario en la estimación de la pose, debido a que la estimación de la fase es poco eficiente. Es significativa la precisión de localización que se consigue al emplear la Firma de Fourier sobre HSV.

- Gist-Color muestra una precisión de localización mayor a Gist-Gabor, aunque también necesita más tiempo en las tareas de creación del mapa y localización.

- Los descriptores Gist requieren más tiempo que los basados en la transformada de Fourier al construir el mapa, pero su tamaño es menor.

Estimación de orientación

- Respecto al cálculo de la orientación, todos los descriptores presentan un error medio inferior a 8^o cuando la imagen de entrada no está afectada por ruido ni por oclusiones.

- La precisión de fase es menor para los descriptores basados en la Transformada de Fourier que en el resto de descriptores.

- Sin embargo, hay que destacar que en los descriptores Gist, HOG y PCA Rotaciones, la resolución angular depende directamente de la información incluida en el descriptor. Si se desea aumentar dicha resolución, lo hará también el tamaño del descriptor, y el coste computacional de creación del descriptor y estimación de la fase. Por lo tanto, la estimación de fase es menos flexible que en el caso de los descriptores basados en Fourier.

Oclusiones

- Cuando las imágenes de test presentan oclusiones, se produce una reducción de la precisión en la localización.

- En general, el efecto de las oclusiones es más notable en los espacios de color que usando las imágenes en escala de grises.
- No obstante, los mejores resultados de localización siguen obteniéndose al usar el espacio HSV y el Histograma de Color.
- Por descriptores, Gist y HOG son los que menos se ven afectados por las oclusiones, destacando HOG sobre BN+HC, y Gist-Color sobre RGB o HSV.
- En cuanto a la estimación de fase, los descriptores de Fourier son los menos robustos al introducir las oclusiones, especialmente Fourier 1D. Aún así, exceptuando este descriptor, el error medio de fase se mantienen igual o inferior a 8^o, aunque el aumento en la varianza sí sea notable.

Ruido Gaussiano

- El ruido Gaussiano afecta notablemente a los canales *Hue* y *Saturation* del espacio HSV. Esto conlleva una pérdida significativa de precisión en la localización al usar los espacios de color HSV y RGB+HSV.
- Los espacios de color que menos se ven afectados por el ruido son BN y RGB.
- Pueden destacarse los descriptores basados en la Transformada de Fourier, PCA Rotaciones y Gist-Color como los más robustos ante ruido.
- En la estimación de fase, únicamente la Firma de Fourier + PCA y Gist-Gabor muestran un aumento claro del error.

CAPÍTULO

6

Análisis Multiescala en Tareas de Navegación Topológica

Este capítulo presenta un método de estimación de desplazamiento relativo entre escenas. Este algoritmo, que aparece bajo el nombre Análisis Multiescala, puede ser utilizado para la mejora de la localización y la construcción de mapas topológicos usando la apariencia global de las imágenes en aplicaciones de navegación robóticas.

El Análisis Multiescala se basa en la introducción de *zooming* o ampliaciones artificiales entre imágenes para, mediante la asociación de las distintas ampliaciones de las escenas, estimar su posición relativa.

Como aportación de este capítulo, se va a presentar su aplicación sobre imágenes campo de visión amplio capturadas con una cámara con lente de ojo de pez, y sobre imágenes omnidireccionales capturadas con un sistema de visión catadióptrico.

En el primer caso, el Análisis Multiescala se usará para construir un mapa topológico compuesto por nodos de imágenes capturados en distintas zonas del área de navegación. Se hará uso de rutas de imágenes tomadas a lo largo de caminos que unen dichos nodos para seleccionar los nodos y determinar su distribución espacial, además de sus relaciones de adyacencia. Posteriormente, aprovecharemos ese mismo mapa para localizar la trayectoria seguidas por las rutas. El Análisis Multiescala permitirá llevar a cabo una localización topológica no sólo en las posiciones de los nodos, sino también en posiciones intermedias.

Se introduce también una comparación de distintas técnicas de apariencia global aplicadas sobre imágenes no omnidireccionales para obtener información de precisión en localización y requisitos computacionales.

Tras ello, extenderemos este sistema de estimación de desplazamiento topológico a imágenes omnidireccionales. Combinaremos el uso del Análisis Multiescala con la información proporcionada por los descriptores basados en apariencia global sobre imágenes panorámicas descritos en el Capítulo 4. De esa forma, obtendremos información sobre la dirección y magnitud del desplazamiento, por lo que se puede construir un sistema de odometría visual topológico.

Este sistema de odometría visual será aplicado dentro de un algoritmo de navegación que permitirá el reconocimiento de zonas por las que anteriormente se ha navegado para mejorar la precisión del mapa construido mediante cierre de bucle.

Por último, se estudia el desempeño de este sistema de navegación topológico utilizando una ruta de imágenes capturadas en entornos reales con cambios de iluminación constantes.

Por tanto, como aportaciones adicionales de esta capítulo, podemos destacar el estudio de sobre descriptores de apariencia global aplicado a imágenes no omnidireccionales, incluyendo la variación de su resolución, y el sistema de odometría visual topológico aplicado a la estimación de rutas que incluyen cierres de bucle.

6.1 Construcción de Mapas y Localización usando el Análisis Multiescala sobre Imágenes Proyectivas

En esta sección se presenta un algoritmo de construcción de mapas y localización topológicos basado en la apariencia global de imágenes no panorámicas.

En concreto, el sistema de visión utilizado es una cámara *GoPro*. Esta cámara está provista de una lente de ojo de pez, por lo que las imágenes presentan distorsión radial (Sección 3.3).

Debido a que los descriptores basados en la apariencia global no pueden aplicarse directamente a imágenes distorsionadas, es necesario corregir las imágenes obtenidas. En la Sección 3.3 se incluye más información de la cámara, del proceso de calibración y corrección de la distorsión de las escenas.

Primero, se propone un algoritmo para la construcción del mapa topológico que describirá el entorno. Dicho mapa se compone de nodos que representan un punto del espacio, y de ejes que denotan conectividad entre los distintos nodos.

Cada nodo está compuesto por un conjunto de 8 imágenes capturadas con un desfase angular de 45^o entre sí, cubriendo el campo de visión completo en una cierta posición.

Usando la información de rutas adquiridas a lo largo del área de navegación, se lleva a cabo la selección de los nodos mediante asociación de las imágenes. Además, gracias al análisis multiescala que se presenta a continación, es posible estimar su posición relativa, permitiendo la construcción del grafo que representa el mapa.

Una vez construido el mapa, se propone un sistema de estimación de la trayectoria de rutas. Este sistema es capaz de localizar al robot no sólo en la posición de los nodos, sino también en posiciones intermedias.

Tanto en la construcción del mapa como en la estimación de la trayectoria, el algoritmo utiliza la información proporcionada por el análisis multiescala en la asociación de imágenes. Este análisis multiescala permite mejorar la asociación entre imágenes y obtener un indicador de la posición relativa entre escenas.

En el apartado de Experimentos y Resultados se incluye una comparación reducida de descriptores basados en apariencia global aplicados a imágenes proyectivas. Estos resultados nos permitirán seleccionar la técnica utilizada para describir las escenas en las tareas de creación del mapa y localización. Además, debido a que el tamaño original de la imagen es de 1004×1817 píxeles, el coste computacional asociado a la obtención del descriptor es alto cualquiera que sea la técnica. Por ello, se estudia también la reducción de la resolución de las imágenes utilizadas.

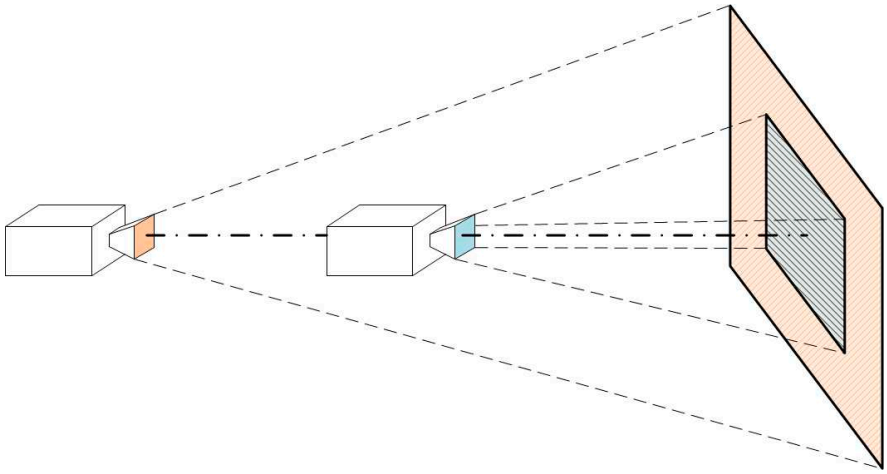


Figura 6.1: Representación de escena capturada por un sistema de visión considerando un desplazamiento perpendicular al plano de proyección.

6.1.1 Análisis Multiescala

El análisis multiescala tiene como objetivo mejorar la asociación entre escenas, y hallar la posición relativa entre dos imágenes haciendo uso de los descriptores de apariencia global.

Durante la asociación entre las imágenes de los nodos (que constituyen la base) y las de ruta, es posible que los nodos se encuentren muy separados para poder obtener una asociación correcta. Esto es especialmente notable en los puntos intermedios entre nodos, donde la apariencia entre las escenas presenta la menor similitud.

Por ello, se propone realizar un análisis basado en una ampliación multiescala para mejorar la apariencia entre las escenas comparadas.

En la Figura 6.1 es posible ver la escena recogida por una misma cámara cuando se desplaza perpendicularmente al plano de proyección. Se puede apreciar que la escena correspondiente a la posición más adelantada, representada en azul, se corresponde con la zona central de la escena naranja, que es la asociada con la posición más atrasada.

Luego si realizamos una ampliación de la zona central de la imagen naranja, la imagen resultante aumenta su apariencia respecto a la escena azul. Esta ampliación se realiza mediante la selección de la región central de la escena y re-escalado al tamaño original de la imagen, simulando un zoom digital.

El factor de escala s es el cociente entre el tamaño original de la escena y el tamaño de la región seleccionada. Suponiendo que nuestra imagen original tiene una resolución de 32×64, una escala $s = 2$ supone seleccionar los 16×32 píxeles centrales.

El aumento de apariencia entre proyecciones conlleva una mejora en la asociación entre imágenes. La Figura 6.2 recoge las curvas *Recall* − *Precision* considerando el vecino más cercano (N.N.) en la asociación entre las imágenes de las rutas y los nodos. En este caso,

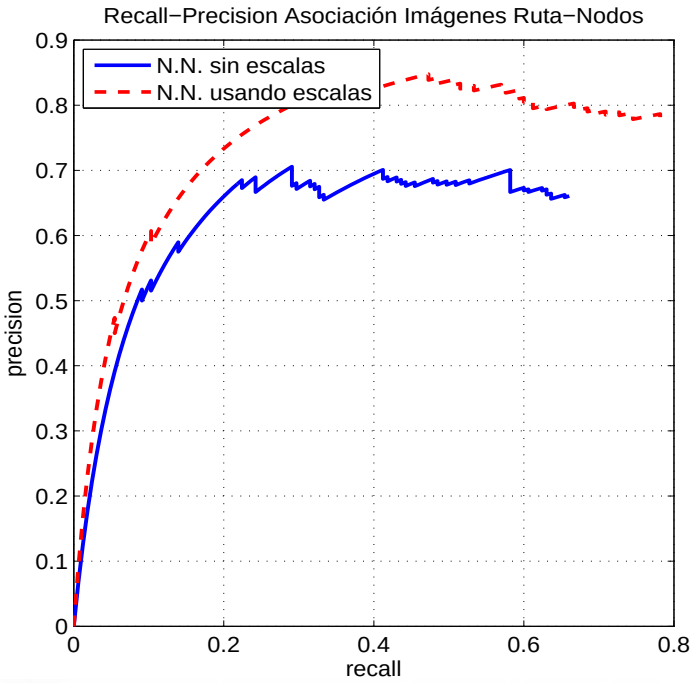


Figura 6.2: Comparación Recall-Precision en precisión de asociación de imágenes al introducir análisis multiescala, utilizando HOG sobres escenas de resolución 32×64 píxeles.

podemos ver que al utilizar el zoom aumentamos la precisión final en un 14%. Estos resultados han sido obtenidos utilizando dos rutas de imágenes distintas, compuestas 172 escenas, e imágenes de nodos que componen el mapa del entorno. Se realiza la asociación entre las imágenes de las rutas y los nodos, utilizando como descriptor HOG. Si la imagen del mapa asociada con la imagen de la ruta corresponde con el nodo más cercano a su posición en el plano del suelo, se considera acierto. En caso contrario, se considera una asociación errónea.

En la Figura 6.3 se presenta un ejemplo que incluye dos imágenes correspondientes a dos nodos distintos, y otras dos de ruta asociadas cada una a uno de los nodos. En el caso del Nodo 1, la imagen de la ruta (escena (b)) se encuentra por delante de la imagen de nodo (Figura 6.3 (a)). Por otro lado, la escena (a') es una ampliación de la parte central de (a). Comparando la imagen original del nodo y la ampliada con (b), se puede comprobar que la apariencia más similar se obtiene entre la imagen (a') y (b), es decir, entre la imagen de nodo ampliada y la imagen de la ruta.

En el caso del Nodo 2, la imagen de la ruta (Figura 6.3 (c)) se encuentra situada por detrás de la posición del nodo. Siendo (c') una ampliación de la escena de la ruta, es posible observar gráficamente que al realizar el zoom de la escena de la ruta se aumenta la apariencia con la imagen del nodo (Figura 6.3 (d)).

Para ver cómo se traduce el aumento de apariencia en la distancia imagen cuando se

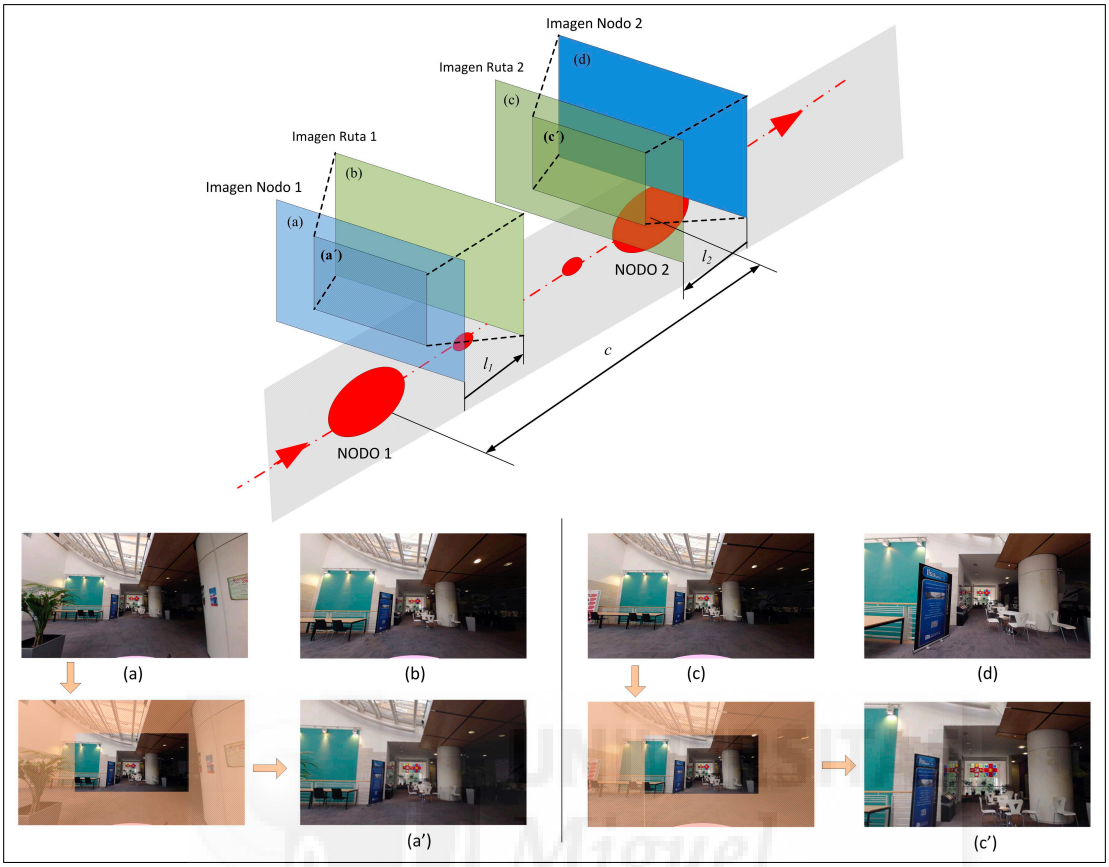
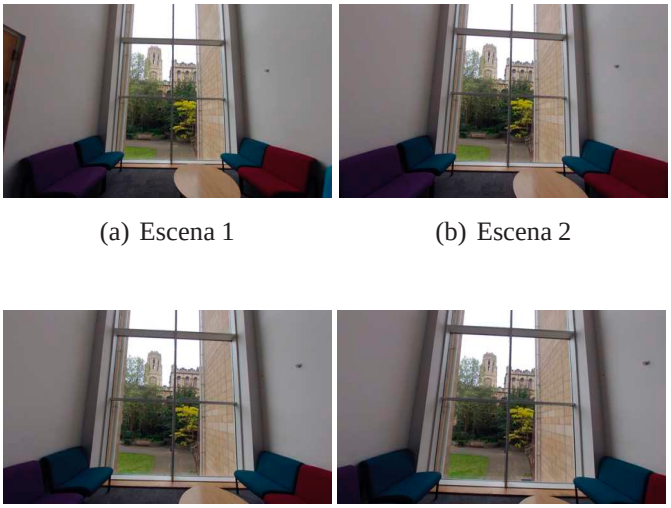


Figura 6.3: Asociación entre imágenes usando el Análisis Multiescala. (a) Imagen del Nodo1, (a') ampliación de imagen (a), y (b) imagen de ruta localizada delante del Nodo1. (c) Imagen de ruta localizada tras el Nodo2, (c') ampliación de imagen (c), y (d) imagen del Nodo2. l_1 y l_2 representan distancias topológicas entre las imágenes de ruta y el nodo más cercano, y c la distancia entre nodos.

introduce el análisis de escalas, la Figura 6.4 incluye 4 ejemplos consecutivas de una ruta. Como se puede apreciar, existe un movimiento de avance a medida que se adquieren las distintas escenas. La gráfica de la figura recoge la distancia imagen entre distintas escalas de la Escena 1 con el resto de imágenes. El mínimo en la curva de distancia imagen denota la escala en la que se obtiene la asociación más similar. Como era de esperar, el mínimo entre la propia escena y las distintas escalas se produce en una escala igual a la unidad, es decir, cuando no se realiza ninguna ampliación. Respecto a las otras escenas, el mínimo se obtiene a escalas mayores a medida que nos alejamos geométricamente de la Escena 1.

De esta forma, a través del análisis multiescala es posible obtener un indicador de la distancia relativa entre dos escenas.



(a) Escena 1 (b) Escena 2

(c) Escena 3 (d) Escena 4

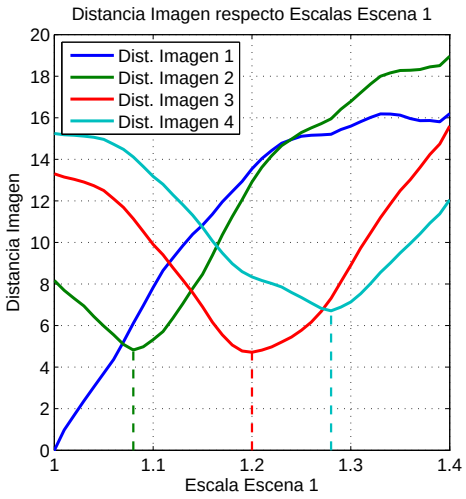


Figura 6.4: Escenas consecutivas de una ruta, y distancia imagen de las escenas respecto a distintas escalas de la Escena 1.

6.1.1.1 Estimación de la posición relativa entre dos escenas usando Análisis Multiescala

En el proceso de asociación de las imágenes de los nodos y de las rutas se emplean distintas escalas de ambos conjuntos de imágenes, es decir, se lleva a cabo la comparación de distintas escalas de la escena de ruta con distintas ampliaciones de las imágenes de los nodos.

Tras esa comparación, el algoritmo selecciona la asociación con menor distancia imagen. Se define s^n y s^r como los valores específicos de escala de la imagen del nodo y de la ruta de dicha asociación para los cuales obtenemos el mínimo.

Estas escalas nos permiten determinar la posición relativa de ambas imágenes. La distancia topológica se define como:

$$l = s^n - s^r. \tag{6.1}$$

Siguiendo el ejemplo de la Figura 6.3, cuando la imagen de ruta se encuentra por delante de la imagen del nodo (ejemplo del nodo 1), la mínima distancia imagen se produce cuando se amplía la imagen del nodo y se mantiene la de ruta. Por lo tanto, $s^n > s^r$, obteniendo una distancia topológica $l > 0$. Por contra, si la imagen de la ruta se encuentra por detrás del nodo, la mínima distancia imagen se obtiene al realizar un zoom de la imagen de la ruta y manteniendo la del nodo. En ese caso, $s^r > s^n$, con lo que $l < 0$.

Esto significa que la distancia topológica l no sólo proporciona información de la distancia relativa entre imágenes, sino también de la dirección de esa distancia.

En la Figura 6.5 se muestran los resultados de asociación entre imágenes de dos nodos y distintas escenas de una ruta que une ambos nodos. Los resultados incluyen el nodo más cercano n, las escalas de la imagen del nodo s^n y de la ruta s^r, y la distancia topológica l. En

este caso, se ha obviado el cálculo de la fase θ, pues el ejemplo se ha simplificado incluyendo una única orientación.

Como se puede comprobar en los resultados, para posiciones de las imágenes de ruta por detrás de los nodos, la distancia topológica l es negativa. Por contra, cuando la imagen de ruta se asocia a un nodo cuya posición métrica es anterior, la distancia topológica es positiva.

Por tanto, con la comparación de las escalas de las distintas escenas se logra obtener una medida del desplazamiento entre imágenes y de la dirección de dicho desplazamiento.

6.1.2 Construcción del Mapa

Cuando se construye un mapa topológico basado en apariencia, se almacenan en los nodos información a medida que se produce una navegación por un entorno determinado. El problema que es preciso considerar y analizar en este apartado consiste en establecer una conexión topológica entre los diferentes nodos que se van incorporando al mapa que está siendo elaborado. Además será necesario estudiar cuándo se deben incorporar nuevos nodos al mapa que se está elaborando.

Por tanto, la construcción del mapa consiste en la selección de nodos repartidos por el espacio de navegación y su conexión topológica. Para ello, se hace uso de rutas de imágenes capturadas a lo largo del entorno.

Conviene recordar que el algoritmo no dispone al inicio de los experimentos información relativa a la localización espacial de los nodos.

Con el uso del análisis multiescala se pretende mejorar la asociación de imágenes entre los nodos y las rutas, y además obtener un grafo que represente la distribución espacial real de los nodos.

6.1.2.1 Asociación de las Imágenes de Nodos y Rutas

A continuación se presenta el proceso de asociación de imágenes que lleva a cabo el algoritmo:

- Primero, se crea la base que contiene las información de los nodos. Para ello, el algoritmo calcula los descriptores $z^n \in \Re^{1\times y}$ del conjunto de imágenes de los nodos incluyendo distintas escalas, siendo y el número de componentes del descriptor.
- Los descriptores se almacenan por columnas en la matriz $\mathbf{Z} = [\, z_1^n \quad z_2^n \ldots z_i^n \ldots z_m^n \,]$, siendo m el número de escenas incluidas en la base. Nótese que el número de escenas que componen la base final es igual al producto del número de nodos, el número de orientaciones por nodo, y las distintas escalas por imagen. Esta matriz $\mathbf{Z}$ constituye la base de información del mapa.

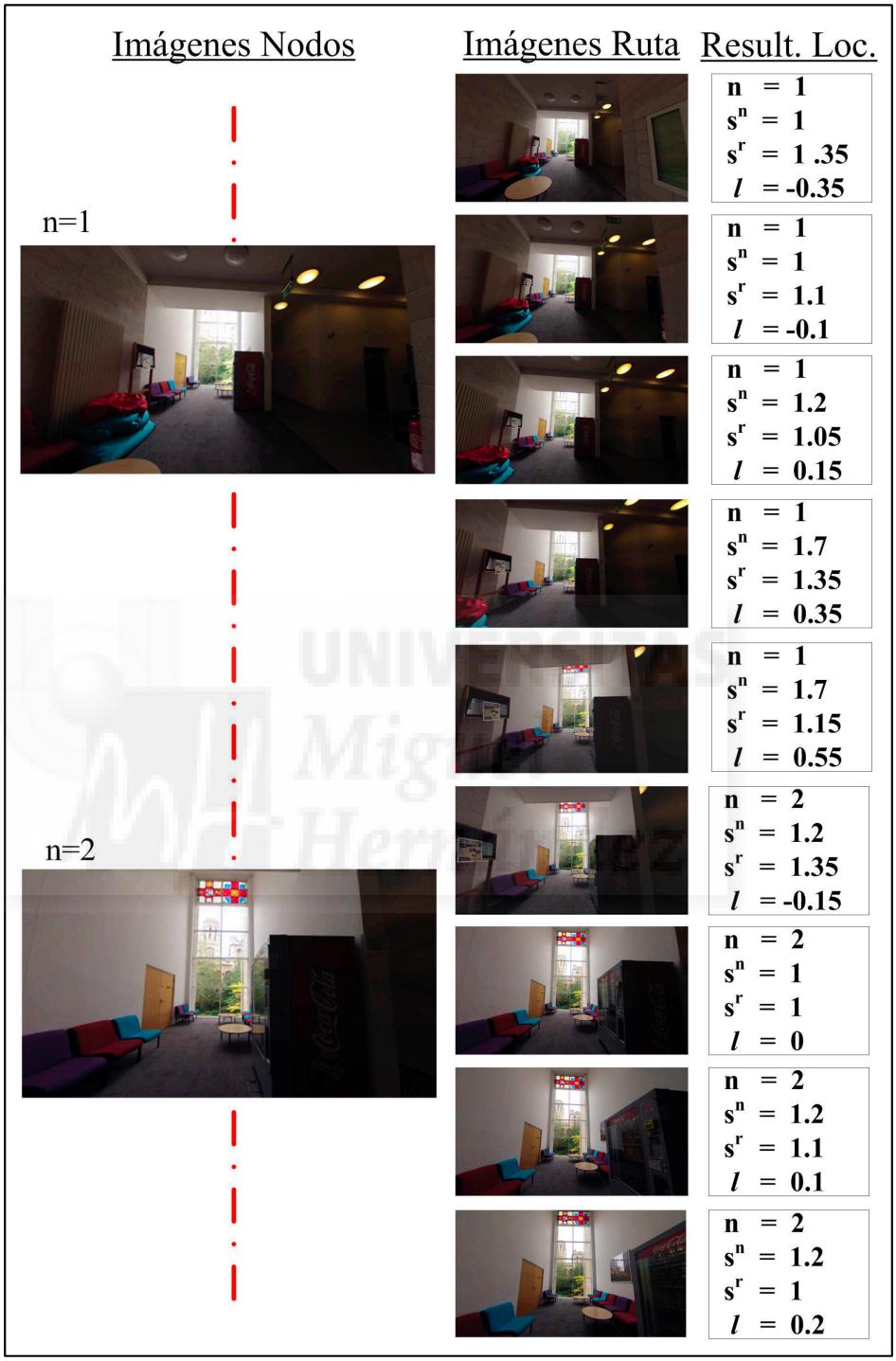


Figura 6.5: Ejemplo de asociación de imágenes usando dos imágenes de nodo y nueve escenas de ruta que conectan ambos nodos. En la parte derecha, los resultados de asociación incluyen el nodo más cercano (n), la escala de la escena de nodo (s^n), la escala de la escena de ruta (s^r) y la distancia topológica (l).

- Como los descriptores de las escenas se almacenan con un orden conocido, es posible asociar el nodo, la orientación en el nodo θ y el factor de escala correspondiente a un descriptor de la base, ya que es función de la columna de **Z** en la que está almacenado:

$$[n, \theta, s^n] = f(i). \tag{6.2}$$

  Cabe resaltar que el orden en que se almacenan los nodos no proporciona información de su distribución espacial. La posición de los nodos es totalmente desconocida para el sistema al comenzar el algoritmo. Es decir, el mapa inicial comienza con una colección de nodos de los que únicamente se conoce su apariencia a través de los descriptores. Sin embargo, no se dispone de información sobre su orden de aparición en el mapa ni su pose en el entorno.

- Cuando llega una nueva imagen de la ruta, primero se obtiene su descriptor z^r y se calcula la distancia imagen (d) con todas las escenas incluidas en la base **Z**. La distancia imagen se define como la distancia Euclídea entre descriptores:

$$d_i^r = \sqrt{\sum_{a=1}^{y} (z_{i,a}^n - z_a^r)^2}, \qquad i = 1 \ldots m. \tag{6.3}$$

- d_i^r se utiliza como clasificador. El algoritmo selecciona el vecino más cercano, y asocia a la distancia imagen correspondiente los valores de n, θ y s^n de la imagen de la base.

- El algoritmo repite este proceso para distintas escalas de la image de ruta (s^r).

- Una vez se obtienen las asociaciones para las diferentes escalas s^r, se vuelven a ordenar los resultados en función de la distancia imagen, y seleccionamos el vecino más cercano.

- Finalmente, de cada imagen de ruta obtenemos el siguiente vector de información de la asociación seleccionada:

$$[n \quad d \quad \theta \quad s^n \quad s^r] \tag{6.4}$$

6.1.2.2 **Construcción del Grafo**

La construcción del grafo emplea como información de partida los vectores obtenidos con las asociaciones de las imágenes de las rutas y del mapa (Ecuación 6.4). Primero, el algoritmo debe decidir cuándo se introduce un nuevo nodo en el mapa. Tras ello, se estima su posición en el grafo a partir de las distancias topológicas y de la orientación de las imágenes de las rutas.

Los vectores de información de asociación se van almacenando en una matriz a medida que llega una nueva imagen de ruta. La decisión de incluir un nuevo nodo tiene en consideración esta información.

Concretamente, comprobamos el número de veces (M) que se repite el nodo moda (n_m) de las últimas 5 asociaciones, y calculamos media (μ) y la desviación estándar (σ) de las distancias imágenes (d) de todas las asociaciones llevadas a cabo hasta ese momento.

El nodo n_m es incluido en el grafo si se cumple una de las siguientes condiciones:

- $M \leq 3$
- $M = 2 \quad$ y $\quad d_{n_m} < \mu - \sigma$

De esa forma, sólo se incluye un nuevo nodo si ha sido seleccionado repetidamente, o si la asociación entre las imágenes de la ruta y del nodo es muy fiable (es decir, cuando la distancia imagen es pequeña).

La elección de los valores que sirven de umbral para añadir un nuevo nodo han sido determinados experimentalmente. El objetivo de estas condiciones es que la inclusión de un nodo no se haga hasta que se haya asociado un mínimo número de veces con las escenas de la ruta. Con ello, aumentamos la certeza de estar seleccionando un nodo correcto, reduciendo los problemas derivados de falsas asociaciones entre la ruta y los nodos. En el caso, por ejemplo, de que las imágenes de la ruta fuesen capturadas con mayor frecuencia, es decir, cada menos centímetros, podríamos considerar unos umbrales más elevados. Sin embargo, unos valores excesivamente altos puede derivar en la no inclusión de un nodo correctamente detectado.

Cuando un vector de información tiene un valor $d > \mu + 2\sigma$, no se tiene en cuenta para tomar la decisión de incluir un nuevo nodo, ya que la asociación tiene una fiabilidad muy baja.

Para conocer las conexiones entre nodos, se crea una matriz de adyacencia $A \in \Re^{NxN}$, con N el número de nodos del mapa. A es una matriz dispersa cuyas filas y columnas se corresponden con los índices de los nodos almacenados en la base $\mathbf{Z}$. Sus valores son binarios, denotando con 1 la existencia de nodos adyacentes. Suponiendo que se ha detectado el nodo n_1 del mapa, y el siguiente nodo encontrado es el n_2, entonces $A_{n_1,n_2} = 1$.

Una vez detectado el nodo, se calculan sus coordenadas en el mapa topológico. Para ello, se estima su distancia topológica y desfase respecto al nodo anterior.

En relación al cálculo de la distancia topológica entre nodos del grafo, se emplea de nuevo la información proporcionada por los factores de escala. Considerando n_i y n_{i+1} dos nodos consecutivos, y una ruta de imágenes que une dichos nodos en dirección $n_i \rightarrow n_{i+1}$, se

consideran las distancia topológicas de la última imagen de ruta donde ha sido detectado n_i ($l^l_{n_i}$), y la primera donde se asocia n_{i+1} ($l^f_{n_{i+1}}$).

La distancia topológica (c) entre esos dos nodos consecutivos se define como:

$$c_{n_i,n_{i+1}} = l^l_{n_i} - l^f_{n_i+1} \tag{6.5}$$

Es importante remarcar que la Ecuación 6.5 tiene en consideración los signos de las distancias topológicas.

Siguiendo el ejemplo de la Figura 6.5, l^l_1 corresponde con la distancia topológica obtenida en la quinta imagen, que es la última en la que el nodo 1 es detectado, y l^f_2 con la distancia topológica de la sexta escena (la primera en la que se asocia el nodo 2). Entonces, $c_{1,2} = 0{,}55 - (-0{,}15) = 0{,}7$.

Para construir el grafo, también es necesario conoce el desfase producido en un nodo. En este caso supondremos que los cambios de dirección se producen únicamente en los nodos.

$\theta^f_{n_i}$ denota la orientación de la primera asociación en la que se encuentra al nodo i, mientras que el ángulo de salida, $\theta^l_{n_i}$ es la dirección de la última imagen en la que ese mismo nodo es detectado.

La diferencia de esos ángulos proporciona el cambio de orientación en el nodo, y por lo tanto el cambio en la dirección que seguirá el robot móvil hasta realizar un nuevo cambio de dirección:

$$\Delta\theta_{n_i} = \theta^l_{n_i} - \theta^f_{n_i} \tag{6.6}$$

Nótese que $\theta^l_{n_i}$ es la dirección que el robot sigue para llegar desde el nodo n_i al siguiente nodo adyacente.

De esa forma, conocemos la dirección en la que se produce el desplazamiento del robot en todo paso del algoritmo.

Es importante resaltar que los nodos, antes de ser incluidos en el mapa, no tienen un sistema de referencia de coordenadas global. Es decir, no conocemos la orientación de las imágenes incluidas en el mapa. Sin embargo, se supone conocida la dirección del desfase entre imágenes consecutivas de los nodos. En nuestro caso, el giro tiene sentido dextrógiro. Como la dirección del robot se obtiene mediante diferencia de fases, es suficiente con esa información.

En la Figura 6.6 se muestra gráficamente la estimación del cambio de fase en un nodo.

En el algoritmo, nosotros fijaremos la orientación del mapa definiendo el ángulo asociado a la dirección de salida del primer nodo. Esa dirección determinará la orientación global del sistema de referencia del grafo.

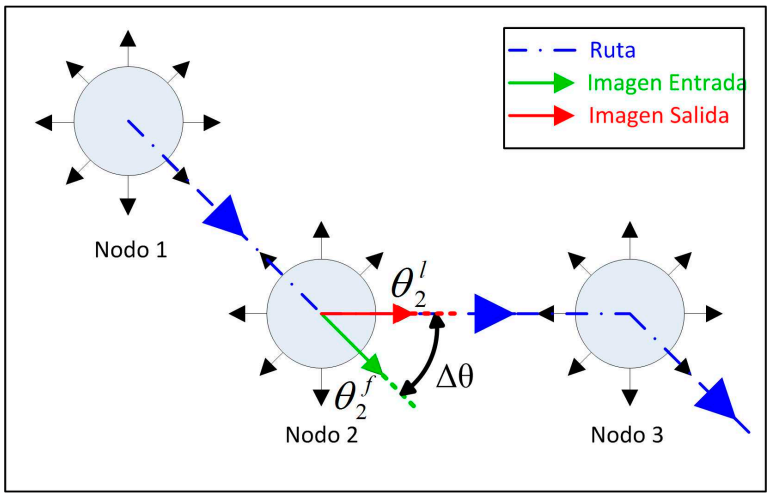


Figura 6.6: Estimación del cambio de fase en un nodo.

En la parte experimental, elegiremos dicha orientación para que coincida con la orientación del grafo de referencia creado sintéticamente. Esto es únicamente para facilitar la comparación visual entre el mapa esperado y el obtenido. El cambio de esta primera orientación supondría un giro del grafo completo.

Después de ese nodo, la orientación de la ruta se actualiza en cada nodo mediante el cálculo del desfase definido en la Ecuación 6.6.

Como en el primer nodo se define el sistema de referencia global del grafo, se puede calcular para cada nodo la diferencia de orientación entre el sistema global y el local. Por ejemplo, si la dirección de entrada de un nodo es 0^o en el sistema global, y corresponde a 90^o en el sistema local del nodo, tenemos un desfase de 90^o entre sistemas para ese nodo.

Cuando se estudia una nueva ruta, el algoritmo inicia un nuevo sistema de coordenadas para sus nodos. Esa ruta se analizará independientemente del grafo global hasta que se encuentre un nodo común. Usando la posición y orientación del nodo común relativo a ambos sistemas, es posible encontrar la diferencia entre ambos sistemas de coordenadas y añadir los nodos de la ruta actual al mapa global.

Si dos rutas comparten un camino común y encuentran nodos que han sido añadidos previamente, las distancias topológicas entre los nodos comunes $c_{n_i,n_{i+1}}$ se estiman de nuevo, y se incluyen en el grafo mediante el cálculo del valor medio con las estimaciones de la distancia entre nodos realizadas previamente. Esta media se ponderará por el número de veces que se ha calculado dicha distancia distancia.

Por lo tanto, el algoritmo desarrollado aprovecha la información proporcionada por distintas rutas con el objeto de seleccionar los nodos del mapa y obtener su posición relativa a través de la asociación de las imágenes. Además, gracias al análisis multiescala, se aumenta

la precisión de asociación y el grafo de la distribución topológica de los nodos presenta una apariencia similar a su distribución espacial.

6.1.3 Estimación de las Rutas sobre el Mapa

Una vez se ha construido el mapa y obtenido el grafo que representa la posición de los diferentes nodos incluidos, nuestro propósito es extender el estudio a la estimación de las rutas en el mapa.

Para ello, se asocian las imágenes de las rutas con la información de las escenas del mapa. Si la localización de nuestro algoritmo se basa únicamente a esta asociación, la precisión del algoritmo está limitada a la posición de los nodos.

Aprovechando de nuevo el análisis multiescala introducido en la Sección 6.1.1, se introduce una mejora de la localización para extender la estimación de la posición no sólo a los nodos sino también a posiciones intermedias entre nodos.

En la estimación de las rutas dentro del grafo del mapa, el primer paso es la asociación entre las imágenes de la ruta y las escenas que componen el mapa. De nuevo se hace uso de la base **Z** que contiene los descriptores de las escenas de los nodos incluyendo distintas escalas, y de múltiples escalas de la imagen de la ruta.

El proceso de asociación y extracción de información a partir de las escenas es muy similar al expuesto en la Sección 6.1.2.1. Sin embargo, como las imágenes provienen de una ruta, se puede suponer que las distancias y los cambios de fase entre imágenes consecutivas no deben ser excesivamente altos. Por ello, se introduce una función de ponderación que penaliza la probabilidad de encontrar la posición y orientación actual de una imagen alejada de los valores de la pose previa.

A continuación se exponen los detalles de la función de ponderación.

6.1.3.1 Función de Ponderación

Tal y como se ha comentado, en el proceso de estimación del camino seguido por las diferentes rutas se va a emplear una función de ponderación. Esta función tiene como objetivo penalizar la probabilidad de que la pose de la imagen actual esté lejos en distancia topológica y orientación a la imagen anterior.

Como el criterio de asociación es el vecino más cercano respecto a la distancia imagen, el decremento de probabilidad que introduce la función de ponderación se representa mediante un incremento de la distancia imagen.

La función de ponderación se compone de dos términos separados: el primero de ellos tiene en consideración la distancia topológica entre la imagen actual y la anterior, mientras que el segundo pondera el desfase entre escenas.

Para calcular la distancia topológica entre dos imágenes, utilizamos la matriz de adyacencia A. Como nuestro grafo es conexo, es posible encontrar siempre un camino que conecte dos puntos del mapa. El coste de llegar de un nodo a otro se corresponde con la distancia topológica entre ambos (c).

En nuestro caso, buscaremos siempre la distancia más corta. Siendo $P_{n_i,n_j} = [n_i, \dots n_j]$ la secuencia de nodos correspondiente al camino más corto que une el nodo i y el nodo j, el coste C_{n_i,n_j} asociado a P_{n_i,n_j} puede ser definido como:

$$C_{n_i,n_j} = \sum_{n_i}^{n_j} c_{n_i,n_{i+1}} \tag{6.7}$$

Introduciendo también el coste asociado al desfase entre imágenes, la función de ponderación queda finalmente como:

$$w(n_i, n_j, \theta_{n_i}, \theta_{n_j}) = \mathrm{w}_1 \cdot C_{n_i,n_j} + \mathrm{w}_2 \cdot \mid \theta_{n_j} - \theta_{n_i} \mid \tag{6.8}$$

donde w_1 y w_2 son constantes que modulan el pesado de la distancia topológica y el desfase respectivamente.

6.1.3.2 Localización dentro del Mapa

Ante una nueva imagen de ruta, el algoritmo calcula la distancia imagen con las escenas del mapa, tal y como se ha definido en la Ecuación 6.3, obteniendo d_i^r, $i = 1 \dots m$.

Ordenamos los resultados por orden creciente de distancia imagen d y seleccionamos los k vecinos más cercanos. A continuación, repetimos el proceso para distintas escalas de la imagen de ruta s^r.

De cada asociación, obtenemos la información descrita en la Ecuación 6.4, es decir, el nodo más cercano n, la orientación estimada de la escena θ, y las escalas de las escenas del nodo (s^n) y de la ruta (s^r).

Estos datos nos permite actualizar el valor de la distancia imagen de los vecinos seleccionados para cada asociación usando la función de pesado:

$$d' = d \times w(n_i, n_{i-1}, \theta_{n_i}, \theta_{n_{i-1}}) \tag{6.9}$$

con n_{i-1} y θ_{i-1} el nos más cercano y orientación de la imagen previa de la ruta, y n_i y θ_i el nodo más cercano y orientación de cada vecino seleccionado en las distintas asociaciones. El valor del pesado puede cambiar para cada vecino, pues n_i y θ_i es particular de cada asociación.

Cuando las distancias imagen de todos los vecinos han sido ponderadas, clasificamos de nuevo esos experimentos utilizando d', y seleccionamos el vecino más cercano.

Con el vector de información asociado al vecino más cercano, el algoritmo es capaz de determinar la pose del robot dentro del mapa. La posición se estimará con el nodo más

cercano, y la distancia relativa l proporcionada por la diferencia de escalas s^n y s^r. Por otro lado, la dirección de avance se obtiene con θ.

El ángulo θ se obtiene respecto al sistema de referencia local del nodo. Para poder expresarlo en el sistema de referencia global debe considerarse el desfase entre el sistema local del nodo y el global, obtenido durante la construcción del mapa.

6.1.4 Experimentos y resultados

En este apartado se presentan los experimentos y resultados de construcción de mapas y localización usando el análisis multiescala sobre imágenes proyectivas.

Primero se incluye una comparativa de descriptores basados en la apariencia global sobre las imágenes proyectivas. El estudio tiene como objetivo medir la precisión de asociación de imágenes y los requerimientos computacionales usando distintas resoluciones de las escenas.

Posteriormente, se describen las bases de imágenes usadas en la parte experimental.

Por último, se muestran los resultados de construcción del mapa para las distintas áreas incluidas en la base de imágenes, y la estimación de los caminos seguidos por las rutas de imágenes sobre el grafo que representa la distribución espacial del mapa.

6.1.4.1 Selección del Descriptor y Resolución de Imagen

Tal y como se ha comentado en la introducción de esta sección, se va a hacer un estudio reducido acerca del descriptor a utilizar sobre las imágenes proyectivas y la resolución de las mismas.

Para ello, se va a llevar a cabo un experimento de reconocimiento de localización a través de asociación de imágenes usando nodos capturados en distintos entornos de interior. Concretamente, se han capturado 26 nodos en localizaciones distintas. En cada posición, además de las imágenes de los nodos, se han adquirido 3 imágenes de test cuya orientación es aleatoria.

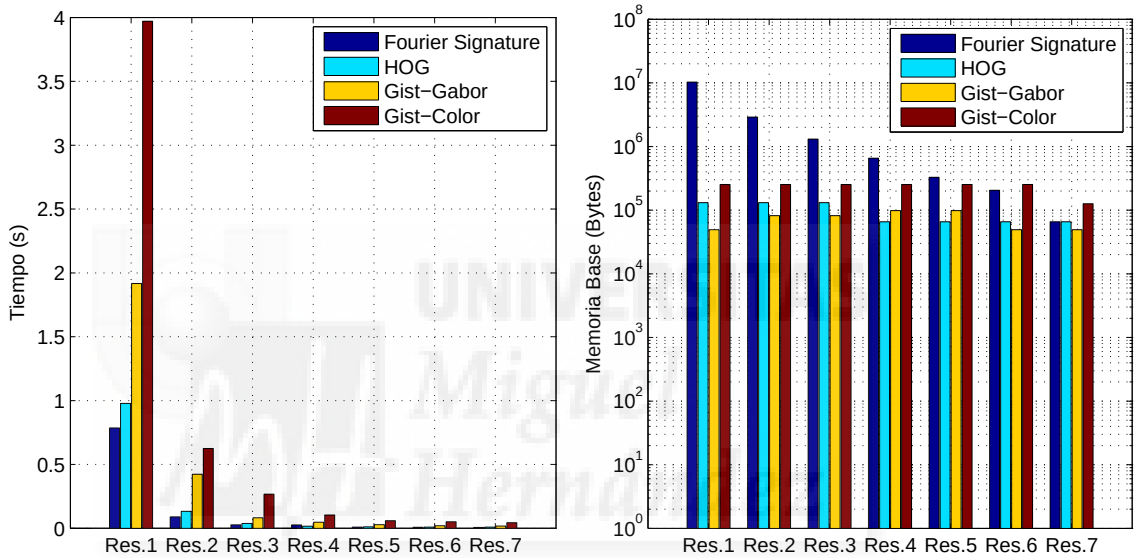
Nótese que la base de imágenes empleada en este análisis es distinta a la usada en los experimentos de construcción del mapa y localización que se presentan posteriormente.

Además, este estudio trata de encontrar el tamaño mínimo de imagen que puede usarse sin que suponga un detrimento de la precisión de localización. Por ello, se repiten los experimentos usando distintas reducciones de las imágenes. En la Tabla 6.1 se pueden ver las distintas resoluciones usadas en los experimentos. La Resolución 1 coincide con el tamaño original de la imagen.

La base de comparación está compuesta por los descriptores de todas las imágenes que componen los nodos, con un total de 26 nodos $\times$ 8 imágenes por nodo. Cuando llega una nueva imagen de test, calculamos su descriptor y lo comparamos con todos los elementos

Tabla 6.1: Resoluciones de Imagen usadas en los experimentos

	Píxeles
Resolución 1	1004×1817
Resolución 2	283×512
Resolución 3	128×256
Resolución 4	64×128
Resolución 5	32×64
Resolución 6	16×32
Resolución 7	8×16



(a) (b)

Figura 6.7: (a) Tiempo y (b) Memoria usando distintos tamaños de imágenes y descriptores.

de la base. La distancia entre las escenas se corresponde con la distancia Euclídea entre sus descriptores, y el clasificador utilizado es el vecino más cercano, es decir, la asociación con menor distancia imagen. Consideramos una asociación correcta cuando el nodo seleccionado es el mismo que el de la imagen de test.

En la Figura 6.7 aparecen los requisitos computacionales de cada descriptor según el tamaño de la escena. Específicamente, en la Figura 6.7 (a) se presenta el tiempo empleado en el cálculo de los descriptores. Los resultados muestran que las técnicas GIST son las que más tiempo requieren para calcular el descriptor, especialmente GIST-Color. También puede apreciarse una disminución exponencial del tiempo cuando se reduce el tamaño de la escena. Esto conlleva que las diferencias en el requerimiento computacional entre descriptores al usar resoluciones menores, disminuya.

Por su parte, la Figura 6.7 (b) muestra el tamaño de la base de comparación para los distintos descriptores. Cabe destacar que la gráfica tiene escala logarítmica en su eje de ordenadas. La Firma de Fourier es el descriptor que más memoria necesita en la mayor parte de resoluciones. Por otro lado, HOG y GIST-Gabor son los descriptores más compactos.

En cuanto a la precisión de localización, la Figura 6.8 muestra las curvas *Recall-Precision* de los descriptores usando las diferentes resoluciones. Las gráficas recogen los resultados considerando los tres vecinos más cercanos (T.N.N.) de las asociaciones. Es significativo que los mejores resultados no tienen por qué coincidir con la máxima resolución.

En las gráficas puede verse que el descriptor más afectado por el cambio de resolución es HOG. También es la técnica que, en general, menos precisión logra. Por otro lado, Fourier es el descriptor que menos varía al modificar la resolución de las imágenes de partida. En cuanto a los descriptores GIST, ambos consiguen una precisión elevada, siendo destacable GIST-Gabor.

El principal criterio para la selección del descriptor es la precisión en la asociación de imágenes de forma correcta. Por esa razón, HOG es menos apropiado que las otras técnicas para esta aplicación, aunque sus requerimientos de tiempo y memoria sean favorables. La Firma de Fourier presenta una elevada precisión en la estimación de la posición para cualquier resolución de imagen. Sin embargo, es menor que con los descriptores GIST. Además, el tamaño del descriptor usando esta técnica es el mayor.

Comparando los dos descriptores GIST, GIST-Gabor mejora a GIST-Color en tiempo de cálculo, además de ser un descriptor más compacto.

Por esa razón, el descriptor seleccionado para llevar a cabo los experimentos es GIST-Gabor. Los resultados obtenidos con la Resolución 5 muestran un buen compromiso entre precisión y requerimientos computacionales. Así pues, la resolución de imagen elegida es de 32×64.

Cabe destacar que con la Resolución 6 también se obtiene una precisión de localización alta. Sin embargo, tal y como se ha visto en la Sección 6.1.1, la comparación multiescala utiliza zoom digital de las imágenes de entrada, lo que supone una reducción de la resolución de la imagen. Esto podría provocar una disminución de la precisión en la localización al usar una resolución insuficiente de la escena.

6.1.4.2 **Bases de Imágenes**

Se han capturado dos bases de imágenes en distintas áreas. Las dos corresponden a zonas comunes del edificio Merchant Venturers de la Universidad de Bristol.

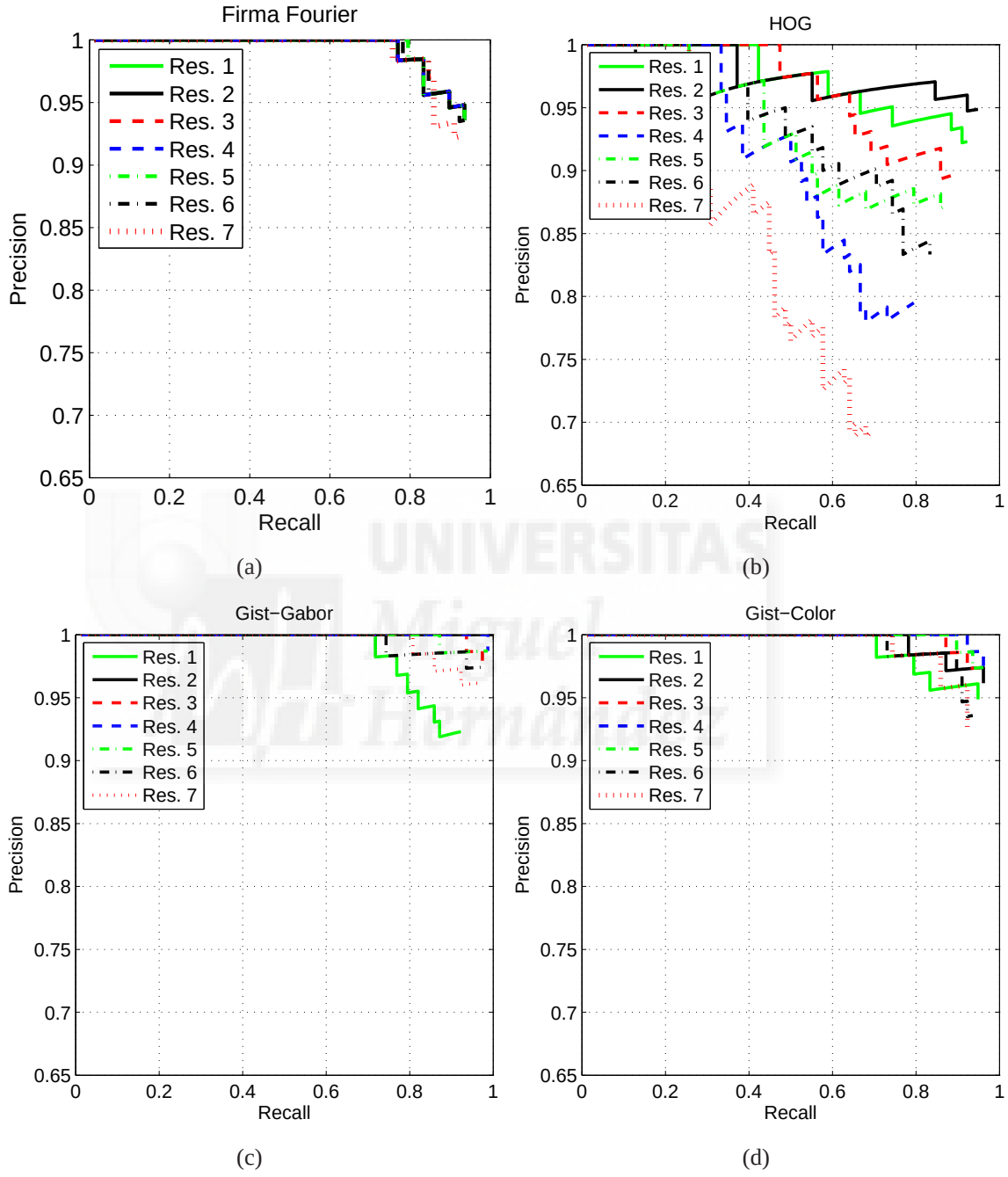


Figura 6.8: Resultados Recall-Precision en precisión de Localización considerando el Tercer Vecino más Cercano (T.N.N.) para distintas resoluciones de imagen usando (a) la Firma de Fourier, (b) HOG, (c) GIST-Gabor y (d) GIST-Color.

Tabla 6.2: Número de imágenes por área de la base experimental.

	# Imágenes Área 1	# Imágenes Área 2
Nodos	352	52
Ruta 1	110	100
Ruta 2	50	72
Ruta 3	67	66
Ruta 4	58	125
Ruta 5	62	-
Ruta 6	46	-
Ruta 7	69	-
Ruta 8	67	-
Ruta 9	40	-

La distorsión radial de las imágenes originales ha sido corregida teniendo en cuenta la calibración del sistema visual. Las distintas simulaciones han sido realizadas usando las escenas sin distorsión.

Cada base se compone de un conjunto de nodos y rutas de imágenes repartidas por el entorno donde se encuentran los nodos. Cabe recordar que cada nodo está compuesto por 8 imágenes, con un desfase de 45^o entre imágenes consecutivas, cubriendo de esa forma el campo de visión completo de la posición en la que se encuentra.

La Figura 6.9 muestra la distribución de los nodos en sendos planos del edificio. Puede observarse que el mapa correspondiente al Área 1 es el más extenso.

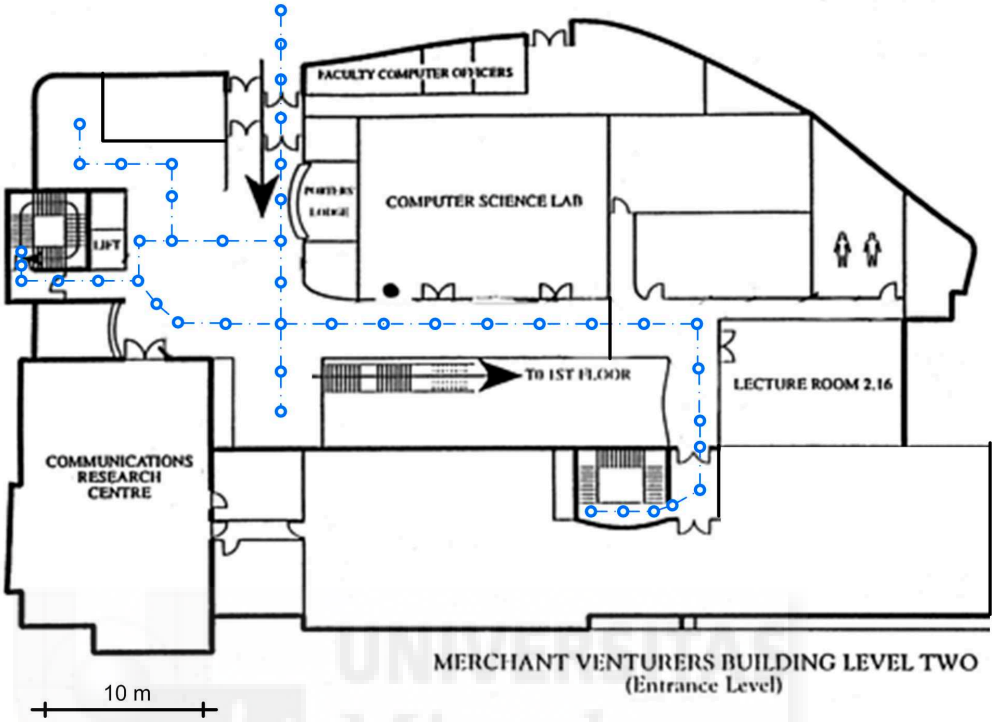
La distancia real entre nodos es de dos metros como norma general, aunque en lugares donde se produce una cambio importante en la apariencia de las imágenes, como por ejemplo al cruzar una puerta, se adquiere un nuevo nodo independientemente de la distancia con el nodo previo. Por esa razón, la distancia entre nodos puede ser menor.

Con respecto a las rutas, la frecuencia de adquisición de imágenes es mayor en el caso del Área 2. Específicamente, las imágenes de las rutas del Área 1 se adquieren cada 0.5 metros, mientras que en el Área 2 esa distancia es de 0.2 metros.

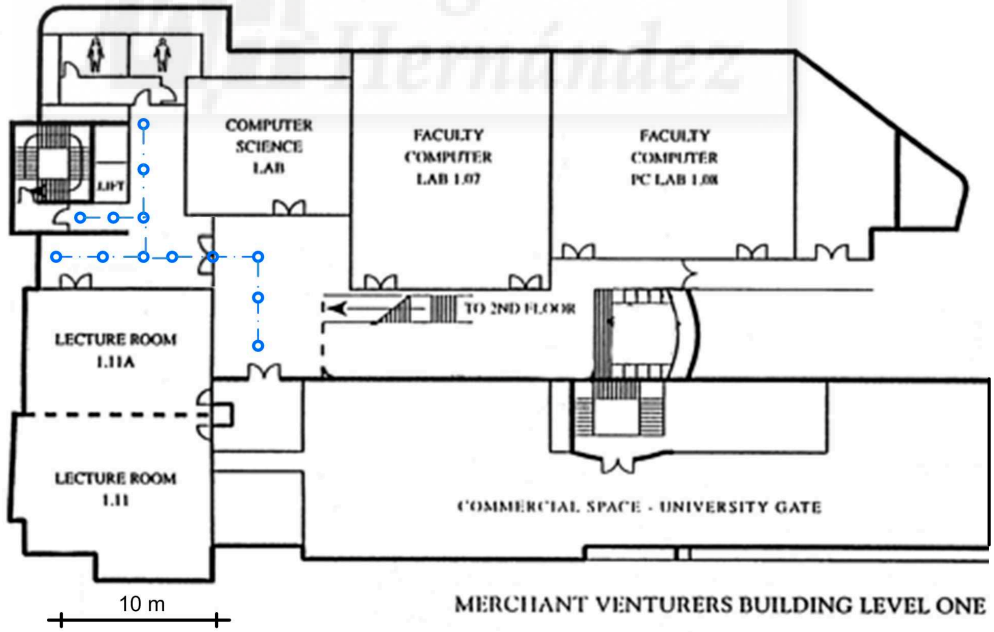
En los puntos de las rutas que se produce un cambio de orientación, se adquiere un mínimo de cuatro imágenes por posición durante el giro en el Área1, aumentando a seis imágenes por posición en el caso del Área 2.

Por otro lado, la Figura 6.10 incluye un esquema sintético de la distribución de los nodos y de las rutas capturadas en el entorno. Esta información se complementa con la Tabla 6.2, que incluye el número de imágenes del mapa y de las distintas rutas capturadas.

Las imágenes han sido adquiridas en entornos reales y dinámicos. Debido a ello, las condiciones de iluminación no son constantes, existiendo además movimiento de elementos

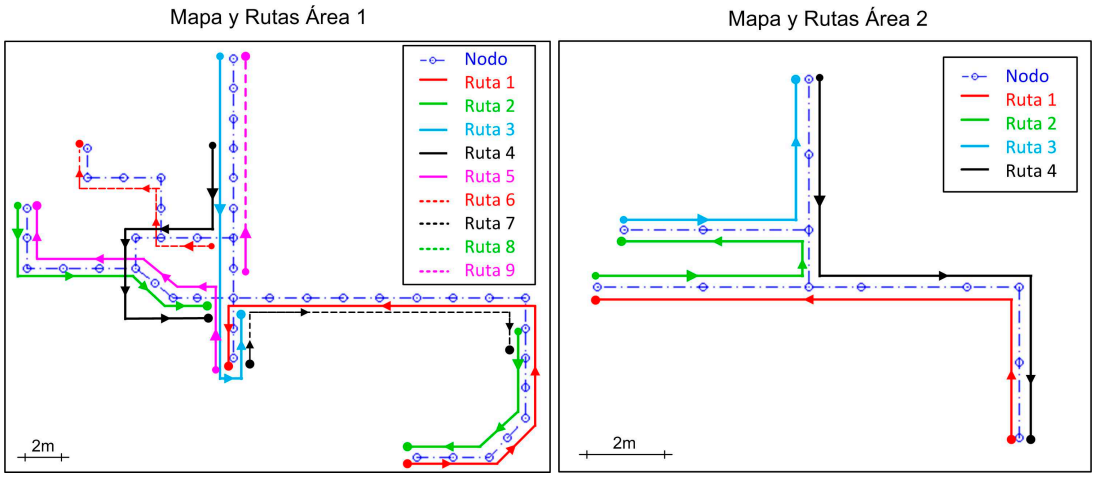


(a) Mapa Área 1



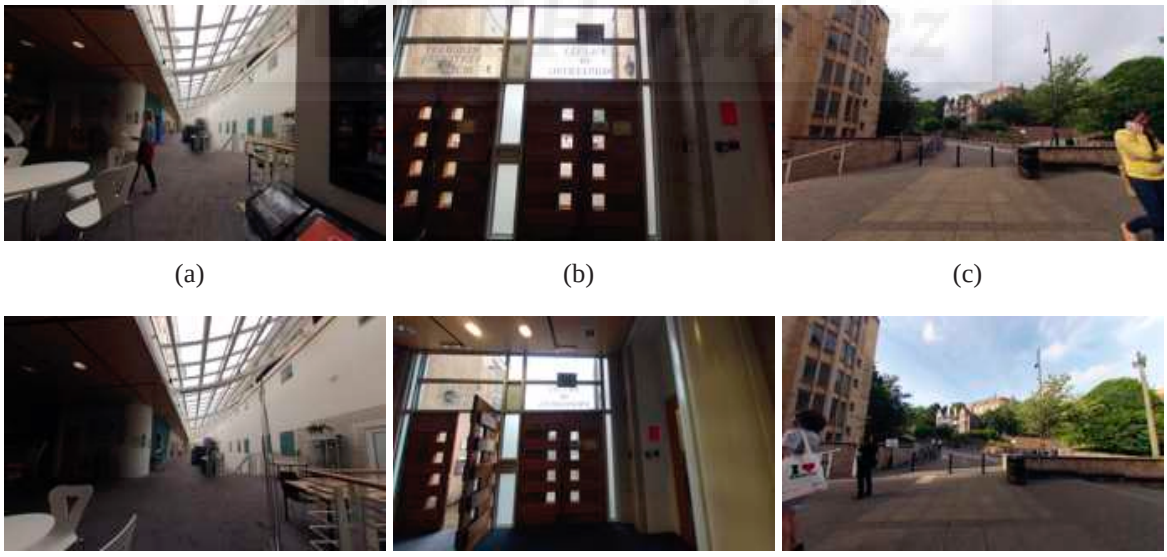
(b) Mapa Área 2

Figura 6.9: Representación de los distintos grafos sobre el plano de cada área de navegación.



(a) (b)

Figura 6.10: Representación de los mapas y las rutas de las distintas áreas.



(a) (b) (c)

(d) (e) (f)

Figura 6.11: (a), (b), (c) Escenas correspondientes a Nodos del Área 1. (d), (e), (f) Escenas correspondientes a Rutas del Área 1.

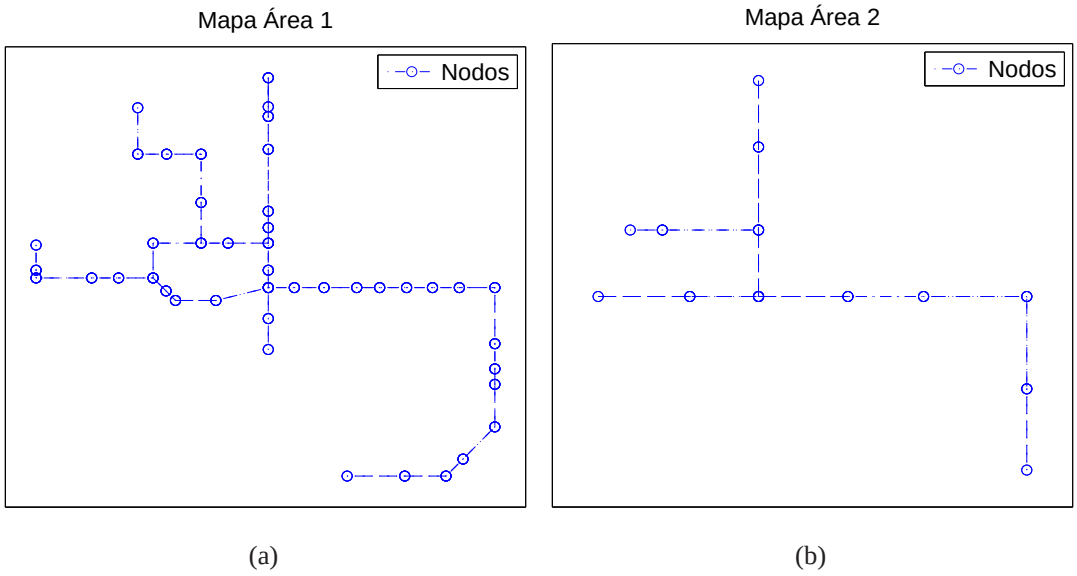


Figura 6.12: Grafos con las distribuciones de los nodos obtenidos experimentalmente en la construcción del mapa del (a) Área 1 y (b) Área 2.

de mobiliario y personas durante la captura.

La Figura 6.11 presenta en la fila superior imágenes de los nodos, y en la fila inferior ejemplos de escenas de rutas cercanas a esos nodos. Estas escenas son ejemplos de los cambios que se producen en el entorno tanto en iluminación como en oclusiones.

6.1.4.3 **Resultados de la Construcción del Mapa**

La Figura 6.12 muestra el mapa obtenido de (a) el Área 1 y (b) el Área 2. En los resultados, podemos apreciar que el algoritmo ha sido capaz de encontrar todos los nodos del mapa, y además estimar las distancias entre ellos de forma correcta. También es posible observar que la distribución de los nodos en el plano y la obtenida en el grafo es muy similar.

Para medir el error entre el plano real y el estimado usamos el análisis Procrustes [50, 100]. La función tiene como datos de entrada la posición de los nodos de ambos grafos. Primero, el algoritmo busca la transformación lineal que mejor hace coincidir los puntos de un grafo con los del otro, teniendo en cuenta translación, reflexión, rotación ortogonal y escalado. Tras ello, se mide la bondad de ajuste a través de la suma de errores cuadrados. El análisis Procrustes devuelve un índice de disparidad estandarizado $\mu \in [0,1]$. Al ser un índice de disparidad, cuanto menor sea μ, más similar es el grafo obtenido a la distribución real.

La principal ventaja del análisis Procrustes es que no es necesario que los datos comparados estén expresados en las mismas unidades, como es nuestro caso, ya que introduce un escalado de la información. Los resultados se muestran en la Tabla 6.3.

	Área 1	Área 2
μ	0.0372	0.0078

Tabla 6.3: Resultados del análisis Procrustes de los grafos de los distintos mapas.

El error geométrico de los grafos es bajo en ambos casos. El Área 1 es la que presenta más dificultad debido al mayor número de nodos, y la existencia de transiciones entre distintos espacios en el mapa y un bucle cerrado.

En la parte izquierda inferior del bucle, el grafo estimado presenta una diferencia geométrica con respecto a la distribución real. Sin embargo, aunque el mapa pierde precisión, su capacidad de ser usado en tareas de navegación no se ve afectado, ya que la información sobre qué dirección debe seguir el robot para llegar de un nodo al siguiente no se ve afectada.

Es importante remarcar que, en el proceso de construcción del mapa, el algoritmo necesita un número mínimo de asociaciones donde se encuentre un mismo nodo para añadirlo al mapa. Si no hay frecuencia suficiente de adquisición de imágenes en las rutas, o los nodos se encuentran situados muy cerca geométricamente, es posible que un nodo existente en la base de imágenes no sea incluido en el mapa.

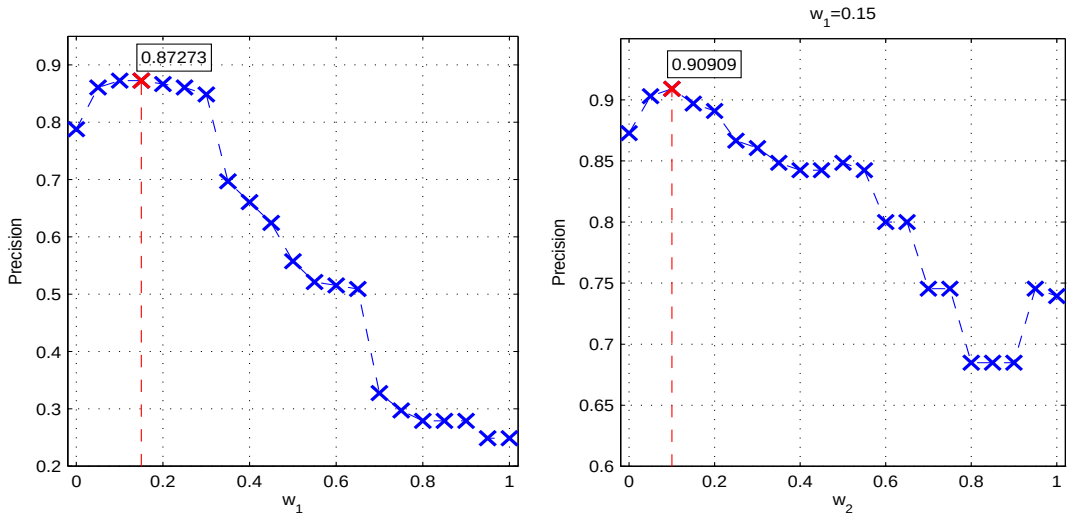
El sistema propuesto es especialmente sensible cuando se produce un cambio de dirección en alguno de los nodos, ya que está basado en la diferencia de ángulo entre la asociación de la primera y última imagen de ruta donde se detecta un nodo. Por esa razón, es recomendable aumentar el número de imágenes adquiridas en las zonas cercanas a los nodos cuando se produce un cambio de dirección.

Con respecto a los parámetros, en los experimentos se ha seleccionado un valor máximo escala de nodo $s^n = 2{,}5$, con un paso de $0{,}1$ entre escalas consecutivas. La escala máxima utilizada en las escenas de ruta es de $s^r = 1{,}4$, con un paso de $0{,}05$. Se han elegido unos pasos reducidos entre escalas consecutivas porque la prioridad en las simulaciones ha sido la precisión del mapa sobre los requerimientos computacionales.

Realizando las simulaciones con un procesador con velocidad 2.8GHz, desarrollando el software con Matlab, el tiempo necesario de análisis por imagen de ruta (incluyendo todas las escalas s^r), es de 725 ms para el Área 1, y de 680 ms para el Área 2.

El descriptor elegido en ambos casos es el mismo, y corresponde al seleccionado en el estudio realizado en la Sección 6.1.4.1, es decir, GIST-Color sobre una imagen de resolución de 32×64 píxeles.

La diferencia de tiempo radica en la asociación de imágenes, ya que la base de comparación $\mathbf{Z}$ del Mapa 1 es más de tres veces superior en tamaño a la del mapa 2. La estimación de la distancia imagen entre la escena de entrada y la base de imágenes de los nodos supone el 65 % del tiempo global en la construcción del mapa.



(a) (b)

Figura 6.13: Precisión de asociación de imágenes respecto a las constantes de ponderación w_1 y w_2. (a) Precisión variando la constante relativa a la distancia topológica (w_1) y (b) precisión variando la constante relativa al cambio de fase (w_2).

6.1.4.4 Resultados de Localización de Rutas en el Mapa

A continuación se presentan los resultados obtenidos de la estimación de las rutas de imágenes sobre el mapa creado. Para ello, primero se introduce un estudio sobre los parámetros de la función de ponderación, y se presentan los resultados representados sobre el grafo del mapa.

Con el propósito de ajustar los valores de las constantes de pesado w_1 y w_2, se mide la precisión de asociación entre las imágenes de ruta y las imágenes de los nodos variando los distintos parámetros. Se considera una asociación correcta cuando la imagen seleccionada del mapa se corresponde al nodo más cercano y tiene la orientación correcta.

La Figura 6.13 muestra los valores variando ambos parámetros. En Figura 6.13 (a) se estudia la precisión respecto a la variación de w_1, que pondera la distancia topológica entre las asociaciones. En estos experimentos, w_2 se mantiene fija igual a 0. El máximo se encuentra en $w_1 = 0{,}15$.

Una vez se ha seleccionado w_1, se estudia la precisión variando la constante relacionada con el desfase entre escenas consecutivas. Los resultados se incluyen en la Figura 6.13 (b). En este caso, el máximo se obtiene para $w_2 = 0{,}1$.

Ambas gráficas muestran que la acción de la función de ponderación mejora la precisión de asociación. No obstante, si la función de ponderación es muy restrictiva ante los cambios de posición y orientación entre imágenes consecutivas, la precisión disminuye. Por esa razón, cuando las constantes tienen valores altos, los resultados empeoran.

Finalmente, en los experimentos de estimación de la trayectoria de las diferentes rutas, las constantes de ponderación son $w_1 = 0{,}15$ y $w_2 = 0{,}1$. Se usa un total de $k = 10$ vecinos más cercanos cuando se realiza la asociación entre cada zoom de las escenas de ruta. Las escalas de los nodos tienen un rango de $s^n = [1,\ 2{,}2]$, con un paso de 0,4 entre imágenes consecutivas, mientras que las imágenes de ruta tienen el mismo rango de escalas pero con un paso de 0,3.

La Figura 6.14 muestra el camino estimado en las diferentes rutas para ambas áreas. Las marcas en los caminos representan la posición de distintas imágenes de las rutas. Como puede apreciarse, el algoritmo es capaz de estimar la posición de las escenas en puntos intermedios entre los nodos usando la información de escalas.

En general, la precisión de localización cuando se produce un cambio en la de dirección en las rutas decrece. También es significativo que, a pesar de introducir la función de ponderación, ante un fallo en la localización, el algoritmo es capaz de volver a encontrar la asociación correcta en un número pequeño de iteraciones. Esto se ve reflejado en las Figuras 6.14 (a) y (c), por ejemplo.

También es notable el resultado de la cuarta ruta del Área 1. Como se puede apreciar en la distribución sintética mostrada en la Figura 6.10 (a), esta ruta presenta en la parte interior del bucle central del mapa una trayectoria que difiere de la del mapa. Esta diferencia puede verse reflejada en los resultados (Figura 6.14 (b)).

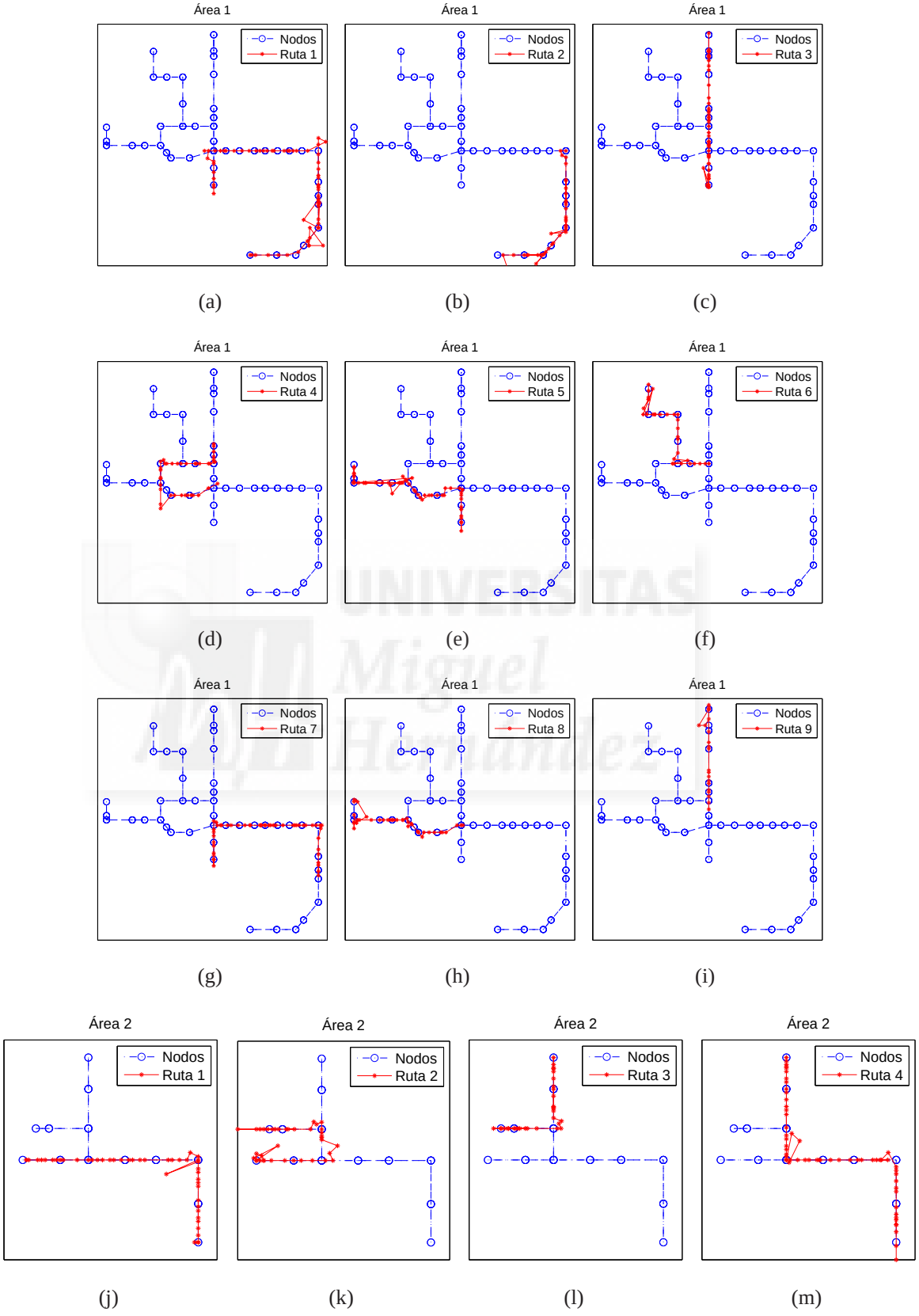


Figura 6.14: Resultados de estimación del camino de las diferentes rutas del Área 1 y 2.

6.2 Aplicación del Análisis Multiescala a Imágenes Omnidireccionales

A continuación, se presenta un algoritmo que extiende el uso del Análisis Multiescala para su uso sobre imágenes omnidireccionales.

Tal y como hemos visto en la sección anterior, el Análisis Multiescala permite extraer información relativos al desplazamiento entre dos escenas mediante el uso de ampliaciones o zooms artificiales de la zona central sobre imágenes proyectivas. De esta forma, es posible aplicarlo a tareas de navegación topológicas.

En este punto, se va estudiar cómo aplicar ese mismo análisis sobre información omnidireccional para crear un sistema de odometría visual topológica. La odometría visual aprovechará la estimación de fase de los descriptores de apariencia global estudiados en el Capítulo 4 y la estimación de desplazamiento a partir del Análisis Multiescala. Con esta información, es posible estimar el desplazamiento entre dos escenas omnidireccionales consecutivas. De esa forma, será posible estimar el camino seguido por el robot usando únicamente información visual.

Además, también se tendrán en cuenta cierres de bucle. El criterio para determinar si la posición actual del robot pertenece a un punto visitado anteriormente es la distancia imagen entre las escenas panorámicas. Si el algoritmo determina que se ha producido un cierre de bucle, se corregirá la trayectoria del robot usando la información almacenada durante su recorrido.

6.2.1 Extracción de Posición Relativa entre dos Imágenes.

En el Capítulo 3 se ha mostrado que, a partir de las imágenes omnidireccionales y la calibración del sistema visual, es posible obtener distintas proyecciones de la información visual. Específicamente, en este caso se utilizarán proyecciones perspectivas (Sección 3.2.3).

Como la información que se dispone es omnidireccional, es posible obtener vistas proyectivas de cualquier punto alrededor del sistema visual. Para poder extraer información útil del análisis multiescala, es necesario que las proyecciones utilizadas se realicen en un plano perpendicular a la dirección de avance. Aprovechando que la información que disponemos es omnidireccional, el algoritmo aplicará el análisis multiescala en dos direcciones distintas, que corresponderán con la dirección de avance, y la contraria. Ambas proyecciones cumplen la condición de ser perpendiculares a la dirección en la que avanza el robot.

La Figura 6.15 muestra las proyecciones que se realizan de una imagen de ruta. La flecha azul indica la dirección de avance muestra la dirección de avance, mientras que la roja corresponde con la contraria.

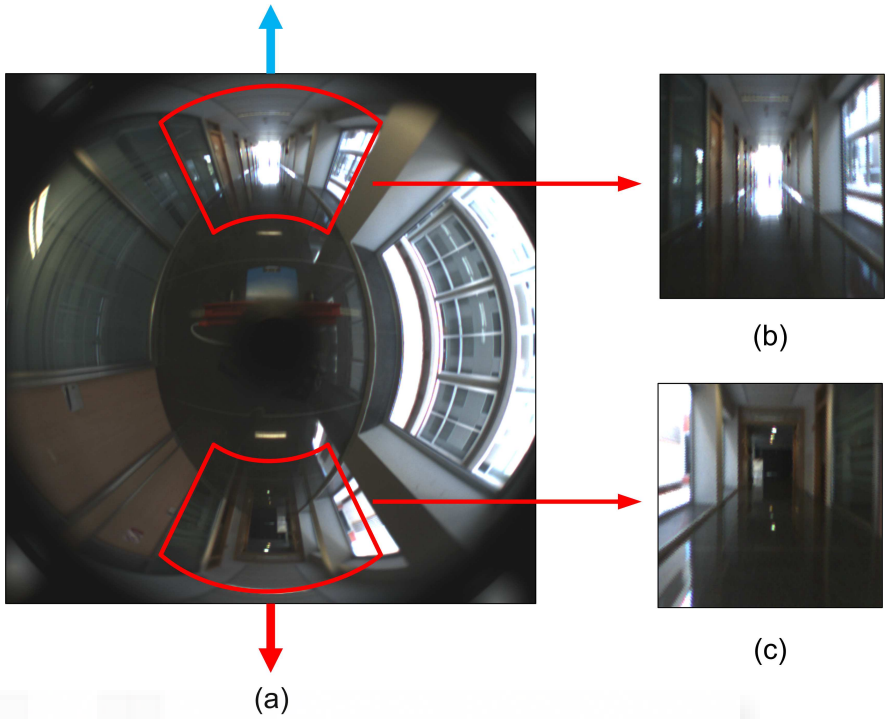


Figura 6.15: Extracción de escenas proyectivas en la dirección de avance y de la opuesta a partir de una imagen omnidireccional.

Hay que tener en cuenta que, cuando el robot avanza, se producirá un acercamiento a las zonas situadas en la dirección de avance, y un alejamiento de las opuestas. Por tanto, las escalas aplicadas a las escenas de una dirección deben ser distintas a las de la otra dirección. Cuando en una se considera un incremento, sobre la otra habrá que aplicar una reducción. Concretamente, se aplicarán incrementos de escala (o de focal, en este caso) iguales, pero de signo opuesto.

En la Figura 6.16 se incluye un ejemplo. La focal central corresponde a $fc = 1{,}1$. Al aplicar el análisis multiescala, la diferencia de focal se sumará en la dirección de avance, y se restará en la opuesta.

Nótese que, a diferencia del sistema propuesto en la sección anterior, este algoritmo utiliza la diferencia de focales en lugar de la diferencia de escalas. Al tratarse de navegación topológica, esto únicamente afecta en el hecho de que el signo de los indicadores es contrario. Mientras que una diferencia de escalas (Δs) positiva suponía una ampliación de la escena, en este caso corresponderá a una diferencia de focales (Δfc) negativa. O dicho de otro modo, al reducir la focal, se produce una ampliación de la imagen.

Sin embargo, en la práctica se ha modificado el signo de la diferencia de focales para que el análisis multiescala devuelva un valor con el mismo signo que el desplazamiento detectado. Por tanto, un desplazamiento en la dirección de avance tendrá signo positivo.

Para cada incremento de focal, se calcula el descriptor $\mathbf{z}_{proy}$. Este vector contendrá los descriptores de las dos proyecciones obtenidas, tanto la de avance como la opuesta. Considerando las diferencias de focales, $\mathbf{z}_{proy}$ se define como:

$$\mathbf{z}_{proy,\Delta fc} = [\ \mathbf{z}_{avance,\, fc_{central}+\Delta fc}\ ,\ \mathbf{z}_{opuesta,\, fc_{central}-\Delta fc}\ ] \tag{6.10}$$

Para estimar el desplazamiento entre dos imágenes, se calcula $\mathbf{z}_{proy,\Delta fc}$ para distintos incrementos de focal de una de esas imágenes. Luego, comparamos con el descriptor correspondiente a $\Delta fc = 0$ de la segunda escena ($\mathbf{z}_{proy,0}$), es decir, sin aplicar ninguna ampliación/-reducción. Dicha comparación se realizará mediante el cálculo de la distancia Euclídea entre los dos descriptores.

Tras ello, se selecciona la comparación que presenta una menor distancia. La diferencia entre las focales de las proyecciones seleccionada proporciona el indicador de desplazamiento relativo entre las dos escenas. Matemáticamente, esta distancia entre descriptores para una imagen i con respecto a la anterior d^i_{proy} se puede expresar como:

$$d^i_{proy} = \text{mín}\left(\sqrt{\sum_{l=1}^{L}\left((\mathbf{z}^i_{proy,\Delta fc}(l))^2 - (\mathbf{z}^{i-1}_{proy,0}(l))^2\right)}\right)\ \forall\, \Delta fc. \tag{6.11}$$

siendo L el número de elementos del vector $\mathbf{z}$.

Obsérvese que, como para la segunda imagen consideramos $\Delta fc = 0$, la diferencia de focales entre las dos imágenes coincidirá con el incremento de focales Δfc asociado a la primera imagen seleccionada en la asociación.

Ese Δfc proporciona información de la magnitud de desplazamiento entre las escenas comparadas mediante su módulo ($|\Delta fc|$), y del sentido del desplazamiento, mediante su signo.

6.2.1.1 **Aplicación a tareas de navegación**

El análisis multiescala sobre la imagen omnidireccional, tal y como se ha explicado anteriormente, no tiene en cuenta los posibles desfases entre dos imágenes consecutivas de la ruta. Sin embargo, tal y como se muestra en la Figura 6.17 (b), el desplazamiento del robot entre la captura de dos escenas no tiene por qué ser rectilíneo. En tal caso, si realizamos el análisis multiescala utilizando las direcciones de avance, se obtendrán proyecciones en dos direcciones no coincidentes. Por lo tanto, las proyecciones comparadas serán distintas.

Nuestra propuesta es aprovechar la capacidad de los descriptores basados en apariencia global para estimar el desfase entre dos escenas panorámicas, como se ha comprobado en el Capítulo 5, para determinar las direcciones de las escenas proyectivas a comparar.

Siguiendo el ejemplo de la Figura 6.17 (b), el análisis multiescala de dos imágenes que presentan desfase entre sí utilizará las proyecciones en la dirección de avance en la última

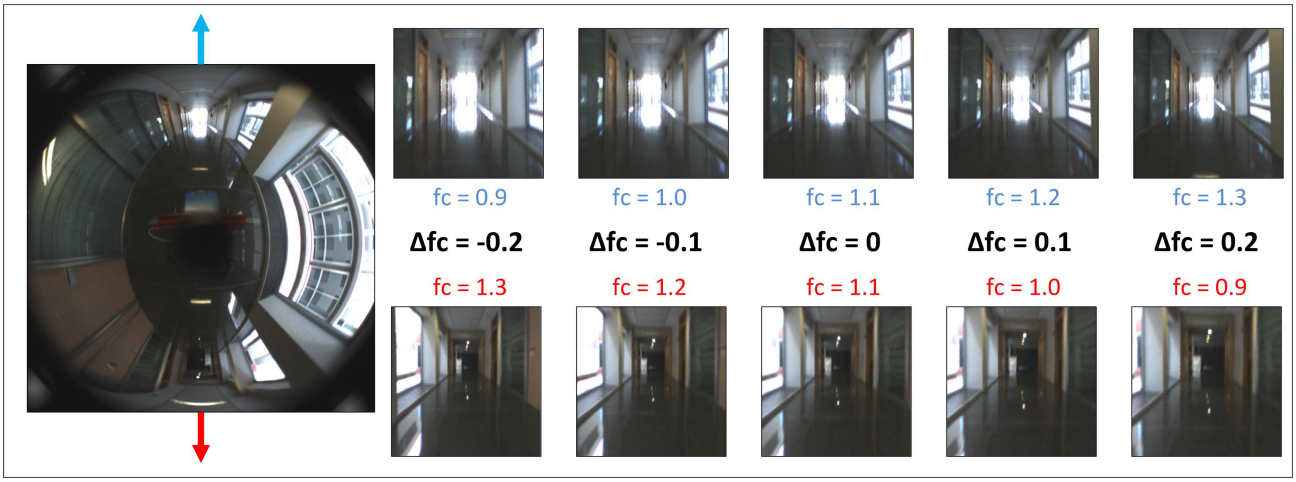


Figura 6.16: Variación de focales de las escenas proyectivas en la dirección de avance y la opuesta para su aplicación en el Análisis Multiescala.

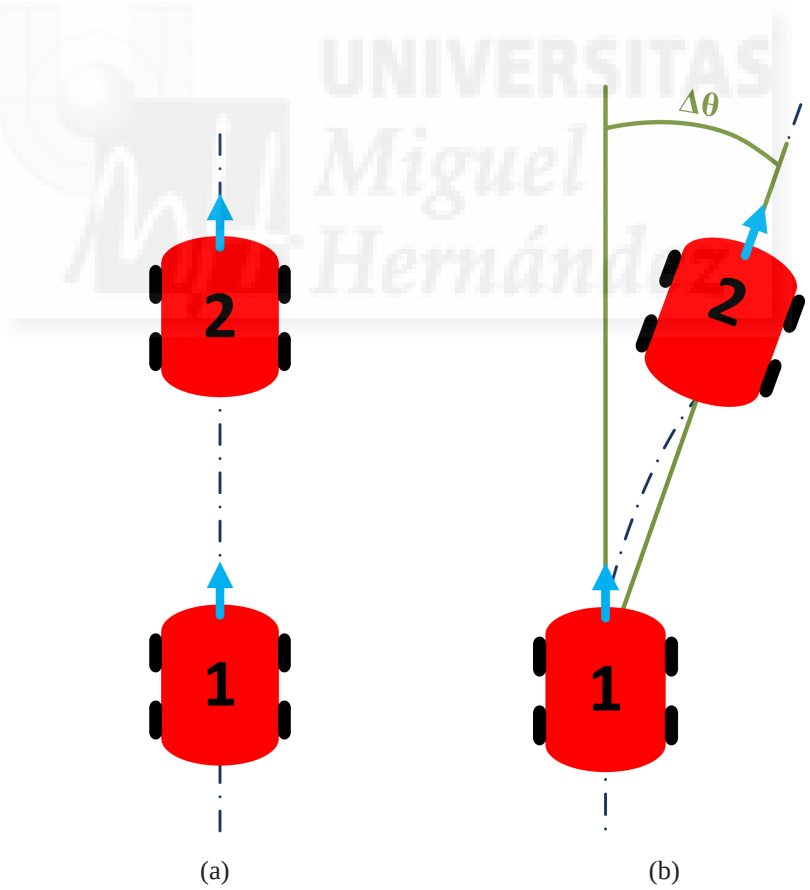


Figura 6.17: Ejemplo de distintas trayectorias seguidas por el robot. (a) Trayectoria rectilínea, y (b) Trayectoria con cambio de dirección.

imagen (la segunda en el ejemplo), y aplicará un desfase de $\Delta\theta$ respecto a la dirección de avance para las proyecciones de la escena anterior (la primera del ejemplo).

Cabe destacar que, cada vez que se calcula el desfase entre dos imágenes consecutivas, se almacenará una medida de la distancia imagen asociada, denotada como d_{fase}. Este valor dependerá del descriptor utilizado. Por ejemplo, para aquellos descriptores que utilizan correlación de ventanas verticales, como HOG o GIST, d_{fase} corresponderá a la distancia imagen entre las ventanas donde se produce la mínima distancia imagen (y que por lo tanto, denota el desfase entre escenas, tal y como se muestra en las Secciones 4.3.2 o 4.4.2). En el caso de Fourier, corresponderá a la diferencia entre los coeficientes de fase para la rotación en la que obtenemos la diferencia mínima.

El objetivo de d_{fase} es obtener una medida de incertidumbre de la estimación de fase. Las distancias serían inversamente proporcionales a la confianza de la estimación, es decir, cuanto mayor sea d_{fase}, menos fiabilidad tendrá la estimación. Estas distancias serán utilizadas por el algoritmo posteriormente.

Nótese que, en ese caso, se está suponiendo que la orientación actual del robot y la de la línea que une la pose actual con la anterior, es la misma. Es decir, estamos aproximando la dirección del desplazamiento al desfase entre escenas (Figura 6.17 (b)). Estas direcciones no tienen por qué coincidir. No obstante, considerar ese desfase es mejor aproximación que suponer únicamente movimientos rectilíneos.

Resumiendo, de este análisis obtenemos un sistema de odometría visual topológica, pues es posible conocer la pose del robot respecto a la imagen anterior, ya que se dispone de la dirección de desplazamiento y magnitud del desplazamiento.

6.2.2 Mejora de la localización dentro de mapa topológico

Hasta este punto, se ha presentado un sistema de odometría visual topológica. Aplicado a una tarea de navegación real, el sistema planteado acumularía error conforme se realiza la estimación de posición de una nueva imagen.

Para poder limitar ese error, se va a incluir un sistema de reconocimiento de zonas anteriormente visitadas, llevando a cabo un cierre de bucle. Cuando el algoritmo detecta que una imagen almacenada durante la navegación es muy similar a la actual, calcula la pose respecto de esa imagen anterior, que tendrá asociado un error menor de odometría. Además, reestima la posición del tramo de la ruta contenido entre las dos imágenes asociadas.

El cierre de bucle se estudiará usando la apariencia global sobre las escenas panorámicas. Concretamente, se hará uso de descriptores invariantes a rotación (Capítulo 4).

El algoritmo estimará la distancia imagen entre las distintas escenas de la ruta y la actual mediante la distancia Euclídea entre descriptores (Ecuación 6.12).

Esta distancia proporciona información sobre la diferencia de apariencia entre dos escenas. El algoritmo determinará que está en una zona anteriormente visitada si la distancia imagen queda por debajo de un umbral que se fijará de forma experimental.

Si comparamos las escenas justamente anteriores a la actual presentan una apariencia muy similar con ella. Por ello, si son tenidas en cuenta en las comparaciones para determinar los cierres de bucle, tienen una alta probabilidad de quedar por debajo del umbral fijado, y por lo tanto, considerar un cierre con esas imágenes. Sin embargo, esas son asociaciones no deseadas. Por ello, en las comparaciones, estos casos son excluidos. En concreto, nuestro algoritmo no considerará las últimas 20 imágenes.

Denotaremos $\mathbf{z}^i_{pano}$ como el descriptor de la imagen panorámica de la i-ésima escena. Considerando que la imagen actual es la número n, calcularemos las distancias imágenes como:

$$d^i_{pano} = \sqrt{\sum_{m=1}^{M} \left((\mathbf{z}^i_{pano}(m))^2 - (\mathbf{z}^n_{pano}(m))^2\right)}, \quad i = 1,\ldots,n-20. \tag{6.12}$$

siendo M el número de elementos del descriptor de la imagen panorámica ($\mathbf{z}_{pano}$).

Obtenidas todas las distancias imagen entre las posiciones del mapa, nos quedamos con la mínima y comprobamos si queda por debajo del umbral fijado, denotado como th_{pano}:

$$\text{mín}\left(d^i_{pano}, i = 1,\ldots,n-20\right) < th_{pano} \tag{6.13}$$

Si la condición no se cumple, se calcula la nueva posición respecto la pose anterior. Para ello, se utilizará el desfase entre escenas ($\Delta\theta$) y el desplazamiento obtenido mediante el análisis multiescala (Δfc). Para una imagen i, la nueva posición se obtendrá como:

$$\begin{bmatrix} x^i \\ y^i \\ \theta^i \end{bmatrix} = \begin{bmatrix} x^{i-1} + \Delta x^i \\ y^{i-1} + \Delta y^i \\ \theta^{i-1} + \Delta\theta \end{bmatrix}, \; i = 1,...,n \tag{6.14}$$

siendo

$$\begin{bmatrix} \Delta x^i \\ \Delta y^i \end{bmatrix} = \begin{bmatrix} \Delta fc \cdot \sin\left(\theta^i\right) \\ \Delta fc \cdot \cos\left(\theta^i\right) \end{bmatrix} \tag{6.15}$$

Por el contrario, si la condición se cumple, el algoritmo interpretará que existe un cierre de bucle entre la posición actual del robot y la posición del mapa con la que se ha hallado la correspondencia (Ecuación 6.13).

Se seguirán tres pasos básicos:

1. Actualización de la pose actual
2. Corrección angular del mapa
3. Corrección de las posiciones XY del mapa

A continuación se detallan estos tres pasos

- Actualización de la pose actual

  Cuando el algoritmo reconoce un cierre de bucle, se considera que la pose actual es cercana a una posición visitada anteriormente. Las coordenadas de ese punto anterior del mapa son conocidas, y el error asociado por la odometría, menor. Por ello, se propone hacer una nueva estimación de la pose actual respecto a esa imagen anterior.

  Es recomendable seleccionar un umbral bajo. De esta forma, si se cumple la condición, hay una mayor probabilidad que la asociación sea correcta, y que la posición real entre las dos imágenes asociadas sea más cercana.

  La pose respecto a la imagen anterior del mapa $\left([x_{actual-mapa},\, y_{actual-mapa},\, \theta_{actual-mapa}]^T\right)$ será calculada de la misma manera que entre imágenes consecutivas, ya que consistirá en la estimación del desfase entre las dos escenas ($\Delta\theta$), y su desplazamiento relativo (Δfc).

- Corrección angular del mapa

  Calculada la fase $\theta_{actual-mapa}$, se estima el error de fase de la ruta. Para ello, se utiliza la fase calculada en la última posición de la ruta, y obtenemos la diferencia con la fase obtenida en el cierre de bucle. Siendo $\left([x_n,\, y_n,\, \theta_n]^T\right)$ la pose de la imagen actual (n) calculada siguiendo la ruta, el error de fase se calcula como:

$$e_{fase} = \theta_{actual-mapa} - \theta^n \tag{6.16}$$

  Propagamos ese error hacia atrás en todo el tramo considerado en el cierre de bucle. El error será ponderado con la incertidumbre asociada a la fase de cada paso de la ruta d_{fase}. Considerando que la imagen actual de la ruta es n, y que la primera imagen a corregir del bucle es n_{bucle}, las orientaciones de las poses de la ruta serán modificadas como:

$$\overline{\Delta\theta^i} = \Delta\theta^i + e_{fase} \cdot \frac{d^i_{fase}}{\sum_{j=n_{bucle}}^{n} d^j_{fase}},\ i = n_{bucle}, \ldots, n. \tag{6.17}$$

  Calculadas los nuevos desfase, se obtienen las posiciones del mapa contenidas en el bucle utilizando las nuevas orientaciones.

$$\begin{bmatrix} \overline{x^i} \\ \overline{y^i} \\ \overline{\theta^i} \end{bmatrix} = \begin{bmatrix} x^{i-1} + \overline{\Delta x^i} \\ y^{i-1} + \overline{\Delta y^i} \\ \theta^{i-1} + \overline{\Delta\theta} \end{bmatrix},\ i = n_{bucle}, \ldots, n. \tag{6.18}$$

  con

$$\begin{bmatrix} \overline{\Delta x^i} \\ \overline{\Delta y^i} \end{bmatrix} = \begin{bmatrix} \Delta fc \cdot \sin(\overline{\theta^i}) \\ \Delta fc \cdot \cos(\overline{\theta^i}) \end{bmatrix}, \; i = n_{bucle}, \ldots, n. \tag{6.19}$$

De esa forma, la orientación final de la ruta coincide con la estimada en el cierre de bucle: $\theta^n = \theta_{actual-mapa}$.

- Corrección de las posiciones XY del mapa

  Una vez corregidas las fases, se realiza una nueva corrección de la ruta dentro del bucle utilizando la posición actual estimada en el cierre de bucle $\left([x_{actual-mapa}, y_{actual-mapa}]^T\right)$. La información de partida será la ruta con las fases ya corregida $\left([\overline{x}, \overline{y}]^T\right)$.

  Considerando de nuevo *n* la posición de la última imagen de la ruta, en la que se detecta el cierre de bucle, el error en la posición se define como:

$$\begin{bmatrix} e_x \\ e_y \end{bmatrix} = \begin{bmatrix} x_{actual-mapa} - \overline{x^n} \\ y_{actual-mapa} - \overline{y^n} \end{bmatrix} \tag{6.20}$$

  De nuevo, la corrección del error será propagada por las distintas posiciones del bucle. El algoritmo asocia a cada posición una incertidumbre igual a la distancia imagen de las proyecciones, d^i_{proy}. La corrección de las distintas posiciones será proporcional a dicha incertidumbre. Obsérvese que esa distancia es obtenida con el Análisis multiescala, asociada a desplazamiento (Δfc) estimado entre escenas. La razón de utilizar esta distancia como media de incertidumbre es que, cuanto mayor sea, menos fiable es la estimación de desplazamiento entre escenas consecutivas.

  Finalmente, las nuevas posiciones de la ruta $[\,\overline{\overline{x}}, \overline{\overline{y}}\,]^T$ se calculan como:

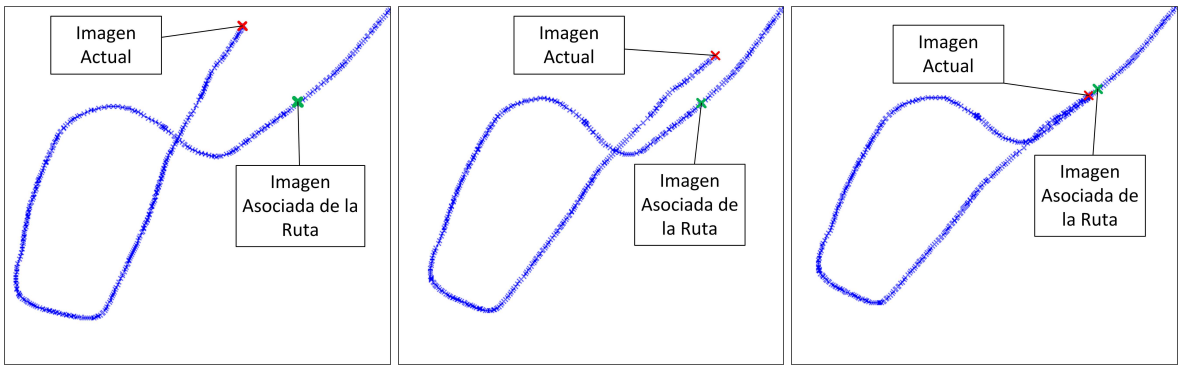
$$\begin{bmatrix} \overline{\overline{x^i}} \\ \overline{\overline{y^i}} \end{bmatrix} = \begin{bmatrix} \overline{x^{i-1}} + \overline{\overline{\Delta x^i}} \\ \overline{y^{i-1}} + \overline{\overline{\Delta y^i}} \end{bmatrix}, \; i = n_{bucle}, \ldots, n. \tag{6.21}$$

  con

$$\begin{bmatrix} \overline{\overline{\Delta x^i}} \\ \\ \overline{\overline{\Delta y^i}} \end{bmatrix} = \begin{bmatrix} \overline{\Delta x^i} + e_x \cdot \dfrac{d^i_{proy}}{\Sigma^n_{j=n_{bucle}} d^j_{proy}} \\ \\ \overline{\Delta y^i} + e_y \cdot \dfrac{d^i_{proy}}{\Sigma^n_{j=n_{bucle}} d^j_{proy}} \end{bmatrix} \tag{6.22}$$

  De esa forma, se corrigen las coordenadas (x,y) de los puntos considerados. Es importante remarcar que, tras hallar las nuevas coordenadas, los ángulos $\overline{\theta^i}$ pueden variar ligeramente. Por ello, habrá que estimarlos como:

$$\overline{\overline{\theta^i}} = \arctan\left(\frac{\overline{\overline{y^i}} - \overline{\overline{y^{i-1}}}}{\overline{\overline{x^i}} - \overline{\overline{x^{i-1}}}}\right) \tag{6.23}$$



(a) (b) (c)

Figura 6.18: Ejemplo de corrección de ruta por cierre de bucle. (a) Detección de escena por la que se ha pasado anteriormente, (b) corrección de los desfases de la odometría, y (c) corrección de las posiciones de la trayectoria.

La ruta final considerará las posiciones corregidas $\left[\overline{\overline{x^i}}, \overline{\overline{y^i}}, \overline{\overline{\theta^i}}\right]^T$ para $i = n_{bucle}, \ldots, n$.

La Figura 6.18 muestra las tres etapas seguidas por el cierre de bucle. En la Figura 6.18(a) se cumple la condición de que la distancia de la imagen actual con una anterior de la ruta queda por debajo del umbral definido 6.13. Esto indica que se ha detectado que el robot se encuentra en una posición muy cercana a otra visitada anteriormente. En ese momento, se calcula la pose del robot respecto a la imagen asociada, obteniendo $\left[x_{actual-mapa}, y_{actual-mapa}, \theta_{actual-mapa}\right]^T$. Siguiendo el proceso, la Figura 6.18(b) muestra la ruta una vez se han corregido los incrementos de fase, mientras que Figura 6.18(c) presenta la apariencia final de la ruta tras llevar a cabo la corrección de las posiciones (x,y).

Cuando se produce el primer cierre de bucle, el tramo de ruta a corregir será el comprendido entre las dos imágenes asociadas, es decir, la anterior del mapa y la actual. Sin embargo, si entre la imagen de cierre del mapa y la actual ha habido una corrección previa de la ruta (debida a un cierre anterior), se corregirá únicamente desde la última posición corregida. Por ejemplo, suponemos un primer cierre entre la imagen 50 y la 150. Si posteriormente se detecta otro cierre entre la imagen 40 y 200, la corrección de la ruta se realizará entre las posiciones 150 y 200. Es decir, $n_{bucle} = 150$. Por el contrario, si el nuevo cierre se asocia las imágenes 160 y 200, n_{bucle} sería igual a 160.

Por último, para evitar llevar a cabo correcciones muy seguidas en el caso que se esté navegando por encima de una trayectoria seguida anteriormente, se fija un número mínimo de escenas para aplicar una nueva corrección de la ruta. Este parámetro es variable, y en los experimentos se utilizará un número mínimo de 5 imágenes antes de aplicar una nueva corrección.

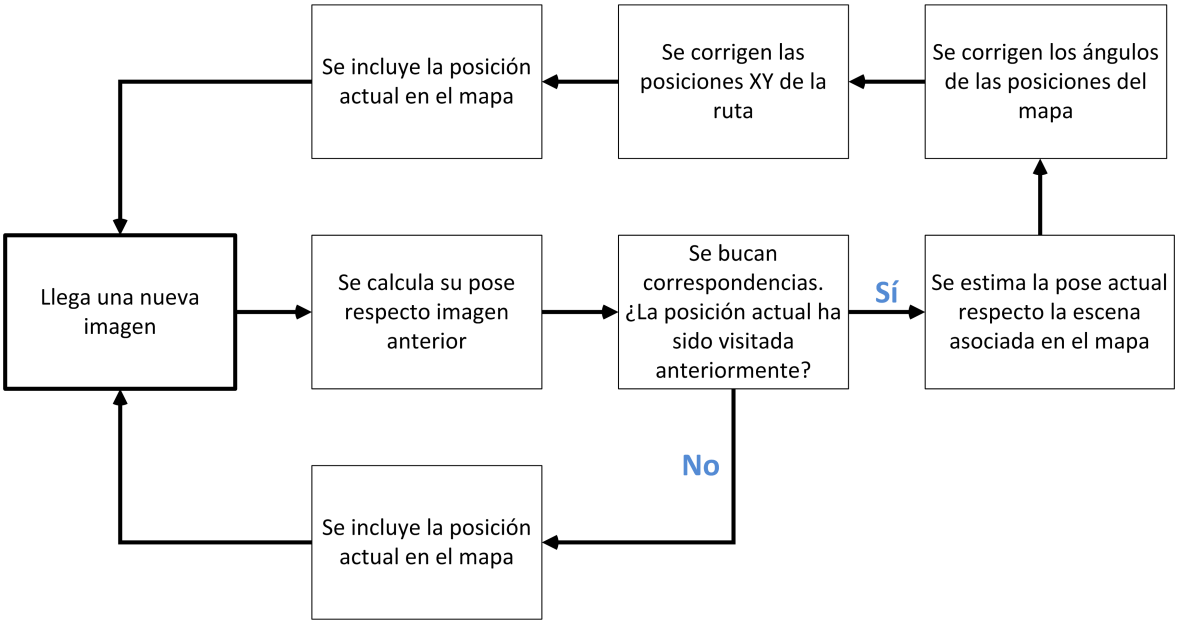


Figura 6.19: Esquema del algoritmo de estimación de la ruta usando la odometría visual topológica.

En la Figura 6.20 se puede encontrar el esquema seguido por el algoritmo cada vez que llega una nueva imagen.

6.2.3 Experimentos y resultados

En este punto, se describen los parámetros y descriptores utilizados en la parte experimental, la base de imágenes usada, y los resultados obtenidos en las simulaciones.

6.2.3.1 Selección del Descriptor y parámetros del Análisis Multiescala

En este algoritmo se utilizan descriptores en dos tareas distintas. La primera de ellas es caracterizar las imágenes de la ruta, con el objetivo de estimar desfases entre escenas consecutivas y permitir su asociación en el cierre de bucle. La segunda, es realizar el análisis multiescala, describiendo las distintas proyecciones que se realizan.

Para la primera tarea, como utiliza la proyección panorámica de la escena omnidireccional, es posible utilizar los datos experimentales obtenidos en el Capítulo 5 para realizar la elección del descriptor. Por compromiso entre coste computacional y precisión, escogemos dos descriptores: la Firma de Fourier sobre HSV, y HOG añadiendo el Histograma de Color.

La escena panorámica tiene un tamaño de 128×512 píxeles. Los parámetros relativos a la localización se variarán en la parte experimental junto con el umbral de cierre de bucle (th_{pano}). Sin embargo, los parámetros de desfase son fijos. Para Fourier sobre HSV, el número

de elementos de fase por fila (N_{rot}) es igual a 32. En cuanto a HOG, la anchura de la celda vertical (S_V) es igual a 16 píxeles, con una distancia entre ventanas consecutivas (D_V) de 2 píxeles.

En cuanto al Análisis Multiescala, el descriptor utilizado es la Firma de Fourier sobre escala de grises. La razón de elegir este descriptor es que se prioriza el tiempo de cálculo. Además, este descriptor sólo se utiliza para la correlación entre las distintas ampliaciones proyectivas, lo cual no es un proceso muy exigente en cuanto a necesidad de distinción de la escena.

El rango de focales utilizadas para obtener las proyecciones es $fc \in [1{,}0 - 1{,}2]$, con un paso entre focales consecutivas de $\Delta fc_{min} = 0{,}01$. El tamaño de la imagen proyectiva es de 256×256 píxeles. Se escoge este tamaño para limitar el coste computacional derivado del procesamiento de la imagen al estimar su descriptor. Es posible destacar que la altura de la proyección es también un parámetro variable. Se ha calibrado para recoger la parte central de la escena, ya que suele contener información más distintiva del entorno.

6.2.3.2 **Base de Imágenes**

La base está compuesta por una ruta de imágenes capturada en la planta baja del Edificio Innova de la Universidad Miguel Hernández. El número de escenas incluidas es de 1211. Las escenas han sido capturadas por un conjunto catadióptrico formado por la cámara DFK-41BF02 y el espejo EizohWide 70 (Sección 3.1.1), obteniendo imágenes omnidireccionales en color de resolución 1280×960 píxeles.

Para la adquisición, se ha utilizado el robot Pioneer P3-AT (Sección 3.4.1). El modelo utilizado estaba equipado con un medidor de distancias laser, además del sistema de visión catadióptrico.

En la Figura 6.20 se muestra el recorrido de la ruta, junto con algunos ejemplos de escenas panorámicas obtenidas en distintas estancias. En concreto, el robot navega por un laboratorio y distintos pasillos dentro de las zonas comunes del edificio. Son constantes los ventanales, que provocan cambios importantes en la iluminación. Además, el recorrido incluye una pendiente, que se recorre tanto en sentido ascendente como descendente.

La ruta que utilizaremos para comparar los resultados de las simulaciones, nuestro *Ground Truth*, ha sido obtenida utilizando la información de odometría del robot y los datos del laser a través del programa *gmapping*.

6.2.3.3 **Resultados de estimación de la ruta**

Los experimentos llevados a cabo consisten en la estimación de la ruta seguida por el robot únicamente la información visual a través del sistema de odometría topológica usando

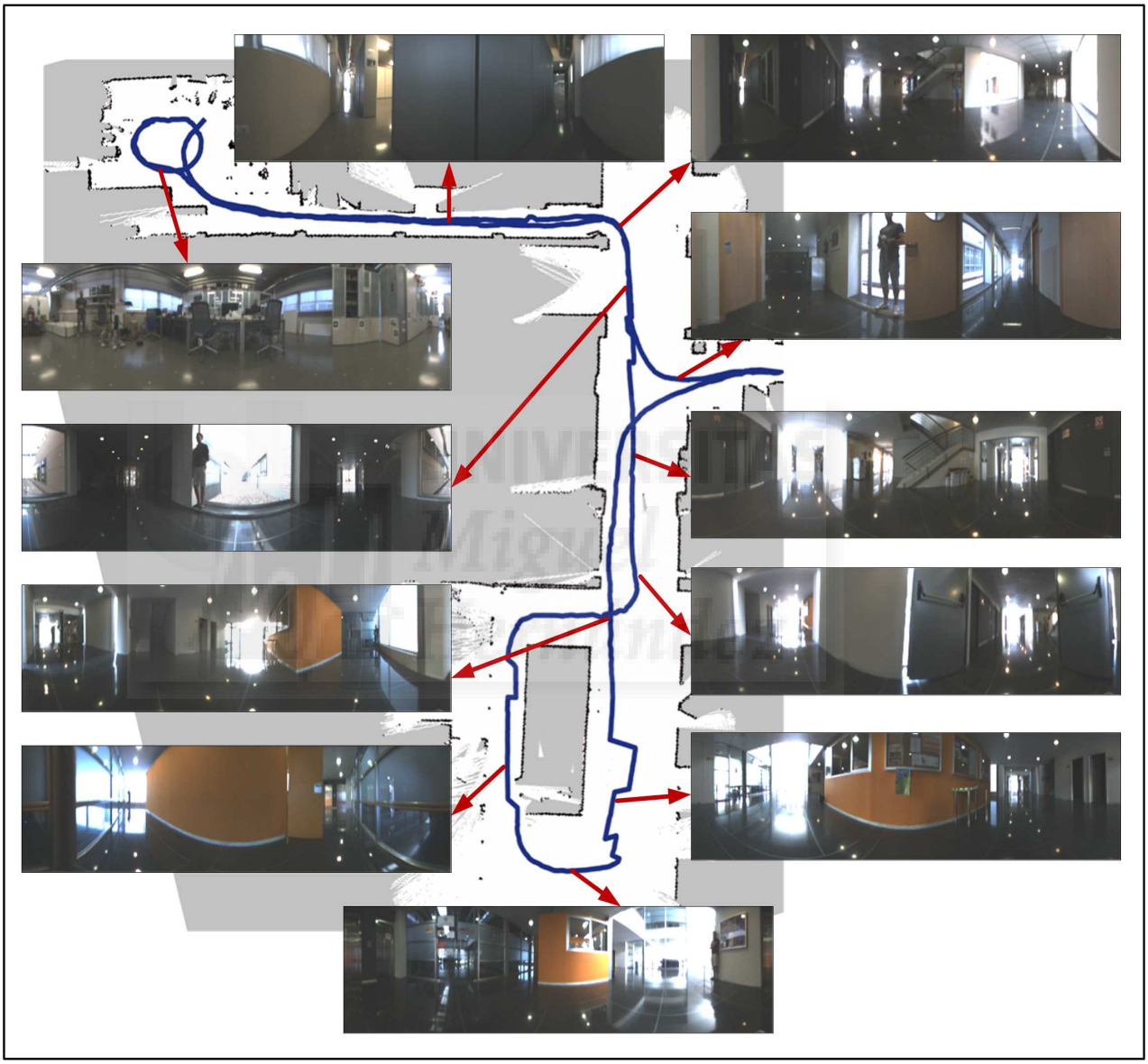
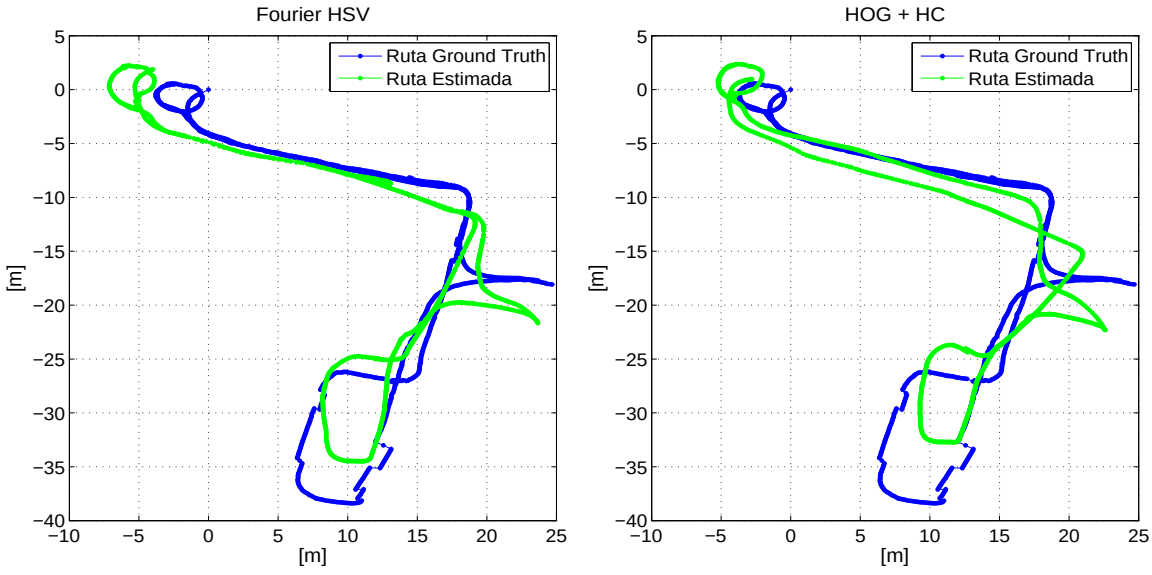


Figura 6.20: Ruta seguida por el robot, junto con ejemplos de escenas de los distintos entornos.



(a) (b)

Figura 6.21: Representación gráfica de la ruta de referencia y la ruta estimada usando (a) Fourier HSV y (b) HOG con HC.

información omnidireccional.

Para medir el error, se usa análisis Procrustes [50, 100], visto anteriormente en la Sección 6.1.4.3. Indicar que ll análisis Procrustes devuelve un índice de disparidad estandarizado $\mu \in [0,1]$. Cuanto menor sea μ, más similar es el grafo obtenido a la distribución real.

Fijados los parámetros de estimación de fase para los descriptores de las imágenes panorámicas, se estudia la variación del umbral al realizar asociaciones de cierre de bucle (th_{pano}), y el parámetro de localización de cada descriptor. En el caso de la Firma de Fourier, este parámetro corresponde al número de elementos seleccionados por fila (N), mientras que para HOG corresponde al número de celdas horizontales utilizadas (C_H)

En las Tablas 6.4 y 6.5 se presentan los resultados para cada descriptor. Los resultados con mínimo error se encuentran marcados en negrita.

Si se utiliza un umbral demasiado bajo, no se asociarán imágenes. Es decir, no se harán cierres de bucle que corrijan la trayectoria, por lo que el resultado final estará basado únicamente en la odometría visual topológica, acumulando el error en la estimación en cada paso. Por contra, si el el umbral es demasiado alto, obtendremos asociaciones con imágenes que realmente no están cerca, por lo que el análisis multiescala aplicado sobre esas asociaciones proporcionará una estimación de posición erróneo.

En la Figura 6.21 se incluyen las gráficas de resultados asociada a los parámetros marcados en negrita en las Tablas 6.4 y 6.5. En ellas se representa la estimación de la ruta usando los distintos descriptores, comparándola con la con la ruta de referencia.

Tabla 6.4: Error en la estimación de la ruta medido con análisis Procrustes (μ) variando el umbral de cierre de bucle (th_{pano}) y el parámetro de localización del descriptor (N) usando la Firma de Fourier sobre HSV.

th_{pano}	N	μ
2,30	8	0,0505
2,30	16	0,0735
2,30	32	0,0574
2,50	8	0,0565
2,50	16	0,0948
2,50	32	0,0578
2,90	8	0,0485
2,90	16	0,0449
2,90	32	0,0935
3,10	8	0,0447
3,10	16	0,0550
3,10	32	0,0949
3,30	**8**	**0,0383**
3,30	16	0,0532
3,30	32	0,0424
3,50	8	0,4656
3,50	16	0,0744
3,50	32	0,0539

Tabla 6.5: Error en la estimación de la ruta medido con análisis Procrustes (μ) variando el umbral de cierre de bucle (th_{pano}) y el parámetro de localización del descriptor (C_H) usando la HOG junto con Histograma de Color (HC).

th_{pano}	C_H	μ
0,020	32	0,1181
0,020	16	0,1181
0,020	8	0,1182
0,022	32	0,0684
0,022	**16**	**0,0582**
0,022	8	0,0610
0,024	32	0,0773
0,024	16	0,0687
0,024	8	0,0691
0,026	32	0,0942
0,026	16	0,0732
0,026	8	0,0670
0,028	32	0,0823
0,028	16	0,0773
0,028	8	0,0757
0,030	32	0,0841
0,030	16	0,0905
0,030	8	0,0814

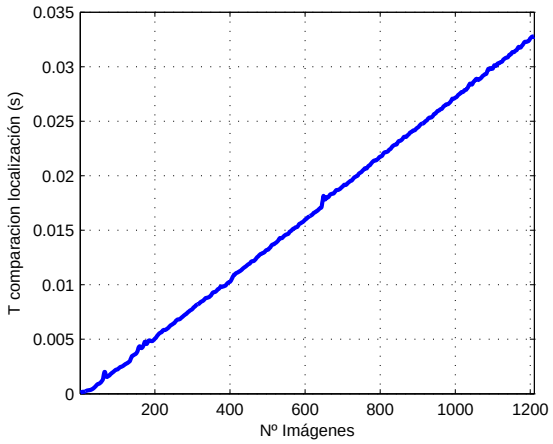


Figura 6.22: Tiempo de comparación entre descriptores para localización variando el número de imágenes incluidas en la base.

Tal y como indican los resultados numéricos, puede observarse que al usar Fourier (Figura 6.21 (a)) la estimación de la ruta es más aproximada a la referencia. En la trayectoria rectilínea de la parte superior, Fourier realiza un mayor número de cierres de bucle, por lo que el resultado es mejor que en el caso de HOG.

Por último, cabe destacar que se planteó construir un mapa que no incluyese todas las imágenes de la ruta, sino sólo aquellas más significativas, entendiendo como imágenes más significativas aquellas en las que se produce cambios de orientación importante, y las que tienen una apariencia menos similar a las otras incluidas en el mapa. Los cierres de bucle, por tanto, se realizarían a partir de estas imágenes de referencia.

El sentido de este mapa era reducir su tamaño por coste computacional. Sin embargo, los descriptores basados en apariencia global son muy compactos, por lo que su almacenamiento en un mapa requiere gran cantidad de memoria. Además, al tener pocos elementos, el tiempo de cálculo de las distancias imagen también es reducido. Como ejemplo, en la Figura 6.22 se muestra el tiempo necesario para comparar los descriptores de localización de las imágenes panorámicas usando Fourier con $N = 32$ elementos por fila.

Como se puede apreciar, en el caso de usar el número máximo de imágenes, el tiempo requerido es inferior a 0.035 segundos. Además, al usar todas las imágenes, los cierres de bucle pueden ser más precisos, por lo que la ventaja de usar todas las escenas es mayor que el inconveniente de incremento de tiempo.

6.3 Conclusiones

En este capítulo se ha presentado un método de estimación de distancia topológica entre dos escenas usando la apariencia global de las imágenes.

En concreto, se estudia su aplicación sobre imágenes proyectivas capturadas con una cámara con lente de ojo de pez, y sobre imágenes omnidireccionales obtenidas con un sistema de visión catadióptrico.

En el primer caso, las conclusiones a las que se llega son:

- Los descriptores basados en apariencia global también pueden ser aplicados sobre escenas no omnidireccionales obteniendo una capacidad de distinción y caracterización de las escenas satisfactoria para su aplicación en tareas de navegación.

- Es posible reducir la resolución de las escenas antes de aplicar los distintos descriptores sin reducir la precisión en la asociación de imágenes de forma significativa. Los descriptores obtenidos siguen presentando una alta precisión, pero los requisitos computacionales se reducen notablemente, sobre todo el tiempo de cálculo. Destacan los resultados de los descriptores GIST, mientras que HOG muestra la mayor dependencia respecto a la resolución de la imagen.

- El Análisis Multiescala permite obtener un indicador de distancia entre dos escenas capturadas en la dirección de avance del robot. Este indicador proporciona información de la magnitud y sentido del desplazamiento entre las dos imágenes.

- Este sistema puede utilizarse en la construcción de un mapa topológico, cuya distribución de nodos es similar a la real.

- Además, también puede ser utilizado para la mejora de la localización dentro del mapa toplógico. Así pues, la localización del mapa no sólo se reduce a las posiciones de los mapas, sino también a localizaciones intermedias.

En el segundo caso, se muestra la aplicación de este Análisis Multiescala sobre información omnidireccional. Las principales conclusiones son:

- Las múltiples posibilidades de vistas que se pueden obtener a partir de las escenas omnidireccionales permite aplicar la aplicación del Análisis Mutiescala sobre este tipo de imágenes. En concreto, se usan escenas proyectivas obtenidas con diferentes distancias focales, lo que proporciona diferentes ampliaciones de las vistas.

- Es posible combinar la información obtenida a partir de la proyecciones panorámicas con la comparación multiescala para obtener un sistema de odometría visual topológica utilizando únicamente la apariencia global de las escenas, ya que se dispone de información de magnitud del desplazamiento y dirección.
- Este sistema de odometría puede ser utilizado en la estimación del camino seguido por un robot en una ruta a partir de las imágenes omnidireccionales capturadas. Además, se puede introducir en el algoritmo el reconocimiento de zonas por las que el robot haya navegado anteriormente. De esa forma, se llevan a cabo cierres de bucle que permiten mejorar la estimación inicial de la ruta.
- Los resultados experimentales muestran que el sistema propuesto es capaz de reconstruir la ruta seguida con un error aceptable.

CAPÍTULO

7

Estimación Topológica de Altura.

Este capítulo tiene como objetivo extender el uso de los descriptores basados en apariencia global de imágenes en aplicaciones en las que existan cambios de altura del robot móvil.

Como aportación de la tesis, se presentan distintas técnicas orientadas a la obtención de un indicador proporcional a la diferencia de altura usando la información de escenas omnidireccionales.

Los métodos utilizados se basan en la comparación de secciones de las imágenes, en la extracción de información en el dominio frecuencial, o en el uso de la geometría epipolar para poder estimar la diferencia de altitud entre dos escenas distintas capturadas en un mismo punto del plano del suelo.

Como se podrá comprobar, cada técnica está basada en una proyección distinta de la información visual, y el indicador obtenido es diferente en cada caso.

En concreto, haremos uso de la transformada panorámica (Sección 3.2.2), la vista de pájaro o proyección ortográfica (Sección 3.2.4) y la proyección sobre la esfera unitaria (Sección 3.2.1).

El descriptor de la imagen depende de la proyección de la información visual empleada. En este análisis nos hemos centrado en el uso de técnicas basadas en la transformada de la imagen al dominio frecuencial.

El estudio incluido en este capítulo parte de la suposición de que el robot mantiene su inclinación constante en todo momento, y por tanto el eje del sistema visual catadióptrico permanece perpendicular al plano del suelo.

Los experimentos han sido realizados usando nuestra propia base de imágenes, incluyendo escenas capturadas en ambientes tanto de exterior como de interior, bajo condiciones de iluminación distintas y en entornos diferenciados.

7.1 Estimación Topológica Altura.

Las distintas técnicas que se detallan a continuación tienen como objetivo proporcionar un indicador del desplazamiento relativo en altura entre dos escenas capturadas en un mismo punto en el plano XY. En concreto, estos indicadores aportan información de la magnitud y sentido del movimiento vertical.

Se han estudiado cuatro métodos distintos. Cada uno de ellos trata de extraer la diferencia entre dos escenas utilizando distinta información.

El primero de ellos busca la mejor correspondencia entre la la parte central de la imagen panorámica con distintos desplazamientos verticales. El segundo se basa en la variación de la información en espacio frecuencial a partir de la fase de la transformada de Fourier Bidimensional. El tercer método trata de aplicar el Análisis Multiescala vista en la Sección 6.1.1. Para ello, se emplea la proyección ortográfica de la escena. Por último, el cuarto algoritmo utiliza la geometría epipolar para simular desplazamientos en el sistema de referencia de la cámara, y con ello, estimar el desplazamiento vertical más probable entre dos imágenes distintas.

7.1.1 Correlación de la Celda Central en Imágenes Panorámicas

La mayor parte de los algoritmos que emplean la información de imágenes omnidireccionales, utilizan la proyección cilíndrica de la información visual, es decir, la imagen panorámica.

En una imagen panorámica, la información más significativa suele situarse en las filas centrales, especialmente en ambientes de exterior, donde los ángulos inferiores corresponden al suelo, y los altos recogen la proyección de rayos correspondientes al cielo.

Otra característica de las filas centrales de la imagen panorámica es que, ante un cambio de altura del robot, representan la zona de la imagen que menos probabilidad tiene de caer fuera del campo de visión del sistema visual.

Tomando todo esto en cuenta, se propone comparar la celda correspondiente a la parte central entre dos imágenes panorámicas para estimar su altitud relativa.

Para ello, se calcula el descriptor asociado a la celda central de la imagen panorámica usando la apariencia global, y se repite el proceso para celdas situadas por encima y por debajo del la posición inicial. La Figura 7.1 muestra un ejemplo de selección de celdas de una imagen panorámica. La celda central aparece con un mayor grosor de línea. También se puede apreciar el resto de celdas adicionales situadas por encima y por debajo de la central.

Dada una nueva imagen, estimaremos el descriptor de la celda central, y lo comparamos sin los descriptores de todas las celdas de la primera imagen.

Figura 7.1: Selección de celdas sobre una escena panorámica para estimación de altura usando la Correlación de la Celda Central.

Buscando la asociación con menor distancia imagen (es decir, con menor distancia Euclídea entre descriptores), el algoritmo identifica la celda de la primera imagen que muestra una apariencia más similar con la celda central de la segunda. La altura (d) asociada a la celda de la primera imagen respecto de la central proporciona un indicador del desplazamiento vertical entre las escenas. El desplazamiento, por tanto, se mide en píxeles.

7.1.2 Desfase Vertical usando FFT2D

En la Sección 4.1.3 se ha presentado la Transformada de Fourier Bidimensional. Aprovechando las propiedades de la transformada de Fourier, se puede estimar la rotación circular en el orden de las filas y las columnas de una matriz de entrada. Tal y como demuestra el Teorema del Desplazamiento (Ecuación 4.9), un desplazamiento en el orden de las filas o las columnas produce un cambio en la fase de los coeficientes de la transformada.

Cuando se trabaja con imágenes panorámicas, una rotación del robot en el plano del suelo produce un desplazamiento de las columnas de la escena. Gracias a este teorema somos capaces de estimar la rotación entre dos imágenes rotadas que han sido capturadas en un mismo punto.

El propósito del método que se describe en este punto es extender el Teorema del Desplazamiento para estimar el desfase en las filas de las imágenes producido por un desplazamiento vertical. Sin embargo, la extrapolación ante un cambio de altura no es directa. A diferencia de una rotación alrededor del eje del sistema catadióptrico, un cambio en la altitud del robot no produce únicamente un desplazamiento de la información de la imagen panorámica, sino que también se modifica la información contenida en la escena.

Ante un movimiento vertical, parte de la información de la imagen en la primera posición queda fuera del nuevo campo de visión de la cámara, al igual que se introduce otra nueva. Dependiendo de la dirección del desplazamiento, las filas superiores o inferiores de la escena desaparecen, mientras que se añaden otras nuevas.

Así pues, no se trata exactamente de una rotación circular de las filas de la imagen. Al modificar la información contenida en la escena, se introducen cambios en los coeficientes de la transformada de Fourier que harán que no se cumpla exactamente el Teorema del Desplazamiento.

Además, si se produce un cambio en la orientación de la cámara al mismo tiempo que en su altitud, el efecto sobre la fase de los coeficientes de Fourier solaparán sus efectos, siendo difícil discernir si el cambio de fase en los coeficientes de Fourier se debe a la rotación o al cambio de altura.

Sin embargo, la mayor parte de la información representada en la escena panorámica sufrirá un desplazamiento vertical, equivalente a una rotación en el orden de las filas de la imagen. Por ello, para estimar el desplazamiento de dos imágenes capturadas en un mismo punto (x,y), se plantea utilizar la fase de los coeficientes de Fourier. En concreto, el algoritmo usa la fase de una submatriz de tamaño $N_F \times N_F$ que recoge los primeros coeficientes, denotada como $ph(F_{N_F \times N_F})$.

Como se ha comentado, un desplazamiento en el dominio espacial produce un cambio de fase de los coeficientes en dominio frecuencial. Es posible simular matemáticamente ese efecto en los coeficientes de Fourier. Siendo R la rotación circular de las filas de la matriz de entrada medida en grados, los coeficientes de fase $ph(F_{N_F \times N_F})$ pueden estimarse como:

$$ph(F_{N_F \times N_F})_R = ph(F_{N_F \times N_F}) + R \cdot MRV \tag{7.1}$$

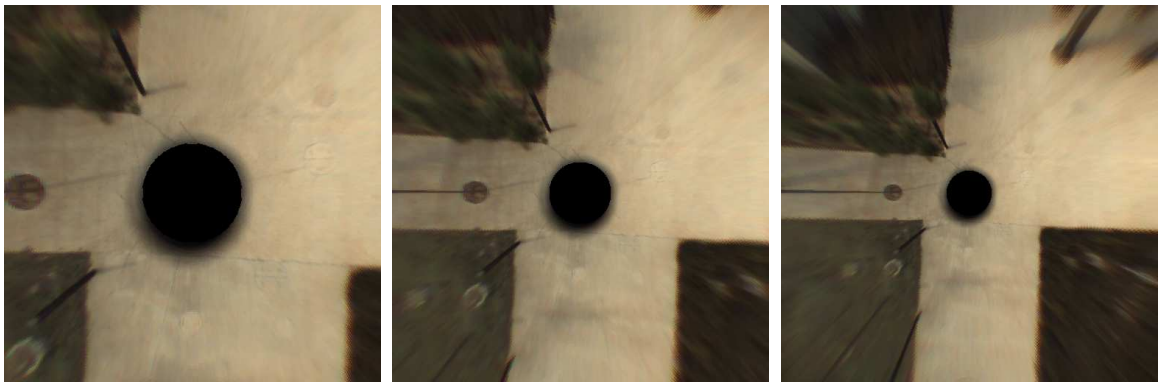
siendo MRV la Matriz de Rotación Vertical, que se define como:

$$VRM = \begin{pmatrix} 0 & 0 & \cdots & 0 \\ 1 & 1 & \cdots & 1 \\ 2 & 2 & \cdots & 2 \\ \vdots & \vdots & \ddots & \vdots \\ N_F & N_F & \cdots & N_F \end{pmatrix}_{N_F \times N_F} \tag{7.2}$$

Dada una imagen de referencia, se estima $ph(F_{N_F \times N_F})_R$ para
$R = [-180, -180 + \Delta R, \ldots, 180]$. En los experimentos, se elige un $\Delta R = 0{,}5$.

Para determinar la altitud relativa entre dos escenas, comparamos los coeficientes de fase $ph(F_{N_F \times N_F})$ de la imagen de la cual se desea obtener el desplazamiento con los distintos $ph(F_{N_F \times N_F})_R$ de la imagen de referencia.

El R para el cual la diferencia de coeficientes de fases es mínima, denota la altitud relativa entre imágenes.



(a) fc=5 (b) fc=8 (c) fc=11

Figura 7.2: Ejemplo de proyección ortográfica utilizando distancias focales distintas.

7.1.3 **Análisis Multiescala sobre la Vista Ortográfica**

Con esta técnica, se propone hacer uso del Análisis Multiescalas para estimar el desplazamiento vertical entre dos escenas.

Como se ha visto en el Capítulo 6, es posible obtener una medida del desplazamiento entre dos imágenes, utilizando ampliaciones artificiales de la parte central de las escenas. Para ello, la proyección sobre la que se aplica el Análisis Multiescala debe ser perpendicular a la dirección de movimiento (Figura 6.1).

Como se considera un desplazamiento vertical del robot, la proyección elegida debe ser la vista ortográfica. La obtención de las distintas escalas (o ampliaciones) de la imagen se realiza mediante la proyección de la información visual sobre planos perpendiculares al eje del sistema catadióptrico a distintas distancias.

En la Figura 7.2 se muestran ejemplos de la misma proyección ortográfica obtenida con tres distancias focales diferentes.

Después de obtener las proyecciones ortográficas, se describe la imagen usando un descriptor de apariencia global.

El algoritmo sigue los siguientes pasos: Primero, se obtienen distintas proyecciones de la imagen de referencia usando diferentes distancias focales, y se calculan sus descriptores. Cuando llega una imagen nueva a comparar, se obtiene una vista ortográfica con distancia focal fija, y se compara su descriptor con los estimados para la imagen de referencia.

La asociación elegida será la que presente una menor distancia imagen. La diferencia entre la focal de la imagen de referencia seleccionada y la focal de la imagen de test proporciona la información del desplazamiento vertical.

En los experimentos, se usa una focal para la imagen de referencia fc_{ref} comprendida entre 4 y 11, mientras que la focal de la imagen de test se fija a $fc_{test} = 7$.

7.1.4 Cambio de Coordenadas del Sistema de Referencia de la Cámara (SRC)

En [191], sus autores muestran que, dada una imagen, se puede simular un desplazamiento del sistema de referencia (SRC) de la cámara usando la geometría epipolar. De esta forma, se obtiene una nueva proyección de la información visual original que refleja el movimiento simulado en la cámara.

Para ello, primero se estiman las coordenadas de la imagen en el mundo real. $m = [m_{x_{pix}}, m_{y_{pix}}]$ son las coordenadas de los píxeles con respecto el centro de la imagen omnidireccional. La función $f(\rho)$ (Ecuación 3.5) obtenida con la calibración del sistema visual nos permite conocer la dirección de incidencia de los rayos en el sistema catadióptrico a partir de las coordenadas en la imagen (m). Con ello, se es posible determinar las coordenadas de la imagen sobre la esfera unitaria $M \in \mathbb{R}^3$.

Una vez obtenidas estas coordenadas, se aplica el cambio en el sistema de referencia de la cámara, que queda como:

$$M' = M + \rho \cdot T, \tag{7.3}$$

siendo T el vector de desplazamiento unitario, y ρ el factor de escala proporcional al desplazamiento del SRC.

Como en estos experimentos se estudian únicamente desplazamientos verticales, el vector T tendrá la dirección del eje z: $T = [0,0,1]^T$.

Una vez obtenidas las nuevas coordenadas imagen en el mundo real (M'), es posible obtener las nuevas coordenadas en el plano imagen (m').

Mediante la asociación de las coordenadas de m con las nuevas coordenadas en la escena m', se obtiene la nueva imagen omnidireccional, que recoge el movimiento del SRC.

Hay que tener en consideración que, en la asociación entre m y m', algunas coordenadas del nuevo sistema pueden caer fuera del plano imagen, y otros píxeles podrían quedar sin valor asociado. Para los píxeles sin asociación, se realiza una interpolación con los 8 vecinos.

Nótese que, tras obtener la nueva imagen omnidireccional, es posible volver a obtener distintas proyecciones de la imagen. Específicamente, en este estudio se incluye la vista ortográfica, la imagen panorámica, y la proyección sobre la esfera unitaria.

La Figura 7.3 muestra el esquema de desplazamiento del sistema catadióptrico con la proyección de un punto del espacio, junto con dos ejemplos de imágenes omnidireccionales y su transformada panorámica suponiendo dos movimientos distintos.

Para estimar la diferencia de altura entre dos escenas, se simulan distintos desplazamientos de la imagen de referencia modificando ρ, y se compara con la imagen de test, la cual tiene $\rho = 0$, es decir, no se le aplica ningún desplazamiento.

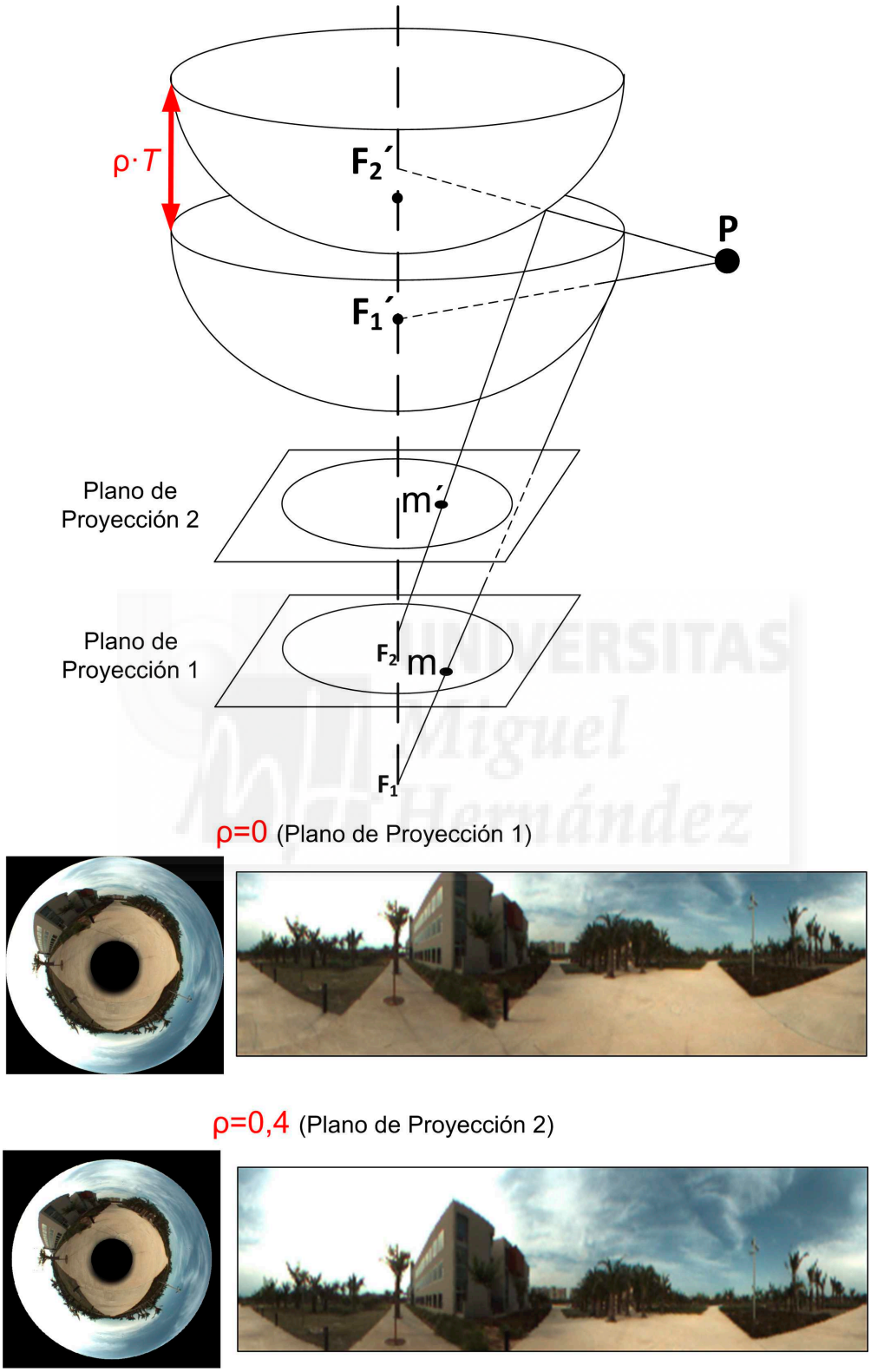


Figura 7.3: Esquema de proyección de un punto al variar el sistema de referencia de la cámara (SRC) usando la geometría epipolar, junto con ejemplo de imagen omnidireccional aplicando desplazamiento vertical, y su transformada panorámica.

Por último, el algoritmo selecciona la comparación con menos distancia imagen. El ρ asociado al desplazamiento de la imagen de referencia seleccionada indicará el desfase vertical entre imágenes.

En los experimentos, los valores de ρ varían de -0.3 a 0.3 para la imagen de referencia, siendo nulo para la imagen de test.

7.2 Base de Imágenes

Con el objeto de comprobar el funcionamiento y eficacia de los algoritmos incluidos en la sección anterior, se ha adquirido una base de datos propia con imágenes de exterior e interior. Para ello, se ha utilizado el sistema catadióptrico detallado en la Sección 3.1.1. En concreto, la cámara utilizada es el modelo DFK-41BF02, por lo que las imágenes obtenidas son omnidireccionales en color con resolución 1280x960 píxeles.

La cámara ha sido acoplada al trípode que se describe en la Sección 3.4.2, permitiéndo adquirir imágenes con un rango de 165 cm en altitud.

Las imágenes han sido capturadas en 10 posiciones distintas en exterior, y 11 en interior. Para exterior, se han adquirido 12 imágenes en altura distinta por posición. La altura mínima es de 125 cm, y la máxima de 290 cm, con un paso de 15 cm entre escenas consecutivas. En interior, no ha sido posible llegar a adquirir imágenes a altura máxima en todas las localizaciones debido a que la altura libre de algunas estancias no lo han permitido.

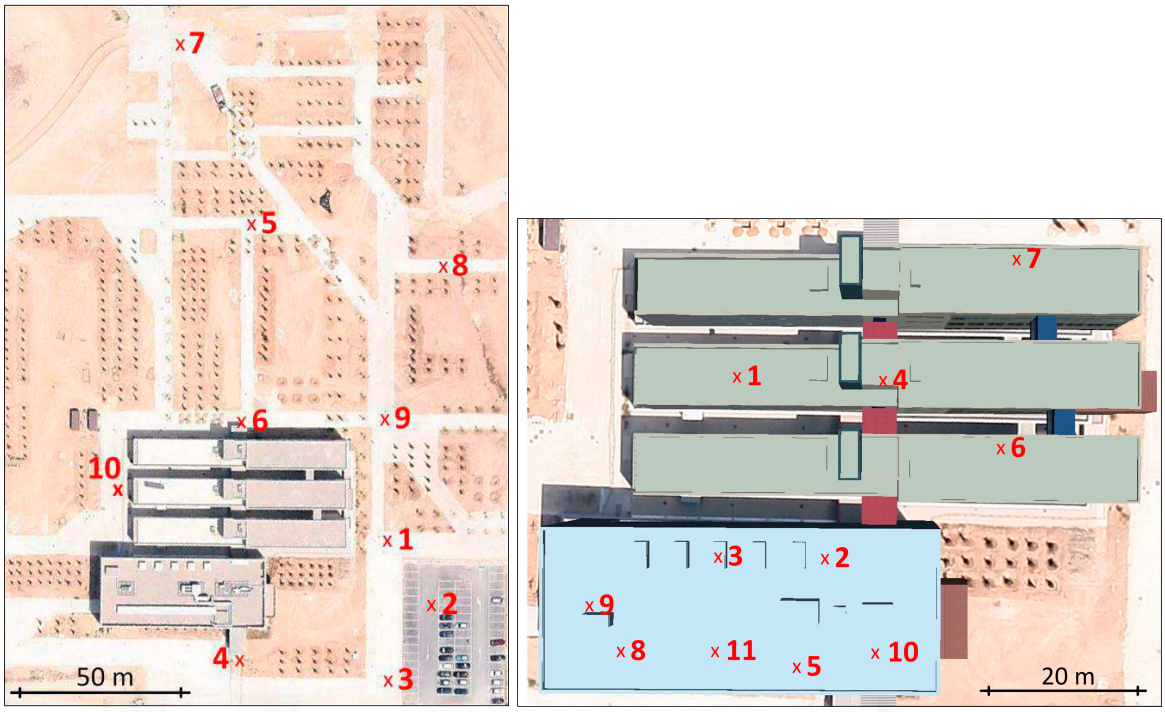
La Tabla 7.1 muestra las alturas de adquisición, y el número de imágenes por altura en la base de exterior e interior. En la Figura 7.4 es posible ver la localización los distintos puntos de captura tanto para las imágenes de exterior como para las de interior.

Tabla 7.1: Altura de cada imagen respecto al plano del suelo, y número de imágenes de cada base por altura.

h	**Altura Imagen (cm)**	**Imágenes Exterior**	**Imágenes Interior**
1	125	10	11
2	140	10	11
3	155	10	11
4	170	10	11
5	185	10	11
6	200	10	11
7	215	10	11
8	230	10	10
9	245	10	8
10	260	10	6
11	275	10	6
12	290	10	5
TOTAL IMÁGENES		**120**	**112**

Las capturas se han realizado a distintas horas para variar las condiciones de iluminación.

La base de exteriores incluye imágenes cerca y lejos de edificios, un parking, y zonas ajardinadas. En la Figura 7.5 se muestran ejemplos pertenecientes a distintas localizaciones.



(a) (b)

Figura 7.4: Plano de localizaciones de las imágenes de (a) exterior y (b) interior capturadas a distintas alturas.

Por su parte, la Figura 7.5 incluye diferentes alturas de una misma localización.

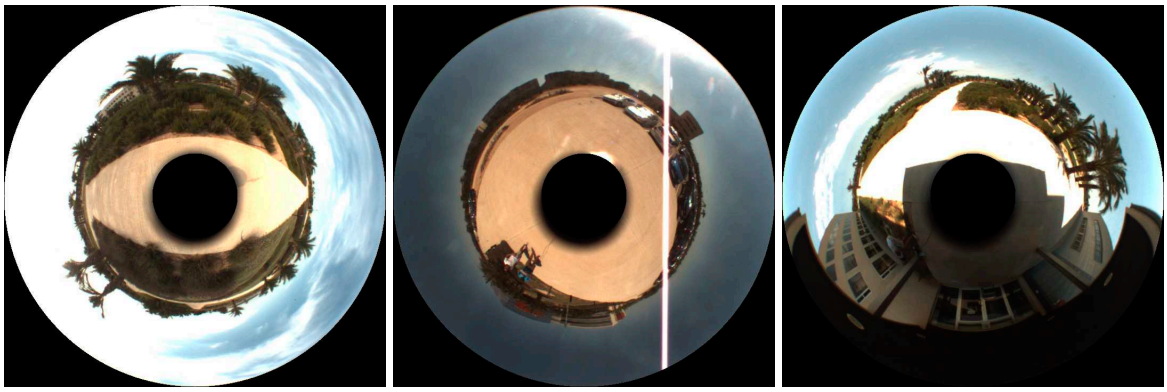
Las escenas de interior han sido adquiridas en distintas estancias del Edificio Innova de la Universidad Miguel Hernández de Elche, incluyendo un laboratorio, pasillos y otras zonas comunes de la planta baja.

En la Figura 7.7 se presentan ejemplos de tres localizaciones distintas de la base de interior, mientras que la Figura 7.8 recoge las imágenes omnidireccionales de una misma posición en distintas alturas. En estas últimas figuras es posible ver como, al aumentar la altura, gran parte de la información visual se reduce al techo.

Como se ha comentado anteriormente, las técnicas de estimación de altura hacen uso de distintas proyecciones de las escenas omnidireccionales. Para ello, ha sido necesaria la calibración del sistema visual.

La transformada panorámica tiene una resolución de 256×1024 píxeles, mientras que la vista de pájaro tiene un tamaño de 256×256 píxeles.

Durante la captura de las imágenes, no se varía la orientación ni la posición respecto al plano del suelo del sistema visual, sólo su altura. Sin embargo, ha sido inevitable pequeños desfases entre imágenes consecutivas y movimientos en la posición XY en el plano del suelo al capturar las imágenes de una misma localización.



(a) (b) (c)

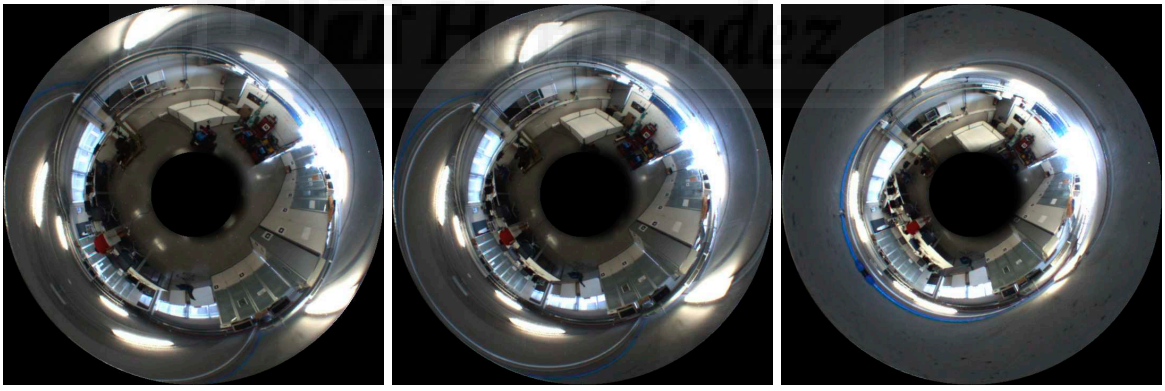
Figura 7.5: Ejemplos de imágenes omnidireccionales capturadas en el entorno de exterior en tres localizaciones distintas variando la posición relativa con los edificios y las condiciones de iluminación .

(a) (b) (c)

Figura 7.6: Ejemplos de imágenes omnidireccionales capturadas en el entorno de exterior en la misma localización variando su altura. (a) Altura 125 cm ($h = 1$), (b) altura 200 cm ($h = 6$) y (c) altura 290 cm ($h = 12$).

(a) (b) (c)

Figura 7.7: Ejemplos de imágenes omnidireccionales capturadas en el entorno de interior con distintas estancias.



(a) (b) (c)

Figura 7.8: Ejemplos de imágenes omnidireccionales capturadas en el entorno de interior en la misma localización variando su altura. (a) Altura 125 cm ($h = 1$), (b) altura 185 cm ($h = 5$) y (c) altura 275 cm ($h = 11$).

7.3 Experimentos y Resultados

Para continuar, se incluyen los resultados experimentales de la estimación de altura topológica con las distintas técnicas desarrolladas, mostrando por separado los resultados obtenidos con la base de imágenes de exterior y la de interior.

En concreto, se han llevado a cabo cuatro experimentos distintos. En el primero de ellos, se estima la altura de las distintas imágenes tomando como referencia la escena situada en la altura más baja ($h = 1$) para cada posición.

El segundo experimento es análogo al primero, pero se toma como imagen de referencia la escena a 180 cm, es decir, $h = 5$.

Por último, se estudian los resultados para distintos gradientes de desplazamiento. Así pues, en cada localización se realizan tantas comparaciones como sea posible tomando distintas altitudes de la imagen de comparación. Por ejemplo, para un gradiente $\Delta h = 2$, que equivale a 30 cm, se compara la primera imagen con la tercera, la segunda con la cuarta, la tercera con la quinta,... Y así sucesivamente hasta realizar todas las combinaciones que permite el rango de alturas. La Tabla 7.2 muestra el número de comparaciones para cada valor de gradiente y base.

El cálculo de la altura con distintos gradientes se realiza tanto para gradientes positivos como para gradientes negativos.

Tabla 7.2: Diferencia de altura para los distintos gradientes de imágenes y número de comparaciones posibles para la base de exterior e interior.

Δh	**Incremento Altura (cm)**	**Comparaciones Exterior**	**Comparaciones Interior**
2	30	100	90
4	60	80	68
6	90	60	46
8	120	40	25

Nótese que se ha supuesto que todas las imágenes comparadas para la estimación de altura tienen la misma orientación. Aunque las imágenes empleadas en estos experimentos tienen un desfase entre sí que suponemos nulo, en un sistema integrado de localización y navegación podría corregirse el desfase entre escenas, pues se ha demostrado la capacidad de localización y estimación de orientación utilizando la información omnidireccional en los capítulos 5 y 6.

Los descriptores utilizados para describir las escenas panorámicas y ortográficas son la Firma de Fourier y Fourier 2D. Se utilizan, por tanto, descriptores basados en la apariencia global de la escena.

La combinación de los algoritmos de estimación de altura con los distintos descriptores utilizados proporcionan 10 métodos distintos. La Tabla 7.3 resume el conjunto de técnicas, proyecciones y descriptores utilizados, junto con las unidades cada estimador.

Los resultados se muestran en gráficas que recogen el valor medio y la varianza de todos los resultados obtenidos para las diferentes localizaciones en cada experimento.

Tabla 7.3: Resumen de Métodos de Estimación de altura junto con las distintas representaciones de la información omnidireccional, el descriptor empleado y el indicador de cambio de altura.

Método Estimación Altura	**Representación de la Escena**	**Descriptor**	**Indicador**
Correlación Celda Central Imagen Panorámica	Imagen Panorámica	Firma Fourier	Píxeles (d)
		FFT 2D	Píxeles (d)
Desfase Vertical FFT2D	Imagen Panorámica	FFT 2D Sig.	$R(^{o})$
Análisis Multiescala	Vista Ortográfica	Firma Fourier	Δfc
		FFT 2D	Δfc
Movimiento Sistema Referencia de Cámara	Imagen Panorámica	Firma Fourier	ρ
		FFT 2D	ρ
	Vista Ortográfica	Firma Fourier	ρ
		FFT 2D	ρ
	Proyección Esfera Unitaria	SFT	ρ

La Figura 7.11 recoge los resultados de estimación de altura de cada localización de exterior respecto a las imágenes con alturas $h = 1$ y $h = 5$. Como se puede apreciar en los resultados, todos los indicadores muestran una tendencia creciente conforme aumentamos la altura de la imagen de test. Además, cuando la altura de la imagen de test se encuentra por debajo de la de referencia (caso de $h_{ref} = 5$), los indicadores adquieren valores negativos.

Para una diferencia de hasta $h = 3$, es decir, hasta 45 cm, todos los descriptores muestran resultados con varianzas aceptables.

En general, la varianza de los resultados aumenta conforme nos distanciamos de la imagen de referencia. Esto es especialmente notable cuando se emplean las vistas panorámicas, tanto utilizando la correlación de la celda central (Figuras 7.11 (a y b)) como el movimiento del sistema de referencia de la cámara (Figuras 7.11 (f y g)). La alta varianza de los resultados para las alturas de test más alejadas de la imagen de referencia, muestran un indicador poco fiable para esos rangos de altura.

De la misma forma, la estimación de la altura usando la estimación de fase vertical con la Transformada de Fourier 2D (Figura 7.11(c)) presenta una varianza elevada para las alturas de test más alejadas de la imagen de referencia. Es importante remarcar, tal y como se ha indicado anteriormente, que este método está basado en el Teorema del Desplazamiento, que supone que la información de la imagen es siempre la misma, y únicamente varía el orden de las filas de la imagen. Sin embargo, al mover verticalmente la cámara, se introduce información nueva a la vez que desaparece otra existente. Esto introduce una diferencia en los coeficientes de la transformada de Fourier que conlleva un error intrínseco del método en la estimación del desfase vertical. A medida que se incremente la diferencia de altura entre imágenes, este error será mayor.

Si se comparan los descriptores utilizados en las escenas panorámicas y ortográficas, no se aprecian diferencias notables entre los resultados obtenidos con la Firma de Fourier y los obtenidos con la Transformada de Fourier 2D. Sin embargo, la varianza de los resultados es algo menor para el primer descriptor.

No es posible comparar la Transformada Esférica de Fourier porque es el único descriptor sobre la esfera unitaria. Sus resultados muestran una marcada tendencia lineal, aunque la desviación respecto de la media es elevada para desfases verticales mayores a 60 cm.

Si se comparan las distintas alturas de referencia, al usar $h_{ref} = 5$, los resultados presentan un mejor comportamiento en general que al realizar la comparación con $h_{ref} = 1$. Conviene recordar que la diferencia de altura máxima usando $h_{ref} = 5$ es menor que en el otro caso.

Por otro lado, los distintos indicadores muestran una precisión similar indistintamente de que la imagen de test se encuentre situada por debajo que por encima de la referencia, manteniendo la tendencia lineal con una desviación respecto de la media similar para incrementos equivalentes positivos y negativos.

La Figura 7.12 muestra los mismos resultados para la base de interior. Los resultados revelan un comportamiento semejante al obtenido con la base de exterior, aunque aumenta en la mayoría de los casos.

Comparando los resultados de interior con los de exterior, podemos apreciar que los indicadores en el caso de la base de interior adquieren valores mayores, sobre todo para las imágenes panorámicas. La razón principal es la distancia relativa de los elementos de la escena con respecto al sensor visual.

En las imágenes de interior, los elementos se encuentran, por lo general, más cercanos al sistema catadióptrico que en el caso de exterior. Esto provoca que al modificar la altura de la cámara, la distribución de los elementos dentro de la escena varíe más, debido a que el ángulo de incidencia de los rayos que recogen los objetos sufren una mayor variación. En la Figura 7.9 se muestra la variación del ángulo de incidencia al modificar la altura del

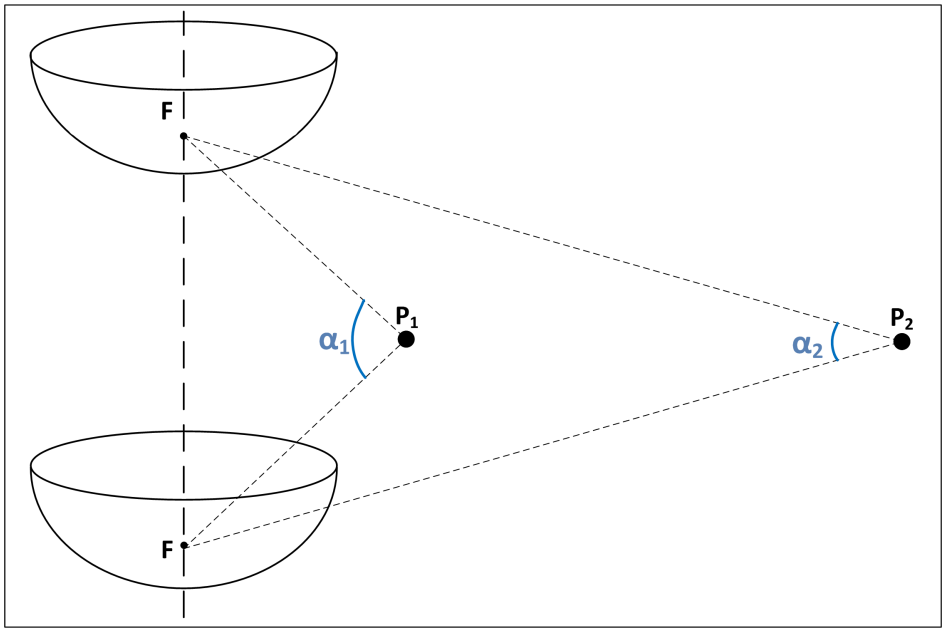


Figura 7.9: Variación del ángulo de incidencia de los rayos provenientes de dos puntos a distinta distancia cuando se produce una variación de la altura del sistema visual.

sistema visual de dos puntos situados a distintas distancias. Siguiendo el ejemplo de la figura, es posible ver que α_1, que representa el cambio en el ángulo de proyección del punto más cercano ($P1$) es mayor que para el caso del punto $P2$ (α_2).

Además, en las imágenes de interior existe por lo general una mayor diferencia de distancia entre elementos recogidos en la escena. Por ello, al modificar la altura de la cámara, los distintos elementos sufren una variación de posición en la escena distinta. La Figura 7.10 muestra la vista panorámica de dos escenas capturadas en un mismo punto a dos alturas diferentes. Se puede observar cómo el elemento remarcado en rojo sufre una mayor variación de altura ($h'_1 - h_1$) que el objeto marcado en verde ($h'_2 - h_2$), que se encuentra más alejado.

Por ello, la varianza de los indicadores de altura aumentan conforme crece el desfase vertical entre imágenes, especialmente en las escenas panorámicas. Los métodos basados en la vista de pájaro sufren menos esta variación al representar elementos que se encuentran a una distancia similar (mayoritariamente situados en el plano del suelo).

Siguiendo con los resultados de estimación de altura de las imágenes de interior (Figura 7.12), se aprecia una pérdida de la tendencia lineal en las alturas más altas. La causa vuelve a estar en la diferencia de movimiento entre los distintos elementos de la escena, y al igual que por la pérdida de información al acercarnos al techo de las estancias.

Esta pérdida es especialmente destacable en el método basado en el desfase vertical de la transformada de Fourier 2D (Figura 7.11 (c)).

Figura 7.10: Desplazamiento de elementos de la escena panorámica situados a distintas distancias respecto del sistema visual al variar la altura de captura.

Las Figuras 7.13 y 7.14 muestran los resultados de los gradientes positivos y negativos respectivamente. Estos resultados evidencian lo expuesto anteriormente: la magnitud de los distintos indicadores siguen una tendencia creciente a medida que aumenta el desplazamiento vertical entre las imágenes comparadas. Para los desplazamientos negativos, su signo también es negativo.

Los indicadores para las escenas de interior tienen valores claramente mayores comparado con los de las escenas de interior, a excepción de los métodos basados en la proyección ortográfica. Las técnicas de análisis multiescala sobre la proyección ortográfica y el movimiento del sistema de referencia de la cámara utilizando la proyección ortográfica son los que menor diferencia presentan entre exterior e interior. Además, la desviación de sus resultados respecto a la media también es reducida.

En general, a medida que aumenta el gradiente, la varianza de los resultados crece. Por último, también se observa que con las imágenes de exterior, las gráficas presentan una tendencia más lineal que en el caso de interior.

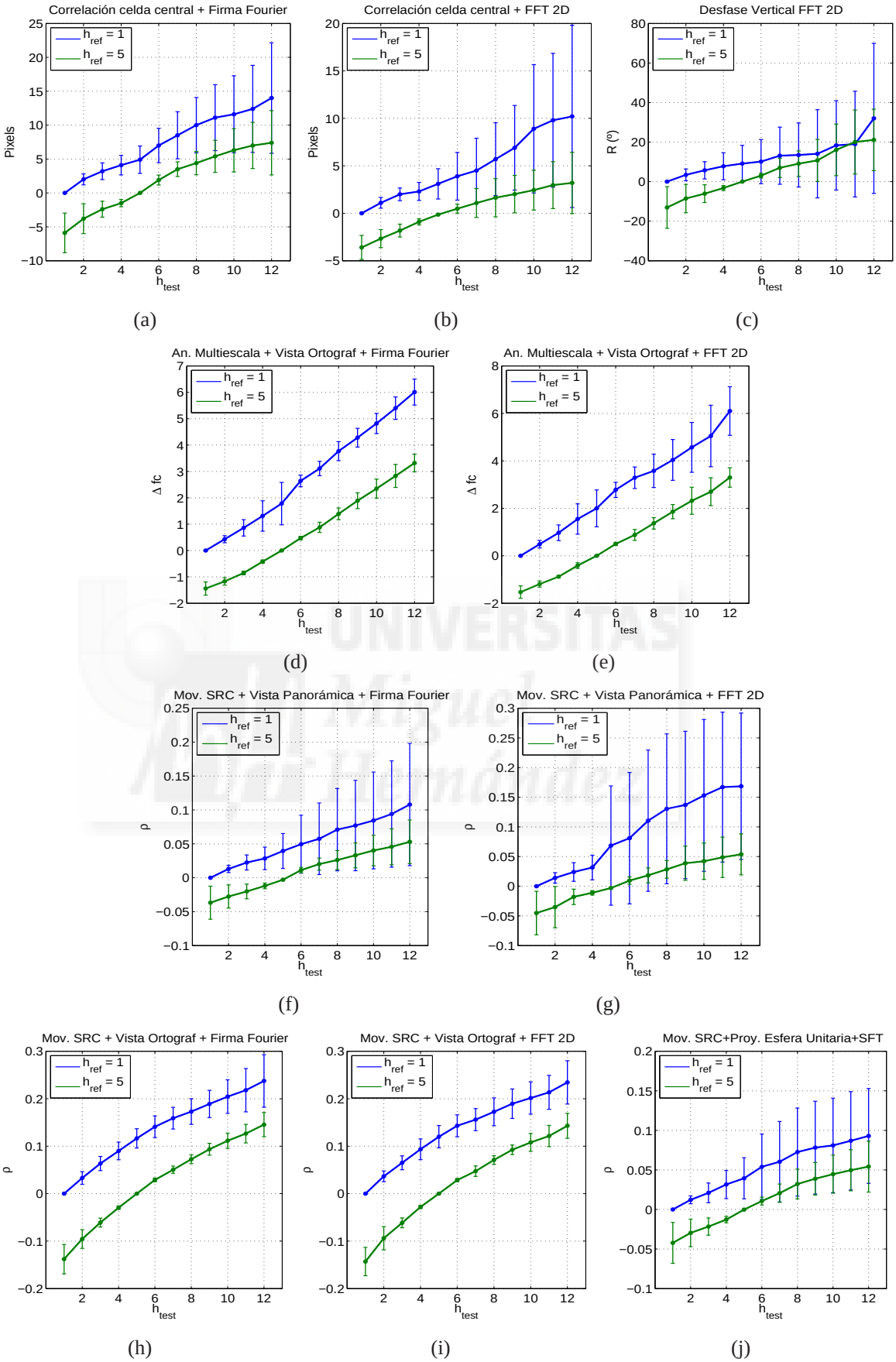


Figura 7.11: Estimación del desplazamiento vertical de las distintas escenas de exterior tomando como referencia la imagen a altura $h = 1$ y $h = 5$.

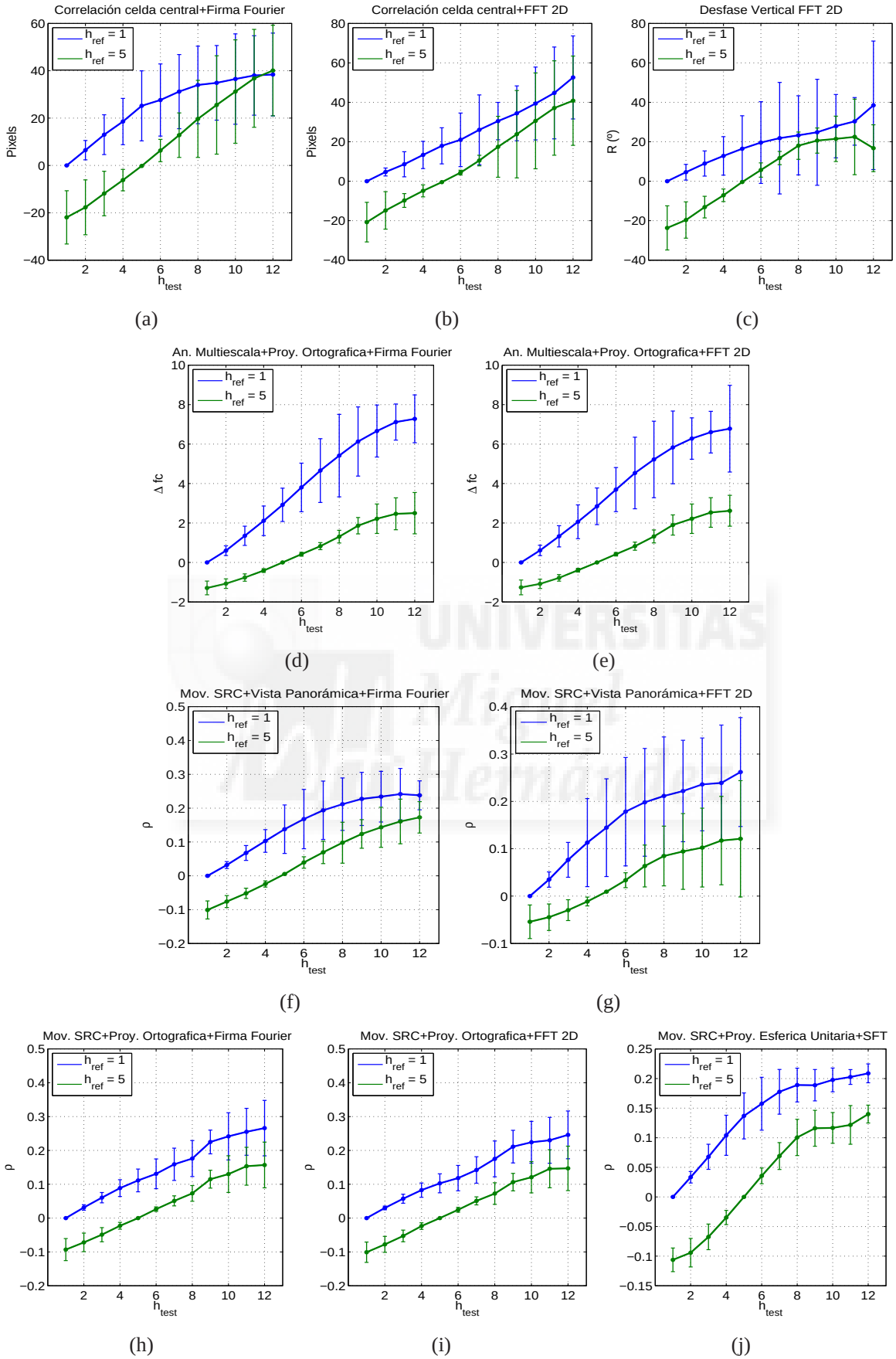


Figura 7.12: Estimación del desplazamiento vertical de las distintas escenas de interior tomando como referencia la imagen a altura $h = 1$ y $h = 5$.

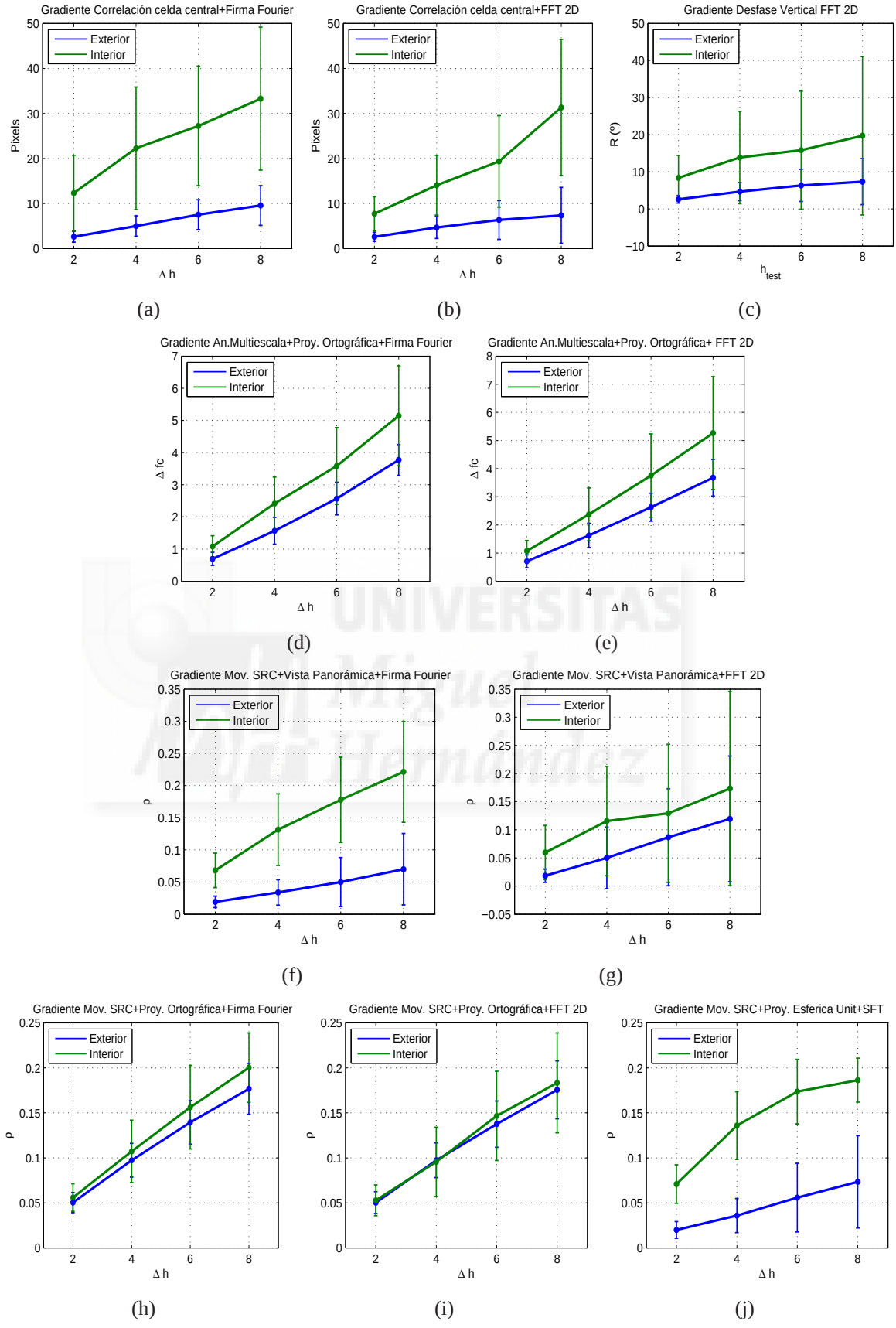


Figura 7.13: Estimación de distintos gradientes de desplazamiento vertical positivos para imágenes de exterior e interior.

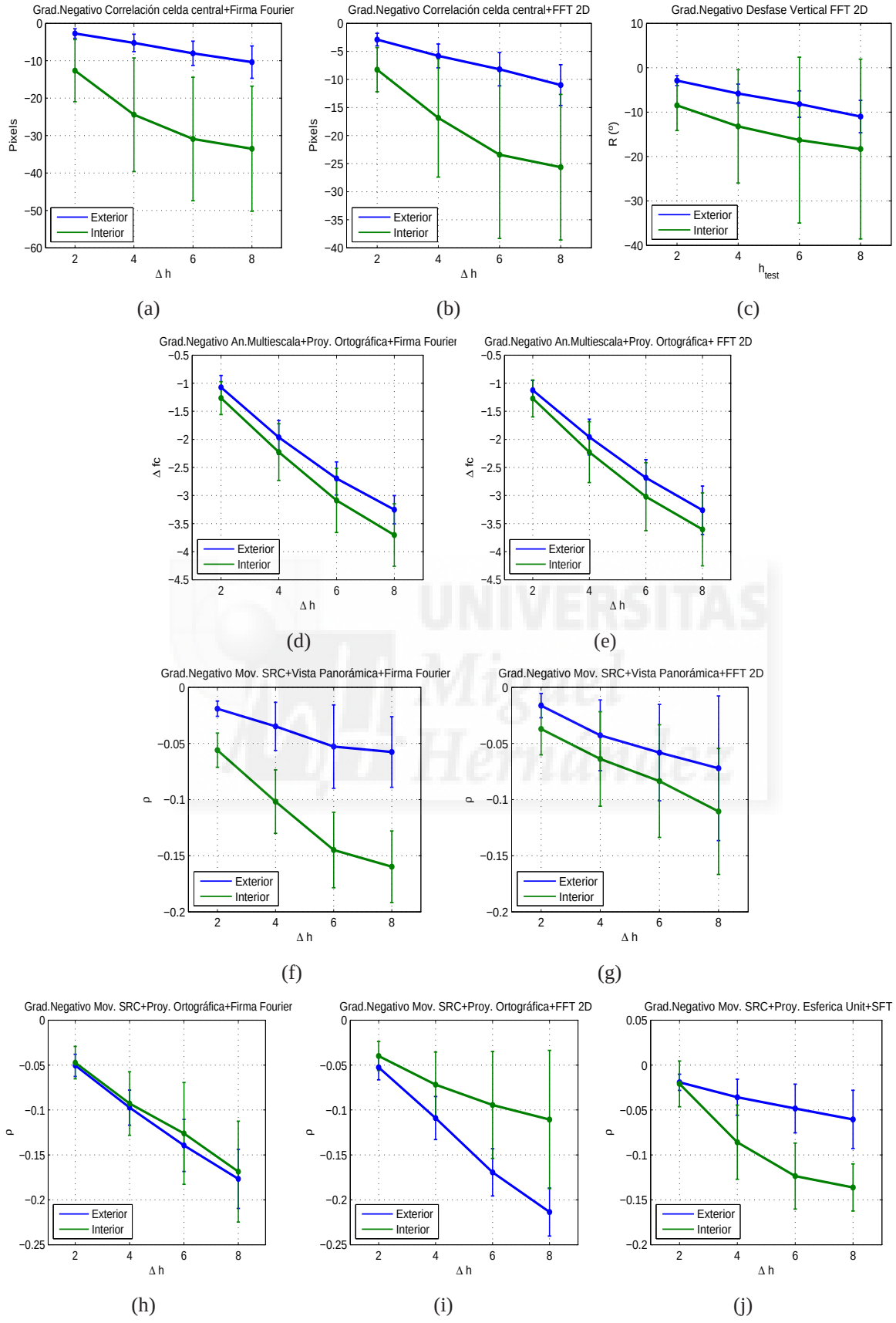


Figura 7.14: Estimación de distintos gradientes de desplazamiento vertical negativos para imágenes de exterior e interior.

7.4 Conclusiones

En este capítulo se ha presentado una comparación de distintos métodos destinados a proporcionar un indicador de diferencia de altura usando escenas omnidireccionales. Las aproximaciones incluidas usan la apariencia global de las imágenes para describir la información visual.

Los experimentos han sido llevados a cabo usando nuestra propia base de imágenes, que incluye escenas de exterior e interior, capturadas en ambientes reales bajo condiciones de iluminación cambiantes.

A continuación se enumeran las principales conclusiones obtenidas:

- Todos los métodos expuestos demuestran ser capaces de detectar una diferencia de altura entre imágenes capturadas en un mimo punto XY, lidiando satisfactoriamente con pequeños desplazamientos y cambios de orientación que se producen durante la captura de las escenas.
- Los indicadores presentan una tendencia lineal respecto a la diferencia de altitud de las imágenes comparadas, siendo más fuerte esta tendencia en las escenas de interior.
- El signo de los indicadores proporciona información sobre la dirección del desplazamiento vertical. Así pues, un signo negativo indica que la imagen comparada está por debajo de la de referencia.
- Conforme aumenta la distancia entre imágenes, también lo hace la varianza de los resultados, especialmente en las escenas de exterior.
- Para diferencias de altura menores a 45 cm, todos los algoritmos presentan resultados que permiten estimar la diferencia de altura con una precisión aceptable.
- Las técnicas basadas en la proyección ortográfica de la escena omnidireccional muestran una mayor linealidad y menor varianza en los resultados, sobre todo cuando se usa el método basado en el movimiento vertical del sistema de referencia de la cámara (SRC) para estimar la altura.
- No existen diferencias significas en la precisión usando la Firma de Fourier o la Transformada de Fourier Bidimensional, aunque el primer descriptor muestra un mejor comportamiento.
- Las distintas técnicas dependen del movimiento de los objetos dentro de la escena. Debido a que dicho movimiento es mayor en las escenas de interior, los indicadores para

esas escenas presentan un valor mayor. Como la proyección ortográfica recoge mayoritariamente información del plano del suelo, se ven menos afectadas por este hecho. Por lo tanto, la magnitud sus indicadores muestran menor dependencia al entorno en el que se ha realizado la captura de las escenas.

- El desplazamiento de los elementos proyectados en las escenas de interior es menos homogéneo que en el caso de las de exterior debido a la mayor diferencia de distancias de los objetos con respecto al sensor visual. Esto se traduce en un aumento de la varianza de los resultados.

- La introducción de nueva información a la escena provoca que la utilización del Desfase vertical sobre la FFT 2D sea poco fiable para incrementos de altura mayores a 45 cm, además de ser especialmente sensible a cambios de orientación de la imagen en el plano del suelo.

- Se produce una pérdida de crecimiento en la tendencia de algunos indicadores para los rangos de altura más altos en imágenes de interior.

- La Transformada Esférica de Fourier es el único descriptor de los incluidos que permitiría estimar cambios en la inclinación del sensor visual.

CAPÍTULO

8

Conclusiones

En este capítulo se resumen las principales conclusiones del trabajo realizado y las posibles líneas futuras de investigación que pueden derivarse de él.

8.1 Aportaciones

Al final de cada capítulo en el que se presentan resultados experimentales, se detallan sus conclusiones de forma más extensa. A continuación, se incluyen las principales aportaciones:

- Se lleva a cabo un estudio acerca de la obtención de distintas vistas a partir de información visual omnidireccionales. En el Capítulo 3 se recogen las distintas proyecciones analizadas en este trabajo, dando una perspectiva del potencial de los sistemas visuales catadióptricos.

- El Capítulo 4 es una recopilación de las principales técnicas basadas en apariencia global que existen actualmente. El estudio se centra en la obtención de descriptores que permiten caracterizar la escena, y que pueden ser utilizados en tareas de asociación de imágenes y estimación de desfase entre escenas. Los distintos métodos utilizan principalmente la proyección cilíndrica de la información omnidireccional, o imagen panorámica, apareciendo también un descriptor basado en la proyección sobre la esfera unitaria. Además, se propone un nuevo descriptor para imágenes panorámicas.

- Se realiza un estudio comparativo de los distintos descriptores de apariencia global añadiendo la información de color de las escenas. El análisis mide la precisión y necesidades computacionales de cada técnica en la tarea de construcción de un mapa

denso y estimación de la pose dentro del mapa creado. La comparación contempla cinco configuraciones distintas de la utilización de los espacios de color: la imagen en escala de grises, aplicando los descriptores sobre los distintos canales RGB, HSV, el uso conjunto de RGB+HSV y, por último, añadiendo la información relativa a color mediante histograma de intensidad de los canales de color en espacio HSV al descriptor calculado sobre escala de grises normalizado. Como muestran los resultados del Capítulo 5, la información de color mejora la precisión de localización de todos los descriptores. De forma general, al utilizar el espacio HSV, la localización presenta una precisión más alta que si se aplica sobre RGB. El cálculo del Histograma de Color es un proceso computacionalmente ligero, y mejora los resultados de localización frente al descriptor sobre escala de grises.

- El descriptor propuesto (Fourier 1D) es, con diferencia, el más compacto, presentando también un menor requerimiento de tiempo para estimación de la pose. Sobre la imagen en escalas de grises, muestra resultados de localización no satisfactorios. Sin embargo, al incluir la información de color, su precisión es similar al resto de técnicas.

- Respecto al comportamiento ante ruido y oclusiones, los descriptores que utilizan la información de color presentan una mayor reducción de la precisión, especialmente los métodos que utilizan HSV cuando la imagen está afectada por ruido Gaussiano. Independientemente, la información de color sigue mejorando la descripción de la escena en la mayoría de técnicas.

- En el Capítulo 6 se muestra la aplicación de la apariencia global para la estimación de desplazamiento topológico visual, con el llamado Análisis Multiescala. En la primera parte, se emplean imágenes capturadas con un sistema visual de amplio campo de visión para construir un mapa topológico basado en un grafo, cuyos nodos representan distintas localizaciones del área de navegación. Los resultados muestran una construcción satisfactoria del mapa, que presenta una distribución espacial similar a la real. Además, se plantea un sistema de estimación de pose del robot dentro del mapa que permite hallar su localización no sólo en las posiciones de los nodos, sino también en puntos intermedios del grafo.

- Otra contribución que se incluye en esta investigación es la adaptación del Análisis Multiescala a información omnidireccional. Aprovechando las propiedades de las escenas omnidireccionales, se mejora la propuesta inicial mediante una doble comparación de escalas, que incluye proyecciones de la imagen en la dirección de avance y la opuesta. Esta información es combinada con la proporcionada por los descriptores

estudiados en el Capítulo 5 para obtener un sistema de odometría visual topológico. Por último, se presenta una aplicación para calcular el camino de rutas visuales, incluyendo el cierre de bucles a través de la asociación de escenas usando la apariencia visual. El cierre de bucle es utilizado para llevar a cabo una reestimación, mejorando el mapa inicial obtenido con el sistema de odometría visual.

- Finalmente, el Capítulo 7 expone distintas técnicas orientadas a la estimación de altura de escenas omnidireccionales mediante la apariencia global. En concreto, se proponen cuatro métodos, que se basan respectivamente en el uso del espacio frecuencial, la correlación de información visual sobre la imagen panorámica, el desplazamiento del sistema de referencia de la cámara a través de la geometría epipolar, y el Análisis Multiescala aplicado sobre la perspectiva ortográfica. Los resultados experimentales muestran que los algoritmos presentados son capaces de proporcionar estimadores topológicos para diferencias de altura de hasta 45 cm de forma bastante precisa. Para desplazamientos mayores, sólo algunas técnicas tienen una varianza suficientemente pequeña para considerar sus estimaciones como fiables. Se pueden destacar especialmente métodos basados en la proyección ortográfica de la imagen. Por último, las escenas de exterior ofrecen mejores resultados, con tendencias más lineales respecto de la variación de altura, con una varianza de las estimaciones generalmente menor a las de interior.

8.2 Líneas Futuras de Investigación

Partiendo de las aportaciones presentadas en esta investigación, es posible plantear objetivos a seguir en futuras investigaciones. Los siguiente puntos incluyen algunas propuestas:

- *Estudio continuo de descriptores basados en apariencia*.

  Los métodos basados en apariencia son una propuesta que podemos considerar reciente. Aunque su aplicabilidad en tareas de navegación visual ha quedado demostrada, todavía queda recorrido para su optimización, especialmente en lo referente a coste computacional. Muestran una dependencia importante con la resolución de la escena sobre la que se aplican. Una posible línea es el estudio de la resolución mínima necesaria de la información omnidireccional sin perder su eficacia. Además, al representar un conjunto de métodos muy amplio, se puede pensar en nuevas formas de describir la información en su conjunto usando otras técnicas.

- *Mejora del sistema de estimación de rutas mediante odometría visual topológica*.

  Los resultados obtenidos a partir del método de odometría visual propuesto nos animan a seguir mejorando la aplicación. Existen distintas líneas a partir de las cuales hacer un sistema más robusto. Por ejemplo, en la estimación de desplazamientos laterales entre imágenes que presentan la misma orientación pero que, sin embargo, sus coordenadas no coinciden exactamente. Se puede plantear un sistema de correlación de subventanas para obtener una estimación más precisa de dicho desplazamiento lateral. Por otro lado, también es posible incluir en el cierre de bucle una reestimación iterativa de las posiciones, utilizando de nuevo la odometría topológica tras corregir los desfases entre escenas, para volver a calcular los los desplazamientos de forma topológica con el Análisis Multiescala. Dicho de otra forma, se puede pensar en recalcular el tramo de ruta comprendido en el cierre de bucle incluyendo la información del error de fase obtenido, estimando de nuevo los desplazamientos con los nuevos ángulos. Finalmente, podría plantearse incluir la información de la odometría interna del robot mediante la utilización de algún filtro probabilístico. De esta manera, se reduciría el número de comparaciones utilizadas en el Análisis Multiescala, además de obtener una primera estimación del desfase entre escenas consecutivas.

- *Sistema de control visual*.

  La información proporcionada por el sistema de odometría puede ser utilizado para crear un control del robot basado en la información visual topológica. Este control puede ser utilizado, por ejemplo, para seguimiento de una ruta anteriormente recorrida. Una vez obtenido el mapa, también puede plantearse la planificación de la trayectoria para llegar a un punto deseado.

- *Estimación de movimientos 6D utilizando la apariencia global*.

  En este trabajo se ha supuesto que la inclinación del sistema visual al estimar la altura del robot se mantiene constante. A partir de descriptores como la Transformada Esférica de Fourier (SFT), se obtienen descriptores invariantes a cambios en la orientación respecto a cualquier eje tridimensional, además de información para estimar dichos cambios de orientación. Combinando los sistemas de estimación de pose en navegación sobre un mismo plano, la estimación de altura con métodos como los presentados en el Capítulo 7 y la estimación de la inclinación del robot, parece viable diseñar un algoritmo que permita obtener información sobre movimientos 6D en el espacio usando la apariencia global visual.

UNIVERSITAS
Miguel
Hernández

UNIVERSITAS
Miguel
Hernández

Bibliografía

[1] Point Grey Research, Inc. Ladybug Cameras. [Online]. Available: http://ww2.ptgrey.com/spherical-vision 20, 32

[2] Woodman Labs Inc. GoPro Hero Camera. [Online]. Available: http://gopro.com/hd-hero2-cameras/ 53

[3] Accowle Company, LTD. Accowle Omnidirectional Vision Sensor. [Online]. Available: http://www.accowle.com/english/products.html 38

[4] M. Achtelik, J. Stumpf, D. Gurdan, and K.-M. Doth, "Design of a flexible high performance quadcopter platform breaking the mav endurance record with laser power beaming," in *Intelligent Robots and Systems (IROS), 2011 IEEE/RSJ International Conference on*, 2011, pp. 5166–5172. 13

[5] Adept MobileRobots LLC. Robot pioneer p3-at. [Online]. Available: http://www.mobilerobots.com/researchrobots/p3at.aspx 56

[6] K. Akimoto, Y. Henmi, H. Shimasaki, K. Yakita, T. Hokamura, T. Masuda, T. Takeuchi, K. Shibata, M. Shibata, K. Nakano, M. Sannoh, and K. Yoshizu, "Autonomous underwater vehicle (auv) investigations to protect valuable invertebrates," in *Underwater Technology (UT), 2011 IEEE Symposium on and 2011 Workshop on Scientific Use of Submarine Cables and Related Technologies (SSC)*, 2011, pp. 1–6. 11

[7] F. Amorós, L. Payá, O. Reinoso, L. Fernández, and D. Valiente, "Towards relative altitude estimation in topological navigation tasks using the global appearance of visual information," in *VISAPP 2014, International Conference on Computer Vision Theory and Applications*. Ed. SciTePress - Science and Technology Publications ISBN: 978-989-758-003-1 - Volume 1, pp. 194-201, 2014.

[8] F. Amorós, L. Payá, L. Fernández, O. Reinoso, M. Ballesta, and M. Juliá, "Construcción de mapas topológicos y estimación de trayectorias usando descriptores de apariencia visual global," in *XXXIV Jornadas de Automática*, Terrassa, 2013, pp. 834–841.

[9] F. Amorós, L. Payá, O. Reinoso, and L. Fernández, "Map building and localization using global-appearance descriptors applied to panoramic images," *Journal of Computer and Information Technology*, vol. 2, no. 1, pp. 55–71, 2012.

[10] F. Amorós, L. Payá, O. Reinoso, L. Fernández, and J. M. Marín, "Visual map building and localization with an appearance-based approach - comparisons of techniques to extract information of panoramic images," in *7th Internacional Conference on Informatics, in Control, Automation and Robotics (ICINCO 2010)*. Funchal (Madeira), Portugal: SciTePress - Science and Technology Publications, 2010, pp. 423–426.

[11] F. Amorós, L. Payá, O. Reinoso, and L. Jiménez, "Comparison of global-appearance techniques applied to visual map building and localization," in *International Conference on Computer Vision Theory and Applications (VISAPP 2012)*, vol. 2. Rome, Italy: SciTePress - Science and Technology Publications, 2012, pp. 395–398.

[12] ——, "Uso de descriptores de apariencia global en tareas de construcción de mapas y localización," in *XXXIII Jornadas de Automática*. Vigo: Ed. CEA-IFAC, 2012, pp. 993–1002.

[13] F. Amorós, L. Payá, O. Reinoso, L. Jiménez, and M. Juliá, "Topological height estimation using global appearance of images," in *ROBOT2013: First Iberian Robotics Conference*, ser. Advances in Intelligent Systems and Computing. Madrid: Springer International Publishing, 2014, vol. 253, pp. 77–89.

[14] F. Amorós, L. Payá, O. Reinoso, W. Mayol-Cuevas, and A. Calway, "Topologial map building and path estimation using global-apperance image descriptors," in *10th International Conference on Informatics in Control, Automation and Robotics (ICINCO 2013)*. Reykjavik, Iceland: SciTePress - Science and Technology Publications, 2013, pp. 385–392.

[15] F. Amorós, O. Reinoso, L. Payá, L. Fernández, and J. M. Marín, "Construcción de mapas visuales y localización mediante métodos basados en apariencia global," in *XXXI Jornadas de Automática*. Jaén: Ed. CEA-IFAC, 2010, pp. 31–39.

[16] M. Artac, M. Jogan, and A. Leonardis, "Mobile robot localization using an incremental eigenspace model," in *Robotics and Automation, 2002. Proceedings. ICRA '02. IEEE International Conference on*, vol. 1, 2002, pp. 1025–1030 vol.1. 82

[17] S. Baker and S. Nayar, "A theory of catadioptric image formation," in *Computer Vision, 1998. Sixth International Conference on*, 1998, pp. 35–42. 33

[18] D. H. Ballard, "Readings in computer vision: Issues, problems, principles, and paradigms," M. A. Fischler and O. Firschein, Eds. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 1987, ch. Generalizing the Hough Transform to Detect Arbitrary Shapes, pp. 714–725. [Online]. Available: http://dl.acm.org/citation.cfm?id=33517.33574 24

[19] P. Batista, C. Silvestre, and P. Oliveira, "Ges integrated lbl/usbl navigation system for underwater vehicles," in *Decision and Control (CDC), 2012 IEEE 51st Annual Conference on*, 2012, pp. 6609–6614. 12

[20] H. Bay, T. Tuytelaars, and L. Gool, "Surf: Speeded up robust features," in *Computer Vision at ECCV 2006*, ser. Lecture Notes in Computer Science, A. Leonardis, H. Bischof, and A. Pinz, Eds. Springer Berlin Heidelberg, 2006, vol. 3951, pp. 404–417. [Online]. Available: http://dx.doi.org/10.1007/11744023_32 23, 61

[21] S. Bogner, "An introduction to panospheric imaging," in *Systems, Man and Cybernetics, 1995. Intelligent Systems for the 21st Century., IEEE International Conference on*, vol. 4, 1995, pp. 3099–3106 vol.4. 33

[22] B. Bonev, M. Cazorla, and F. Escolano, "Robot navigation behaviors based on omnidirectional vision and information theory," *Journal of Physical Agents*, vol. 1, no. 1, 2007. [Online]. Available: http://www.jopha.net/index.php/jopha/article/view/10/9 50

[23] F. Bonin-Font, A. Ortiz, and G. Oliver, "Visual navigation for mobile robots: A survey," *Journal of Intelligent and Robotic Systems*, vol. 53, no. 3, pp. 263–296, 2008. [Online]. Available: http://dx.doi.org/10.1007/s10846-008-9235-4 31

[24] D. Bradley, A. Brunton, M. Fiala, and G. Roth, "Image-based navigation in real environments using panoramas," in *Haptic Audio Visual Environments and their Applications, 2005. IEEE International Workshop on*, 2005, pp. 3 pp.–. 18

[25] A. J. Briggs, C. Detweiler, Y. Li, P. C. Mullen, and D. Scharstein, "Matching scale-space features in 1d panoramas," *Computer Vision and Image Understanding*, vol. 103, no. 3, pp. 184 – 195, 2006, special issue on Omnidirectional Vision and Camera Networks. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1077314206000683 65

[26] A. J. Briggs, C. Detweiler, P. C. Mullen, and D. Scharstein, "Scale-space features in 1d omnidirectional images," in *in Omnivis 2004, the Fifth Workshop on Omnidirectional Vision*, 2004, pp. 115–126. 65

[27] A. J. Briggs, Y. Li, and D. Scharstein, "Feature matching across 1d panoramas," in *In Omnivis 2005, the sixth Workshop on Omnidirectional Vision*, 2005. 65

[28] A. J. Briggs, D. Scharstein, and S. D. Abbott, "Reliable mobile robot navigation from unreliable visual cues," in *ALGORITHMIC AND COMPUTATIONAL ROBOTICS: NEW DIRECTIONS*, 2000, pp. 349–362. 62

[29] M. Brown, D. Burschka, and G. Hager, "Advances in computational stereo," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 25, no. 8, pp. 993–1008, 2003. 18

[30] E. Cabral, J. de Souza, and M. Hunold, "Omnidirectional stereo vision with a hyperbolic double lobed mirror," in *Pattern Recognition, 2004. ICPR 2004. Proceedings of the 17th International Conference on*, vol. 1, 2004, pp. 1–9 Vol.1. 33

[31] J. Canny, "A computational approach to edge detection," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. PAMI-8, no. 6, pp. 679–698, 1986. 24, 87

[32] Y. Cha and D. Kim, "Omni-directional image matching for homing navigation based on optical flow algorithm," in *Control, Automation and Systems (ICCAS), 2012 12th International Conference on*, 2012, pp. 1446–1451. 30

[33] C.-K. Chang, C. Siagian, and L. Itti, "Mobile robot vision navigation and localization using gist and saliency," in *Intelligent Robots and Systems (IROS), 2010 IEEE/RSJ International Conference on*, 2010, pp. 4147–4154. 105

[34] W.-C. Chang and C.-Y. Chuang, "Vision-based robot navigation and map building using active laser projection," in *System Integration (SII), 2011 IEEE/SICE International Symposium on*, 2011, pp. 24–29. 17

[35] C.-H. Chen and K.-T. Song, "Complete coverage motion control of a cleaning robot using infrared sensors," in *Mechatronics, 2005. ICM '05. IEEE International Conference on*, 2005, pp. 543–548. 17

[36] D. Choi, J. Kim, S. Cho, S. Jung, and J. Kim, "Rocker-pillar : Design of the rough terrain mobile robot platform with caterpillar tracks and rocker bogie mechanism," in *Intelligent Robots and Systems (IROS), 2012 IEEE/RSJ International Conference on*, 2012, pp. 3405–3410. 15

[37] J. Choi, S. Ahn, M. Choi, and W. K. Chung, "Metric slam in home environment with visual objects and sonar features," in *Intelligent Robots and Systems, 2006 IEEE/RSJ International Conference on*, 2006, pp. 4048–4053. 17

[38] J. Choi, S. Ahn, and W. K. Chung, "Robust sonar feature detection for the slam of mobile robot," in *Intelligent Robots and Systems, 2005. (IROS 2005). 2005 IEEE/RSJ International Conference on*, 2005, pp. 3415–3420. 16

[39] D. Cobzas and H. Zhang, "Cylindrical panoramic image-based model for robot localization," in *Intelligent Robots and Systems, 2001. Proceedings. 2001 IEEE/RSJ International Conference on*, vol. 4, 2001, pp. 1924–1930 vol.4. 45, 50

[40] G. Conte and P. Doherty, "An integrated uav navigation system based on aerial image matching," in *Aerospace Conference, 2008 IEEE*, 2008, pp. 1–10. 13, 14

[41] J. Courbon, Y. Mezouar, L. Eck, and P. Martinet, "A generic fisheye camera model for robotic applications," in *Intelligent Robots and Systems, 2007. IROS 2007. IEEE/RSJ International Conference on*, 2007, pp. 1683–1688. 53

[42] M. Cummins and P. Newman, "Fab-map: Probabilistic localization and mapping in the space of appearance," in *International Journal of Robotics Research 27 (6), pp. 647-665*, 2008. 62

[43] ——, "Appearance-only SLAM at large scale with FAB-MAP 2.0," *The International Journal of Robotics Research*, 2010. [Online]. Available: http://ijr.sagepub.com/content/early/2010/11/11/0278364910385483 62

[44] S. da Costa Botelho, P. Drews, G. Oliveira, and M. da Silva Figueiredo, "Visual odometry and mapping for underwater autonomous vehicles," in *Robotics Symposium (LARS), 2009 6th Latin American*, 2009, pp. 1–6. 13

[45] N. Dalal and B. Triggs, "Histograms of oriented gradients fot human detection." in *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, San Diego, USA. Vol. II, pp. 886-893*, 2005. 87

[46] A. J. Davison and N. Kita, "Simultaneous localisation and map-building using active vision," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 865–880, 2002. 24

[47] S. M. M. Dehghan, M. Zarezadeh, N. Farhadian, and H. Moradi, "The design, modeling and control of a tethered aerial robot for serach and rescue missions," in *Robotics and Biomimetics (ROBIO), 2011 IEEE International Conference on*, 2011, pp. 1302–1307. 13

[48] G. DeSouza and A. Kak, "Vision for mobile robot navigation: a survey," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 24, no. 2, pp. 237–267, 2002. 31

[49] J. Driscoll and D. Healy, "Computing fourier transforms and convolutions on the 2-sphere," *Advances in Applied Mathematics*, vol. 15, no. 2, pp. 202 – 250, 1994. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0196885884710086 71, 75

[50] I. Dryden and K. Mardia, *Statistical shape analysis*, ser. Wiley series in probability and statistics. Wiley, 1998. 177, 194

[51] R. O. Duda and P. E. Hart, "Use of the hough transformation to detect lines and curves in pictures," *Commun. ACM*, vol. 15, no. 1, pp. 11–15, Jan. 1972. [Online]. Available: http://doi.acm.org/10.1145/361237.361242 24

[52] Eizoh Company, LTD. Eizoh Omnidirectional Vision Sensor. [Online]. Available: http://www.eizoh.co.jp/mirror/wide70.html 38, 119

[53] M. J. Er, S. Yuan, and N. Wang, "Development control and navigation of octocopter," in *Control and Automation (ICCA), 2013 10th IEEE International Conference on*, 2013, pp. 1639–1643. 13

[54] L. Fernández, L. Payá, M. Ballesta, F. Amorós, and O. Reinoso, "Odometría visual y construcción de un mapa topológico a partir de la apariencia global de imágenes omnidireccionales," in *XXXII Jornadas de Automática*, Sevilla, 2011.

[55] ——, "Localización monte carlo a partir de la apariencia global de imágenes omnidireccionales," in *XXXIII Jornadas de Automática*. Vigo: Ed. CEA-IFAC, 2012, pp. 743–750.

[56] L. Fernández, L. Payá, M. Juliá, F. Amorós, and O. Reinoso, "Visual odometry with an appearance-based method," in *ROBOT 2011. Robótica Experimental*. Sevilla: Ed. Universidad de Sevilla, 2011.

[57] L. Fernández, L. Payá, O. Reinoso, and F. Amorós, "Appearance-based visual odometry with omnidirectional images. a practical application to topological mapping," in *8th Internacional Conference on Informatics, in Control, Automation and Robotics (ICINCO 2011)*. Noordwijkerhout, The Netherlands: SciTePress - Science and Technology Publications, 2011.

[58] L. Fernández, L. Payá, O. Reinoso, A. Gil, and D. Valiente, "Visual hybrid slam: An appearance-based approach to loop closure," in *ROBOT2013: First Iberian Robotics Conference*, ser. Advances in Intelligent Systems and Computing, M. A. Armada, A. Sanfeliu, and M. Ferre, Eds. Springer International Publishing, 2014, vol. 252, pp. 693–701. [Online]. Available: http://dx.doi.org/10.1007/978-3-319-03413-3_51 30

[59] M. A. Fischler and R. C. Bolles, "Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography," *Commun. ACM*, vol. 24, no. 6, pp. 381–395, Jun. 1981. [Online]. Available: http://doi.acm.org/10.1145/358669.358692 24

[60] W. T. Freeman, M. Roth, and M. Roth, "Orientation histograms for hand gesture recognition," in *In International Workshop on Automatic Face and Gesture Recognition*, 1994, pp. 296–301. 87

[61] A. Friedman, "Framing pictures: The role of knowledge in automatized encoding and memory for gist." in *Journal of Experimental Psychology: General, 108:316-355.*, 1979. 95

[62] H. Friedrich, D. Dederscheck, K. Krajsek, and R. Mester, "View-based robot localization using spherical harmonics: Concept and first experimental results," in *Pattern Recognition*, ser. Lecture Notes in Computer Science, F. Hamprecht, C. Schnarr, and B. Jahne, Eds. Springer Berlin Heidelberg, 2007, vol. 4713, pp. 21–31. [Online]. Available: http://dx.doi.org/10.1007/978-3-540-74936-3_3 76

[63] H. Friedrich, D. Dederscheck, M. Mutz, and R. Mester, "View-based robot localization using illumination-invarant spherical harmonics descriptors," in *In VISAPP 2008*, 2008. 76

[64] D. Fries, G. Barton, G. Hendrick, B. Gregson, L. Hotaling, J. Paul, A. Sanderson, and R. Blidberg, "Solar robotic material sampler system for chemical, biological and physical ocean observations," in *OCEANS 2011*, 2011, pp. 1–5. 11

[65] J. Gaspar, N. Winters, and J. Santos-Victor, "Vision-based navigation and environmental representations with an omnidirectional camera," *Robotics and Automation, IEEE Transactions on*, vol. 16, no. 6, pp. 890 –898, dec 2000. 50

[66] L. Gerstmayr-Hillen, O. Schluter, M. Krzykawski, and R. Moller, "Parsimonious loop-closure detection based on global image-descriptors of panoramic images," in *Advanced Robotics (ICAR), 2011 15th International Conference on*, 2011, pp. 576–581. 45

[67] C. Geyer and K. Daniilidis, "A unifying theory for central panoramic systems and practical implications," in *Computer Vision â ECCV 2000*, ser. Lecture Notes in Computer Science, D. Vernon, Ed. Springer Berlin Heidelberg, 2000, vol. 1843, pp. 445–461. [Online]. Available: http://dx.doi.org/10.1007/3-540-45053-X_29 70

[68] A. Gil, O. Reinoso, O. Mozos, C. Stachniss, and W. Burgard, "Improving data association in vision-based slam," in *Intelligent Robots and Systems, 2006 IEEE/RSJ International Conference on*, 2006, pp. 2076–2081. 61, 87

[69] A. Gil, O. Reinoso, M. Ballesta, M. Julia, and L. Paya, "Estimation of visual maps with a robot network equipped with vision sensors," *Sensors*, vol. 10, no. 5, pp. 5209–5232, 2010. 132

[70] A. Gil, O. M. Mozos, M. Ballesta, and O. Reinoso, "A comparative evaluation of interest point detectors and local descriptors for visual slam," *Mach. Vision Appl.*, vol. 21, no. 6, pp. 905–920, Oct. 2010. [Online]. Available: http://dx.doi.org/10.1007/s00138-009-0195-x 24, 62

[71] A. Gil, O. Reinoso, M. Ballesta, and M. Juliá, "Multi-robot visual SLAM using a rao-blackwellized particle filter," *Robotics and Autonomous Systems*, vol. 58, no. 1, pp. 68 – 80, 2010. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0921889009001158 28

[72] G. Giralt, R. Sobek, and R. Chatila, "A multi-level planning and navigation system for a mobile robot: a first approach to hilare," in *Proceedings of the 6th international joint conference on Artificial intelligence - Volume 1*, ser. IJCAI'79. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 1979, pp. 335–337. 31

[73] S. Goto, A. Yamashita, R. Kawanishi, T. Kaneko, and H. Asama, "3d environment measurement using binocular stereo and motion stereo by mobile robot with omni-directional stereo camera," in *Computer Vision Workshops (ICCV Workshops), 2011 IEEE International Conference on*, 2011, pp. 296–303. 21

[74] V. Grassi Junior and J. Okamoto Junior, "Development of an omnidirectional vision system," *Journal of the Brazilian Society of Mechanical Sciences and Engineering*, vol. 28, pp. 58 – 68, 03 2006. [Online]. Available: http://www.scielo.br/scielo.php?script=sci_arttext&pid=S1678-58782006000100007&nrm=iso 33, 43

[75] A. Gurtner, D. Greer, R. Glassock, L. Mejias, R. Walker, and W. Boles, "Investigation of fish-eye lenses for small-uav aerial photography," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 47, no. 3, pp. 709–721, 2009. 20, 53

[76] K. Hara, S. Maeyama, and A. Gofuku, "Navigation path scanning system for mobile robot by laser beam," in *SICE Annual Conference, 2008*, 2008, pp. 2817–2821. 17

[77] C. Harris and M. Stephens, "A combined corner and edge detector," in *In Proc. of Fourth Alvey Vision Conference*, 1988, pp. 147–151. 22

[78] A. Hata and D. Wolf, "Outdoor mapping using mobile robots and laser range finders," in *Electronics, Robotics and Automotive Mechanics Conference, 2009. CERMA '09.*, 2009, pp. 209–214. 17

[79] D. Healy, Jr., D. Rockmore, P. J. Kostelec, and S. S. B. Moore, "Ffts for the 2-sphere - improvements and variations," *The Journal of Fourier Analysis and Applications*, vol. 9, pp. 341–385, 1996. 75

[80] E. Hering, *Outlines of a Theory of the Light Sense*. Harvard University Press, Jan. 1964. 105

[81] G. Hoffmann, C. Tomlin, D. Montemerlo, and S. Thrun, "Autonomous automobile trajectory tracking for off-road driving: Controller design, experimental validation and racing," in *American Control Conference, 2007. ACC '07*, 2007, pp. 2296–2301. 15

[82] H. Hotelling, “Analysis of a complex of statistical variables into principal components,” *J. Educ. Psych.*, vol. 24, 1933. 77

[83] C.-H. Hsieh, M.-L. Wang, L.-W. Kao, and H.-Y. Lin, “Mobile robot localization and path planning using an omnidirectional camera and infrared sensors,” in *Systems, Man and Cybernetics, 2009. SMC 2009. IEEE International Conference on*, 2009, pp. 1947–1952. 27

[84] Y. Huang, K. Huang, L. Wang, D. Tao, T. Tan, and X. Li, “Enhanced biologically inspired model,” in *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*, 2008, pp. 1–8. 105

[85] K. M. Huffenberger and B. D. Wandelt, “Fast and exact spin-s spherical harmonic transforms,” *The Astrophysical Journal Supplement Series*, vol. 189, no. 2, p. 255, 2010. [Online]. Available: http://stacks.iop.org/0067-0049/189/i=2/a=255 75

[86] B. Huhle, T. Schairer, A. Schilling, and W. Strasser, “Learning to localize with gaussian process regression on omnidirectional image data,” in *Intelligent Robots and Systems (IROS), 2010 IEEE/RSJ International Conference on*, 2010, pp. 5208–5213. 73

[87] H. Ishiguro and S. Tsuji, “Image-based memory of environment,” in *Intelligent Robots and Systems ’96, IROS 96, Proceedings of the 1996 IEEE/RSJ International Conference on*, vol. 2, 1996, pp. 634–639 vol.2. 68

[88] L. Itti, C. Koch, and E. Niebur, “A model of saliency-based visual attention for rapid scene analysis,” *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 20, no. 11, pp. 1254–1259, 1998. 105

[89] Y. Jia, M. Li, L. An, and X. Zhang, “Autonomous navigation of a miniature mobile robot using real-time trinocular stereo machine,” in *Robotics, Intelligent Systems and Signal Processing, 2003. Proceedings. 2003 IEEE International Conference on*, vol. 1, 2003, pp. 417–421 vol.1. 18

[90] M. Jogan and A. Leonardis, “Robust localization using eigenspace of spinning-images,” in *Omnidirectional Vision, 2000. Proceedings. IEEE Workshop on*, 2000, pp. 37–44. 83

[91] ——, “Robust localization using an omnidirectional appearance-based subspace model of environment,” in *Robotics and Autonomous Systems, vol. 45, no. 1, pp. 51-72*, 2003. 83

[92] M. Joshima, K. Kisimoto, and K. Nishimura, "Sea floor mapping using the data of forward looking sonar and side-scan sonar around the hydrothermal sites, south mariana trough," in *Underwater Technology and Workshop on Scientific Use of Submarine Cables and Related Technologies, 2007. Symposium on*, 2007, pp. 621–626. 11

[93] M. Juliá, A. Gil, and O. Reinoso, "A comparison of path planning strategies for autonomous exploration and mapping of unknown environments," *Autonomous Robots*, vol. 33, no. 4, pp. 427–444, 2012. [Online]. Available: http://dx.doi.org/10.1007/s10514-012-9298-8 27

[94] ——, "Searching dynamic agents with a team of mobile robots," *Sensors*, vol. 12, no. 7, pp. 8815–8831, 2012. [Online]. Available: http://www.mdpi.com/1424-8220/12/7/8815 29

[95] C. Junhua and L. Jing, "Research on color image classification based on hsv color space," in *Instrumentation, Measurement, Computer, Communication and Control (IMCCC), 2012 Second International Conference on*, 2012, pp. 944–947. 116

[96] A. Karimi, M. Danesh, A. Tabibian, and A. Nouri, "Dynamic analysis and path planning for a redundant actuated biped robot," in *Control, Instrumentation and Automation (ICCIA), 2011 2nd International Conference on*, 2011, pp. 1074–1079. 15

[97] M. Karimi, M. Bozorg, and A. Khayatian, "A comparison of dvl/ins fusion by ukf and ekf to localize an autonomous underwater vehicle," in *Robotics and Mechatronics (ICRoM), 2013 First RSI/ISM International Conference on*, 2013, pp. 62–67. 13

[98] T. Kashima, A. Asada, and T. Ura, "The positioning system integrated lbl and ssbl using seafloor acoustic mirror transponder," in *Underwater Technology Symposium (UT), 2013 IEEE International*, 2013, pp. 1–4. 13

[99] M. Kazhdan, T. Funkhouser, and S. Rusinkiewicz, "Rotation invariant spherical harmonic representation of 3d shape descriptors," in *Proceedings of the 2003 Eurographics/ACM SIGGRAPH symposium on Geometry processing*, ser. SGP '03. Aire-la-Ville, Switzerland, Switzerland: Eurographics Association, 2003, pp. 156–164. [Online]. Available: http://dl.acm.org/citation.cfm?id=882370.882392 73

[100] D. Kendall, "A survey of the statistical theory of shape," vol. 4, no. 2, 1989, pp. 87–99. 177, 194

[101] J. Kim and S. Sukkarieh, "Real-time implementation of airborne inertial-slam," *Robot. Auton. Syst.*, vol. 55, no. 1, pp. 62–71, Jan. 2007. 15

[102] König & Meyer GmbH. Stand. [Online]. Available: http://produkte.k-m.de/en/product?info=461&xec51a=um6ioaf8vfl2bl46hf853dnok558

[103] K. Konolige, E. Marder-Eppstein, and B. Marthi, "Navigation in hybrid metric-topological maps," in *Robotics and Automation (ICRA), 2011 IEEE International Conference on*, 2011, pp. 3041–3047. 30

[104] P. Kostelec and D. Rockmore, "Ffts on the rotation group," *Journal of Fourier Analysis and Applications*, vol. 14, no. 2, pp. 145–179, 2008. [Online]. Available: http://dx.doi.org/10.1007/s00041-008-9013-5 74

[105] G. Krishnan and S. Nayar, "Cata-fisheye camera for panoramic imaging," in *Applications of Computer Vision, 2008. WACV 2008. IEEE Workshop on*, 2008, pp. 1–8. 33

[106] C. Kunz, C. Murphy, R. Camilli, H. Singh, J. Bailey, R. Eustice, M. Jakuba, K. Nakamura, C. Roman, T. Sato, R. Sohn, and C. Willis, "Deep sea underwater robotic exploration in the ice-covered arctic ocean with auvs," in *Intelligent Robots and Systems, 2008. IROS 2008. IEEE/RSJ International Conference on*, 2008, pp. 3654–3660. 12, 13

[107] G. H. Lee, F. Faundorfer, and M. Pollefeys, "Motion estimation for self-driving cars with a generalized camera," in *Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on*, 2013, pp. 2746–2753. 15

[108] J. Leonard and H. Durrant-Whyte, "Mobile robot localization by tracking geometric beacons," *Robotics and Automation, IEEE Transactions on*, vol. 7, no. 3, pp. 376–382, 1991. 17

[109] Q. Li, N. Zheng, L. Ma, and H. Cheng, "True single view point multi-resolution catadioptric system for intelligent vehicle," in *Intelligent Transportation Systems, 2004. Proceedings. The 7th International IEEE Conference on*, 2004, pp. 155–160. 33

[110] S. Li, "Full-view spherical image camera," in *Pattern Recognition, 2006. ICPR 2006. 18th International Conference on*, vol. 4, 2006, pp. 386–390. 20

[111] S. Li, M. Nakano, and N. Chiba, "Acquisition of spherical image by fish-eye conversion lens," in *Virtual Reality, 2004. Proceedings. IEEE*, 2004, pp. 235–236. 20

[112] P. Liljeback, O. Stavdahl, K. Pettersen, and J. Gravdahl, "A modular and waterproof snake robot joint mechanism with a novel force/torque sensor," in *Intelligent Robots and Systems (IROS), 2012 IEEE/RSJ International Conference on*, 2012, pp. 4898–4905. 15

[113] A. Lingua, D. Marenchino, and F. Nex, "Performance analysis of the sift operator for automatic feature extraction and matching in photogrammetric applications," *Sensors*, vol. 9, no. 5, pp. 3745–3766, 2009. 61, 87

[114] M. Liu, C. Pradalier, and R. Siegwart, "Visual homing from scale with an uncalibrated omnidirectional camera," *Robotics, IEEE Transactions on*, vol. PP, no. 99, pp. 1–13, 2013. 45

[115] D. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vision*, vol. 60, no. 2, pp. 91–110, Nov. 2004. 23, 61, 87

[116] ——, "Object recognition from local scale-invariant features," in *Computer Vision, 1999. The Proceedings of the Seventh IEEE International Conference on*, vol. 2, 1999, pp. 1150–1157 vol.2. 61

[117] P. Luis, O. Reinoso, A. Gil, J. Pedrero, and M. Ballesta, "Appearance-based multi-robot following routes using incremental pca," in *Knowledge-Based Intelligent Information and Engineering Systems*, ser. Lecture Notes in Computer Science, B. Apolloni, R. Howlett, and L. Jain, Eds. Springer Berlin Heidelberg, 2007, vol. 4693, pp. 1170–1178. [Online]. Available: http://dx.doi.org/10.1007/978-3-540-74827-4_14682

[118] M. Magnabosco and T. P. Breckon, "Cross-spectral visual simultaneous localization and mapping (slam) with sensor handover," *Robotics and Autonomous Systems*, vol. 61, no. 2, pp. 195 – 208, 2013. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0921889012001777 61

[119] I. Mahon and S. Williams, "Slam using natural features in an underwater environment," in *Control, Automation, Robotics and Vision Conference, 2004. ICARCV 2004 8th*, vol. 3, 2004, pp. 2076–2081 Vol. 3. 13

[120] A. Makadia and K. Daniilidis, "Direct 3d-rotation estimation from spherical images via a generalized shift theorem," in *Computer Vision and Pattern Recognition, 2003. Proceedings. 2003 IEEE Computer Society Conference on*, vol. 2, 2003, pp. II–217–24 vol.2. 76

[121] ——, "Rotation recovery from spherical images without correspondences," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 28, no. 7, pp. 1170–1175, 2006. 76

[122] A. Makadia, L. Sorgi, and K. Daniilidis, "Rotation estimation from spherical images," in *Pattern Recognition, 2004. ICPR 2004. Proceedings of the 17th International Conference on*, vol. 3, 2004, pp. 590–593 Vol.3. 76

[123] P. Márquez-Valle, D. Gil, R. Mester, and A. Hernàndez-Sabaté, "Local analysis of confidence measures for optical flow quality evaluation," in *VISAPP 2014, International Conference on Computer Vision Theory and Applications*. Ed. SciTePress - Science and Technology Publications ISBN: 978-989-758-003-3 - Volume 3, pp. 450-457, 2012. 30

[124] J. Matas, O. Chum, M. Urban, and T. Pajdla, "Robust wide baseline stereo from maximally stable extremal regions," in *Proc. BMVC*, 2002, pp. 36.1–36.10, doi:10.5244/C.16.36. 23

[125] J. McEwen, "Fast, exact (but unstable) spin spherical harmonic transforms," *All Res. J. Phys.*, vol. 1, no. 1, 2011. 75

[126] J. McEwen and Y. Wiaux, "A novel sampling theorem on the sphere," *Signal Processing, IEEE Transactions on*, vol. 59, no. 12, pp. 5876–5887, 2011. 75

[127] E. Menegatti, T. Maeda, and H. Ishiguro, "Image-based memory for robot navigation using properties of omnidirectional images," *Robotics and Autonomous Systems*, vol. 47, no. 4, pp. 251 – 267, 2004. 68

[128] K. Mikolajczyk and C. Schmid, "Indexing based on scale invariant interest points," in *Computer Vision, 2001. ICCV 2001. Proceedings. Eighth IEEE International Conference on*, vol. 1, 2001, pp. 525–531 vol.1. 22

[129] ——, "A performance evaluation of local descriptors," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 27, no. 10, pp. 1615–1630, 2005. 23

[130] K. Mikolajczyk, T. Tuytelaars, C. Schmid, A. Zisserman, J. Matas, F. Schaffalitzky, T. Kadir, and L. V. Gool, "A comparison of affine region detectors," *Int. J. Comput. Vision*, vol. 65, no. 1-2, pp. 43–72, Nov. 2005. [Online]. Available: http://dx.doi.org/10.1007/s11263-005-3848-x 24

[131] M. Milford and G. Wyeth, "Persistent navigation and mapping using a biologically inspired slam system," *The International Journal of Robotics Research*, 2009. [Online]. Available: http://ijr.sagepub.com/content/early/2009/07/21/0278364909340592.abstract 45

[132] J. Min, J. Kim, I.-S. Kweon, and Y.-W. Park, "Vision-based metric topological slam," in *Control, Automation and Systems (ICCAS), 2012 12th International Conference on*, 2012, pp. 2171–2176. 30

[133] I. Mondragón, M. Olivares-Méndez, P. Campoy, C. Martínez, and L. Mejias, "Unmanned aerial vehicles uavs attitude, height, motion estimation and control using visual systems," *Autonomous Robots*, vol. 29, no. 1, pp. 17–34, 2010. 14

[134] H. Moravec and A. Elfes, "High-Resolution Maps from Wide-Angle Sonar," in *IEEE International Conference on Robotics and Automation (ICRA)*. Los Alamitos, Calif.: CS Press, 1985. 31

[135] H. Murase and S. Nayar, "Visual learning and recognition of 3-d objects from appearance," *International Journal of Computer Vision*, vol. 14, no. 1, pp. 5–24, 1995. [Online]. Available: http://dx.doi.org/10.1007/BF01421486 78, 82

[136] A. Murillo, J. Guerrero, and C. Sagues, "Surf features for efficient robot localization with omnidirectional images," in *Robotics and Automation, 2007 IEEE International Conference on*, april 2007, pp. 3901 –3907. 61

[137] L. Najman and M. Schmitt, "Watershed of a continuous function," *Signal Processing*, vol. 38, no. 1, pp. 99 – 112, 1994, mathematical Morphology and its Applications to Signal Processing. 23

[138] S. Nayar, "Omnidirectional video camera," in *In Proceedings of the 1997 DARPA Image Understanding Workshop*, 1997, pp. 235–241. 18, 33

[139] S. Nayar, S. Nene, and H. Murase, "Subspace methods for robot vision," *Robotics and Automation, IEEE Transactions on*, vol. 12, no. 5, pp. 750–758, 1996. 82

[140] A. Nemra and N. Aouf, "Robust cooperative uav visual slam," in *Cybernetic Intelligent Systems (CIS), 2010 IEEE 9th International Conference on*, 2010, pp. 1–6. 15

[141] C. H. Nguyen, K. N. Vu, and D. H. Dao, "Applying order reduction model algorithm for balancing control problems of two-wheeled mobile robot," in *Industrial Electronics and Applications (ICIEA), 2013 8th IEEE Conference on*, 2013, pp. 1302–1307. 15

[142] A. Ohte, O. Tsuzuki, and K. Mori, "A practical spherical mirror omnidirectional camera," in *Robotic Sensors: Robotic and Sensor Environments, 2005. International Workshop on*, 2005, pp. 8–13. 33, 34

[143] A. Ohya, A. Kosaka, and A. Kak, "Vision-based navigation by a mobile robot with obstacle avoidance using single-camera vision and ultrasonic sensing," *Robotics and Automation, IEEE Transactions on*, vol. 14, no. 6, pp. 969–978, 1998. 18

[144] K. Okuyama, T. Kawasaki, and V. Kroumov, "Localization and position correction for mobile robot using artificial visual landmarks," in *Advanced Mechatronic Systems (ICAMechS), 2011 International Conference on*, 2011, pp. 414–418. 21

[145] A. Oliva, "Gist of the scene," in *The Encyclopedia of Neurobiology of Attention*, L. Itti, G. Rees, and J. K. Tsotsos, Eds. San Diego, CA: Elsevier, 2005, pp. 251–256. 95

[146] A. Oliva and P. Schyns, "Coarse blobs or fine edges? evidence that information diagnosticity changes the perception of complex visual stimuli." in *Cogn. Psychol. 34, 72-107.*, 1997. 95

[147] A. Oliva and A. Torralba, "Modeling the shape of the scene: a holistic representation of the spatial envelope." in *International Journal of Computer Vision, Vol. 42(3): 145-175.*, 2001. 95, 100

[148] ——, "Building the gist of ascene: the role of global image features in recognition." in *Progress in Brain Reasearch: Special Issue on Visual Perception.Vol. 155.*, 2006. 95

[149] A. Oliva and P. Schyns, "Diagnostic colors mediate scene recognition," *Cognitive Psychology*, vol. 41, no. 2, pp. 176 – 210, 2000. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0010028599907284 95

[150] B. Papalia, W. Prendin, and G. Veruggio, "Sara, an autonomous underwater vehicle for researches in antarctica," in *OCEANS '94. Óceans Engineering for Today's Technology and Tomorrow's Preservation.'Proceedings*, vol. 3, 1994, pp. III/617–III/620 vol.3. 11

[151] L. Payá, F. Amorós, O. Reinoso, L. Fernández, and A. Gil, "An educational software to compare appearance image descriptors in robot localization," in *7th International Technology, Education and Development Conference (INTED 2013)*. Valencia: Ed. IATED, 2013, pp. 3097–3105.

[152] L. Payá, F. Amorós, O. Reinoso, and L. Jiménez, "An educational software to develop robot mapping and localization practices using visual information," in *Advances in Control Education*, vol. 10. Sheffield, United Kingdom: Ed. International Federation of Automatic Control (IFAC), 2013, pp. 174–179.

[153] L. Payá, L. Fernández, A. Gil, and O. Reinoso, "Map building and monte carlo localization using global appearance of omnidirectional images," *Sensors*, vol. 10, no. 12, pp. 11 468–11 497, 2010. [Online]. Available: http://www.mdpi.com/1424-8220/10/12/11468 45

[154] L. Payá, L. Fernández, O. Reinoso, F. Amorós, and L. Jiménez, "A new resource in the teaching of a computer vision and robotics subject," in *7th International Technology, Education and Development Conference (INTED 2013)*. Valencia: Ed. IATED, 2013, pp. 3074–3082.

[155] L. Payá, O. Reinoso, F. Amorós, L. Fernández, and A. Gil, *Multi-Robot Systems, Trends and Development*. Ed. INTECH, 2011, ch. 11: Probabilistic Map Building, Localization and Navigation of a Team of Mobile Robots. Application to Route Following.

[156] L. Payá, A. Vicente, O. Reinoso, C. Fernández, and A. Gil, "Navegación continua de un robot móvil basada en apariencia," in *XXVIII Jornadas de Automática (Huelva)*, 2007. 82

[157] K. Pearson, "On lines and planes of closest fit to systems of points in space," *Philosophical Magazine*, vol. 2, no. 6, pp. 559–572, 1901. 77

[158] V. Peri and S. Nayar, "Generation of Perspective and Panoramic Video from omnidirectional Video," in *DARPA Image Understanding Workshop (IUW)*, May 1997, pp. 243–246. 32

[159] D. M. W. Powers, "Evaluation: From Precision, Recall and F-Factor to ROC, Informedness, Markedness & Correlation," School of Informatics and Engineering, Flinders University, Adelaide, Australia, Tech. Rep. SIE-07-001, 2007. 132

[160] S. Prasad and J. Domke, "Gabor Filter Visualization," 2005. [Online]. Available: http://www.cs.umd.edu/class/spring2005/cmsc838s/assignment-projects 100

[161] D. W. Rees, "Panoramic television viewing system," in *US Patent application 3505465*, April 1970. 33

[162] R. Reid and T. Braunl, "Large-scale multi-robot mapping in magic 2010," in *Robotics, Automation and Mechatronics (RAM), 2011 IEEE Conference on*, 2011, pp. 239–244. 27

[163] S. Roebert, T. Schmits, and A. Visser, "Creating a bird-eye view map using an omnidirectional camera," in *BNAIC 2008: Proceedings of the twentieth Belgian-Dutch Conference on Artificial Intelligence*. Universiteit Twente, Faculteit Elektrotechniek, Wiskunde en Informatica, 2008. 50

[164] S. Sablak and T. E. Boult, "Multilevel color histogram representation of color images by peaks," 1999. 115

[165] D. Scaramuzza, A. Martinelli, and R. Siegwart, "A flexible technique for accurate omnidirectional camera calibration and structure from motion," in *Computer Vision Systems, 2006 ICVS '06. IEEE International Conference on*, jan. 2006, p. 45. 38

[166] D. Scaramuzza and R. Siegwart, "Appearance-guided monocular omnidirectional visual odometry for outdoor ground vehicles," *Robotics, IEEE Transactions on*, vol. 24, no. 5, pp. 1015–1026, 2008. 29

[167] D. Scaramuzza, R. Siegwart, and A. Martinelli, "A robust descriptor for tracking vertical lines in omnidirectional images and its use in mobile robotics," *The International Journal of Robotics Research*, vol. 28, no. 2, pp. 149–171, 2009. [Online]. Available: http://ijr.sagepub.com/content/28/2/149.abstract 24, 43, 62

[168] T. Schairer, B. Huhle, and W. Strasser, "Increased accuracy orientation estimation from omnidirectional images using the spherical fourier transform," in *3DTV Conference: The True Vision - Capture, Transmission and Display of 3D Video, 2009*, 2009, pp. 1–4. 76

[169] ——, "Application of particle filters to vision-based orientation estimation using harmonic analysis," in *Robotics and Automation (ICRA), 2010 IEEE International Conference on*, 2010, pp. 2556–2561. 76

[170] T. Schairer, B. Huhle, P. Vorst, A. Schilling, and W. Strasser, "Visual mapping with uncertainty for correspondence-free localization using gaussian process regression," in *Intelligent Robots and Systems (IROS), 2011 IEEE/RSJ International Conference on*, 2011, pp. 4229–4235. 76

[171] X. Shouzhang and W. Fengwen, "Generation of panoramic view from 360 degree fisheye images based on angular fisheye projection," in *Distributed Computing and Applications to Business, Engineering and Science (DCABES), 2011 Tenth International Symposium on*, 2011, pp. 187–191. 53

[172] W. Shukowsky, "A quadrature formula over the sphere with application to high resolution spherical harmonic analysis," *Bulletin géodésique*, vol. 60, no. 1, pp. 1–14, 1986. [Online]. Available: http://dx.doi.org/10.1007/BF02519350 75

[173] C. Siagian and L. Itti, "Rapid biologically-inspired scene classification using features shared with visual attention," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 2, pp. 300–312, Feb 2007. 105

[174] ——, "Biologically inspired mobile robot vision localization," *Robotics, IEEE Transactions on*, vol. 25, no. 4, pp. 861–873, 2009. 105

[175] K. Siemes, J.-P. Hermand, M. Snellen, and D. G. Simons, "A multi-sensor approach for remotely modeling and mapping sediment properties," in *Acoustics in Underwater Geosciences Symposium (RIO Acoustics), 2013 IEEE/OES*, 2013, pp. 1–5. 11

[176] R. Sim and G. Dudek, "Effective exploration strategies for the construction of visual maps," in *Intelligent Robots and Systems, 2003. (IROS 2003). Proceedings. 2003 IEEE/RSJ International Conference on*, vol. 4, oct. 2003, pp. 3224 – 3231 vol.3. 29

[177] A. R. Smith, "Color gamut transform pairs," *SIGGRAPH Comput. Graph.*, vol. 12, no. 3, pp. 12–19, Aug. 1978. [Online]. Available: http://doi.acm.org/10.1145/965139.807361 116

[178] S. M. Smith and J. M. Brady, "Susan - a new approach to low level image processing," *International Journal of Computer Vision*, vol. 23, pp. 45–78, 1995. 23

[179] P. Suhasini, K. Krishna, and I. Krishna, "Combining sift and invariant color histogram in hsv space for deformation and viewpoint invariant image retrieval," in *Computational Intelligence Computing Research (ICCIC), 2012 IEEE International Conference on*, 2012, pp. 1–4. 116

[180] F. Sun, J. Yu, and D. Xu, "Visual measurement and control for underwater robots: A survey," in *Control and Decision Conference (CCDC), 2013 25th Chinese*, 2013, pp. 333–338. 13

[181] Y.-R. Tang and Y. Li, "Design of an optimal flight control system with integral augmented compensator for a nonlinear uav helicopter," in *Intelligent Control and Automation (WCICA), 2012 10th World Congress on*, 2012, pp. 3927–3932. 13

[182] J. D. Tardós, J. Neira, P. M. Newman, and J. J. Leonard, "Robust mapping and localization in indoor environments using sonar data," *Int. J. Robotics Research*, vol. 21, pp. 311–330, 2002. 16

[183] The Imaging Source Europe GmbH. Color CCD Camera DFK 21BF04. [Online]. Available: http://www.theimagingsource.com/es_ES/products/cameras/firewire-ccd-color/dfk21bf04/ 38, 119

[184] ——. Color CCD Camera DFK 41BF02. [Online]. Available: http://www.theimagingsource.com/es_ES/products/cameras/firewire-ccd-color/dfk41bf02/ 38

[185] C. Tomasi and T. Kanade, "Detection and tracking of point features," International Journal of Computer Vision, Tech. Rep., 1991. 14

[186] R. Triebel and W. Burgard, "Improving simultaneous localization and mapping in 3d using global constraints," in *ASSOCIATION FOR THE ADVANCEMENT OF ARTIFICIAL INTELLIGENCE*, 2005. 17

[187] M. Turk and A. Pentland, "Eigenfaces for recognition," *J. Cognitive Neuroscience*, vol. 3, no. 1, pp. 71–86, Jan. 1991. [Online]. Available: http://dx.doi.org/10.1162/jocn.1991.3.1.71 77, 78, 81

[188] M. Ueonara and T. Kanade, "Optimal approximation of uniformly rotated images: relationship between karhunen-loeve expansion and discrete cosine transform." in *IEEE Transactions on Image Processing. Vol. 7, No. 1, pp. 116-119.*, 1998. 83

[189] D. Valiente, A. Gil, F. Amorós, and O. Reinoso, "SLAM of view-based maps using SGD," in *10th International Conference on Informatics in Control, Automation and Robotics (ICINCO 2013)*. Reykjavik, Iceland: SciTePress - Science and Technology Publications, 2013, pp. 385–392.

[190] D. Valiente, A. Gil, L. Fernández, L. Payá, and O. Reinoso, "Appearance-guided monocular omnidirectional visual odometry for outdoor ground vehicles," *Journal of Robotics*, p. 13, 2012. 29

[191] D. Valiente, A. Gil, L. Fernández, and O. Reinoso, "View-based SLAM using omnidirectional images," in *ICINCO (2)*, 2012, pp. 48–57. 205

[192] ——, "A comparison of EKF and SGD applied to a view-based SLAM approach with omnidirectional images," *Robotics and Autonomous Systems*, vol. 62, no. 2, pp. 108 – 119, 2014. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0921889013002261 28

[193] Q. Wang, O. Ronneberger, and H. Burkhardt, "Rotational invariance based on fourier analysis in polar and spherical coordinates," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 31, no. 9, pp. 1715–1722, 2009. 73

[194] E. P. Wigner, "Group Theory and Its Application to the Quantum Mechanics of Atomic Spectra," *American Journal of Physics*, vol. 28, 1960. 75

[195] O. Wijk and H. Christensen, "Localization and navigation of a mobile robot using natural point landmarks extracted from sonar data," *Robotics and Autonomous Systems*, vol. 31, no. 1â2, pp. 31 – 42, 2000. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0921889099000858 16

[196] N. Winters, J. Gaspar, G. Lacey, and J. Santos-Victor, "Omni-directional vision for robot navigation," in *Omnidirectional Vision, 2000. Proceedings. IEEE Workshop on*, 2000, pp. 21 –28. 29

[197] R. Wood, "Xxiii. fish-eye views, and vision under water," *Philosophical Magazine Series 6*, vol. 12, no. 68, pp. 159–162, 1906. 20, 53

[198] Y. Xu, M. Sun, Z. Cao, J. Liang, and T. Li, "Multi-object tracking for mobile navigation in outdoor with embedded tracker," in *Natural Computation (ICNC), 2011 Seventh International Conference on*, vol. 3, 2011, pp. 1739–1743. 53

[199] Y. Yagi, S. Fujimura, and M. Yachida, "Route representation for mobile robot navigation by omnidirectional route panorama fourier transformation," in *Robotics and Automation, 1998. Proceedings. 1998 IEEE International Conference on*, vol. 2, 1998, pp. 1250–1255 vol.2. 45

[200] Y. Yagi and S. Kawato, "Panorama scene analysis with conic projection," in *Intelligent Robots and Systems '90. 'Towards a New Frontier of Applications', Proceedings. IROS '90. IEEE International Workshop on*, 1990, pp. 181–187 vol.1. 33

[201] T. Yairi and H. Kanazaki, "Bearing-only mapping by sequential triangulation and multi-dimensional scaling," in *Robotics and Automation, 2008. ICRA 2008. IEEE International Conference on*, 2008, pp. 1449–1454. 27

[202] K. Yamazawa, Y. Yagi, and M. Yachida, "Obstacle detection with omnidirectional image sensor hyperomni vision," in *Robotics and Automation, 1995. Proceedings., 1995 IEEE International Conference on*, vol. 1, 1995, pp. 1062–1067 vol.1. 33

[203] L.-J. Yang, C.-K. Hsu, J.-Y. Ho, H.-H. Wang, and G.-H. Feng, "The micro aerial vehicle (mav) with flapping wings," in *Mechatronics, 2005. ICM '05. IEEE International Conference on*, 2005, pp. 811–815. 13

[204] Z. Yong-guo, C. Wei, and L. Guang-liang, "The navigation of mobile robot based on stereo vision," in *Intelligent Computation Technology and Automation (ICICTA), 2012 Fifth International Conference on*, 2012, pp. 670–673. 18

[205] K. Yoshida, H. Nagahara, and M. Yachida, "An omnidirectional vision sensor with single viewpoint and constant resolution," in *Intelligent Robots and Systems, 2006 IEEE/RSJ International Conference on*, 2006, pp. 4792–4797. 33

[206] C. Yu and D. Zhang, "A new 3d map reconstruction based mobile robot navigation," in *Signal Processing, 2006 8th International Conference on*, vol. 4, 2006, pp.–. 17

[207] Z. Yuyi, G. Zhenbang, W. Lei, Z. Ruiyong, and L. Huanxin, "Study of underwater positioning based on short baseline sonar system," in *Artificial Intelligence and Computational Intelligence, 2009. AICI '09. International Conference on*, vol. 2, 2009, pp. 343–346. 12

[208] L. Zhang and B. Ghosh, "Line segment based map building and localization using 2d laser rangefinder," in *Robotics and Automation, 2000. Proceedings. ICRA '00. IEEE International Conference on*, vol. 3, 2000, pp. 2538–2543 vol.3. 17

[209] Q. Zhu, M.-C. Yeh, K.-T. Cheng, and S. Avidan, "Fast human detection using a cascade of histograms of oriented gradients," in *Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on*, vol. 2, 2006, pp. 1491–1498. 87

[210] Z. Zhu, S. Bhattacharya, M. de Haag, and W. Pelgrum, "Using single-camera geometry to perform gyro-free navigation and attitude determination," in *Position Location and Navigation Symposium (PLANS), 2010 IEEE/ION*, 2010, pp. 858–867. 18